Patho-adaptive evolution of *Haemophilus* spp. bacterial respiratory colonizing opportunistic pathogens

TESIS DOCTORAL

Javier Moleres Apilluelo

Pamplona, 2018

Directora: Junkal Garmendia García

Agradecimientos

Quiero agradecer a todas las personas e instituciones que, cada cual a su manera, me han ayudado durante esta etapa.

A mi directora, Junkal. Gracias por transmitirme esa pasión por la ciencia. Gracias por todo tu apoyo y paciencia. Por tanto tiempo, esfuerzo y dedicación empleado en mi formación. Gracias

A mis padres, mi mayor e incondicional apoyo en todos los sentidos. Mis mayores ídolos y fans. A ti amaxi. Sin ti no hubiera llegado a ningún lado, tanto personal como académicamente. Gracias por tu comprensión, todos tus ánimos, y por ser la mejor, simplemente la mejor. Gracias por estar siempre ahí. Os quiero.

A mis hermanos y familia. Gus y Ves, por ser mis elder bros y un apoyo incondicional y brutal, en todos y cada unos de las etapas de mi vida. Que durante la tesis han pasado muchos... A Pol, por ser mi kalandraka favorito. Va por ti bro! A Sar, por ser imprescindible, y a Nachete, un trozo de pan. A mis tías Patty y Shelma. Sois las mejores, siempre me habéis ayudado incondicionalmente, y apoyado en todo. Desde que tengo uso de la razón. Zuk ere, Nieves, eskerrik asko zaudelako. Ama bat niretzat, adibide bat eta poza handia. De corazón, gracias.

A mis amigos. Todo el Félix, sin excepción. Especialmente gracias a tí, Mutis. Por estar siempre ahí. Por acogerme en los malos momentos y no dejarme en paz. Por contar siempre conmigo en los buenos, por todas esas charlas y consejos que no tienen precio. Por arrastrarme al Box. Por conseguirme trabajo, uno tras otro. Por ser como eres. Un fenómeno. También a ti Gori. Te debo mucho. Por esas largas discusiones con Brutus, no te imaginas lo que me han ayudado, en todo. A Dito, Mehdi y Nachete, a todos.

A mis compañeros de trabajo. Desde los Paus, pasando por Cris, sus risas y villancicos, Begoñita y su meticuloso buen hacer y sus constantes ánimos, Lucius y sus chistes chistosos y su perpetua sonrisa, Nahi por ser una pequeña gran compañera en todos los sentidos. A Falete, por ser como eres, un personaje. Siempre un apoyo y unas risas. A Ariadnita, por las risas que nos echamos cuando sacas el hacha, por responder con una sonrisa siempre que te he pedido ayuda. Y a ti Irenota. Por ser mi constante apoyo. Mi psicóloga un poco macarra. Por aguantarme. Entenderme. Ayudarme. Siempre. Por todo lo que hemos pasado en estos años. Y por ser como eres, gracias. A todos, gracias.

A tí, Esther. Te lo debo todo. Nunca hubiese llegado hasta aquí si no me hubieses contagiado un poco de tu sentido de la responsabilidad, de la organización. Me has ayudado a llegar hasta donde estoy y a ser quien soy. Tengo muchísimo que agradecerte y nunca voy a dejar de hacerlo. También a Charo y a Luismari, Elena y Edu. Os debo mucho. Muchísimo. Eternamente agradecido.

A todos mis compañeros del IdaB. A todos los ILes, especialmente a Kike que me ayudaste a llegar hasta aquí. A mi segunda madre, Edurne. A Paquito, a Ram, a Cris Solano, Victor, un crack. Idoia, Igoriko. A tod@s.

A Alfonsito, Andreas, Bruno y todos los demás. Sois increíbles. Colaborar con vosotros ha sido una gran suerte. Me siento muy afortunado.

To Josh. Thanks dude, for everything you have made for me. I have had the opportunity to learn so much from you, that I only miss having been able to spend more time learning from you. Thanks also to Rachel, Josh Earl, Carol, Sergei, Garth. To all. Thanks for making it possible.

A isidro, porque el capítulo 1 de esta Tesis no hubiese sido posible sin ti.

A mis sozios de la frathouse. Asier, David y Juan Carlos. Que me adoptasteis cuatro meses en Philly, me tratasteis como un hermano. Y la de historietas que pasamos por allá. Las recuerdo frecuentemente, y espero veros pronto y recordarlas.

A todos los que habéis hecho que pudiera llegar hasta aquí, y ser como soy. Eskerrik asko. Bihotzetik.

Dedicado a mis gorditos, Asier & Lucía.

Este trabajo de Tesis se ha llevado a cabo mediante el disfrute de un contrato de Formación de Personal Investigador (FPI) con referencia BES-2013-062644 por parte de Javier Moleres Apilluelo, adscrito a los proyectos SAF2012-31166 (enero 2014-diciembre 2015), SAF2015-66520-R (enero 2016-diciembre 2017) concedidos a la Dra. Junkal Garmendia García.

La **Dra. Junkal Garmendia García**, Científico Titular del Consejo Superior de Investigaciones Científicas (CSIC), adscrita al grupo de Sanidad Animal del Instituto de Agrobiotecnología,

INFORMA,

Que la presente memoria de Tesis Doctoral "Patho-adaptive evolution of *Haemophilus* spp. bacterial respiratory colonizing opportunistic pathogens" elaborada por **Javier Moleres Apilluelo** ha sido realizada bajo su dirección, y que cumple las condiciones exigidas por la legislación vigente para optar al grado de Doctor.

Y para que así conste, firma la presente en Pamplona/Iruña, a 7 de Febrero de 2018

Fdo. Dra Junkal Garmendia García

Index

Abbreviations

A	Adenine
Aa	Aminoacid
AA	Arachidonic acid
AATD	Alpha-1 antitrypsin
AECOPD	Acute exacerbation of Chronic Obstructive Pulmonary Disease
AMC	Amoxicillin-clavulanic acid
AMP	Antimicrobial peptide
Amp	Ampicillin
BALF	Bronco-alveolar lavage fluid
Bcc	*Burkholderia cepacia* complex
BHi	Brain Heart Infusion
Bp	Base pair
C	Cytosine
CB	Chronic bronchitis
CEC	Cefaclor
ChoP	Phosphorylcholine
CID	Cefonicid
CIP	Ciprofloxacin
CF	Cystic Fibrosis
C.F.U.	Colony forming unit
CFU	Cefuroxime
CLSI	Clinical Laboratory Standards Institute
cm	centimeter
Cm	Chloramphenicol
CTX	Cefotaxime
COPD	Chronic Obstructive Pulmonary Disease
COP	Colonizing opportunistic pathogen
CT	Clonal type
ºC	Celsius degree
DAM	Deoxyadenosine methylase
DLCO	Carbon monoxide diffusion in the lung
dN	Non-synonymous genetic changes
dNTP	deoxynucleotide triphosphate
DNA	Deoxyribonucleic acid
dS	Synonymous genetic changes
DTT	Dithiothreitol
3D	3 dimensions
EBSS	Earle´s balance salt solution
ELISA	Enzyme-linked immunoabsorbent assay
ERIC	Enterobacterial repetitive intergenic consensus
FA	Fatty acid
FAM13A	Family with sequence similarity 13 member A gene
FCS	Fetal calf serum
G	Guanine
GC-MS	Gas chromatography-mass spectrometry
gDNA	Genomic DNA
GWAS	Genome wide association studies
μg	microgram
h	hour
hBD-1	Human β-defensin 1
hCEACAM1	Human carcinoembryonic antigen-related cell adhesion molecule 1

HGT	Horizontal gene transfer
Hib	*Haemophilus influenzae* type b
Hip	Hedgehog interacting protein
HIV	Human immunodeficiency virus
HMW	High molecular weight protein
hpi	Hours post-infection
HTM	*Haemophilus* test medium
HUB	Hospital Universitario de Bellvitge
HUGTiP	Hospital Universitario German Trias i Pujol
Ig	Immunoglobulin
ICAM-1	Intercellular adhesion molecule-1
ID	Identity
IL-8	Interleukin 8
Indel	Insertion/deletion
Kb	Kilobase
KDa	KiloDalton
ºK	Kelvin degree
Lamp-1	Lysosome associated membrane protein-1
LB	Luria Bertani
LBP	Lipopolysaccharide-binding protein
LCFA	Long chain fatty acid
LDA	Lauryl dimethylamine-N-oxide
LOS	Lipooligosacharide
LPS	Lipopolysaccharide
µl	microliter
mA	miliAmpere
MALDI-TOF	Matrix-associated laser desorption ionization-time of flight
Mb	Megabase
MD	Molecular dynamics
MEM	Minimum essential medium eagle
MH-F	Mueller Hinton fastidious
MIC	Minimum inhibitory concentration
min	minute
MLST	Multilocus sequence typing
mM	milimolar
MM-FFA	Minimal medium free of fatty acids
MMP	Matrix metalloproteinase
mMRC	Modified Medical Research Council scale
MOI	multiplicity of infection
µM	micromolar
mRNA	messenger RNA
NAD	Nicotinamide adenine dinucleotide
NMA	Normal mode analysis
Nal	Nalidixic Acid
NCBI	National Center for Biotechnology Information
NET	Neutrophils extracellular trap
Nov	Novobiocin
NTHi	Non-typeable *Haemophilus influenzae*
NTHi-CV	NTHi-containing vacuole
nm	nanomolar
ns	nanosecond
OA	Oleic acid

OCD	Chronic coughing
OD	Optical density
OMP	Outer membrane protein
ORF	Open reading frame
PacBio	Pacific Biosciences
PAF	Platelet-activating factor
PAF-R	Platelet-activating factor receptor
PAGE	Polyacrilamide gel electrophoresis
PBD	Post-bronchodilator test
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PFGE	Pulse field gel electrophoresis
PLA_2	Phospholipase A_2
PxB	Polymixin B
PxE	Polymixin E
ps	picosecond
QTL	Quantitative trait locus
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
sBHI	Supplemented BHI
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SEM	Standard error of the mean
SGH	Supragenome hybridization
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SOP	Simple opportunistic pathogen
Spc	Spectinomycin
ST	Sequence type
SSR	Single sequence repeat
Str	Streptomycin
SXT	Cotrimoxazole
T	Thymine
TET	Tetracycline
TIMP	Tissue inhibitor of metalloproteinase
TREP	Transformed recombinant enriched profiling
Tris-HCl	(hidroximetil)aminomethane-chloridic acid
USS	Uptake signal sequence
vtaA	Virulence associated trimeric autotransporter A
w/v	Weight/volume
WGS	Whole genome sequencing
WT	Wild type

LIST OF FIGURES:

General Introduction:

Chapter 1:

Chapter 2:

Chapter 3:

LIST OF TABLES:

General Introduction:

Chapter 1:

Chapter 2:

Chapter 3:

study.

Table S3.4. FadL allelic variants encoding full-length proteins.

Table S3.5. FadL allelic variants encoding truncated proteins.

Table S3.6. Primers used in this study.

Table S3.7. Genome sequencing, assembly, and annotation metrics.

Table S3.8. DNA modification/motif report from PacBio assemblies. List of of NTHi database and sequencing metrics used in this study.

Table S3.9. List of of NTHi database and sequencing metrics used in this study.

Table S3.10. Coding regions SNVs in strains from CTs 9 and 76.

ABSTRACT

This PhD Thesis work has been focused on three independent aspects of the patho-adaptation by two closely related bacterial species of host-restricted respiratory colonizing opportunistic pathogens. Bacterial patho-adaptive evolution has been considered in both experimental (Chapter 2) and within-host (Chapters 1 and 3) settings.

Chapter 1 presents the identification and characterization of a novel natural plasmid responsible for acquisition and transmission of resistance to β-lactamic antibiotics. Antibiotics are often used in swine disease control, and resistance to several antibiotics has been described in *Haemophilus parasuis*, an early colonizer of the nasal mucosa of piglets and the causative agent of Glässer's disease. We isolated 86 *H. parasuis* strains from nasal swabs collected from healthy piglets at weaning in farms without disease. A novel 2,661-bp plasmid, named pJMA-1, bearing the bla_{ROB-1} β-lactamase was detected among colonizing *H. parasuis*. pJMA-1 was shown to be stable and to have a low biological cost. pJMA-1 was not found in *Haemophilus influenzae* strains from different clinical origins. Altogether, commensal *H. parasuis* represents a reservoir of β-lactam resistance genes, which could be transferred to other bacterial pathogens.

Chapter 2 presents a genetic screening method based on gain-of-function named transformed recombinant enrichment profiling (TREP), where bacterial natural transformation is used to generate complex pools of recombinants, phenotypic selection is used to enrich for specific recombinants, and deep sequencing is used to survey for the genetic variation responsible for the gained phenotypic trait. We applied TREP to investigate the genetic architecture of intracellular invasion by the human-restricted opportunistic pathogen *Haemophilus influenzae*. TREP identified the HMW1 adhesin as a crucial factor for epithelial entry, and invasion rates were enhanced by the *hmw1A* allele of Hi 86-028NP strain. Bacterial self-aggregation and adherence to airway cells were also improved in recombinants, suggesting that the high invasiveness induced by $hmw1A_{86\text{-}028NP}$ might be a consequence of these phenotypes. Finally, intracellular $hmw1A_{86\text{-}028NP}$ bacteria likely invaded as groups, suggesting an emergent invasion-specific consequence of HMW1A-mediated self-aggregation.

Chapter 3 describes patho-adaptive evolution by nontypeable *H. influenzae* (NTHi) during lower airways infection of patients suffering Chronic Obstructive Pulmonary Disease (COPD). We sequenced the genomes of 94 NTHi isolates that were serially collected from sputum samples of 13 COPD patients over a 14 years window time. Recurrent coding mutations were seen in several genes. Most prominently, about a third of clonal types were polymorphic for loss-of-function alleles of the *fadL* gene, which encodes a bifunctional membrane protein that both directly binds the hCEACAM1 receptor and transports exogenous fatty acids. We showed that this is a likely case of antagonistic pleiotropy, in which loss of FadL reduces bacterial ability to attach to- and invade human cells, but also increases its resistance to bactericidal fatty acids enriched within the COPD lung.

RESUMEN

Esta Tesis Doctoral se ha centrado en tres aspectos de la pato-adaptación de dos especies bacterianas relacionadas filogenéticamente, que colonizan de forma asintomática y también son patógenos oportunistas en los sistemas respiratorios porcino y humano, respectivamente. Para ello, hemos analizado rasgos de evolución pato-adaptativa bacteriana en un modelo de trabajo experimental (Capítulo 2) y de forma natural dentro del hospedador (Capítulos 1 y 3).

El Capítulo 1 presenta la identificación y caracterización de un nuevo plásmido natural, responsable de la adquisición y transmisión de resistencia a antibióticos β-lactámicos. El uso de antibióticos en el sector porcino está extendido para el control de infecciones de origen bacteriano. La resistencia a varios antibióticos ha sido descrita en *Haemophilus parasuis*, una bacteria colonizadora de la mucosa nasal porcina, y a su vez causante de la enfermedad de Glässer. En este trabajo, se aislaron 86 cepas de *H. parasuis* a partir de muestras tomadas de la cavidad nasal de lechones sanos próximos al destete, en granjas porcinas con ausencia de enfermedad de Glässer. En esta colección de aislados, se identificó un *subset* de cepas de *H. parasuis* portadoras de un plásmido de 2,661 pb, que denominamos pJMA-1, que contiene la β-lactamasa ROB-1 (bla_{ROB-1}). pJMA-1 se mantiene de manera estable y su presencia conlleva un bajo coste biológico. pJMA-1 no fue encontrado tras la realización de un muestreo en aislados de la bacteria *Haemophilus influenzae* con diferentes origenes patológicos. En conjunto, este trabajo muestra que *H. parasuis* representa un reservorio de genes de resistencia a antibióticos β-lactámicos, en elementos genéticos móviles potencialmente transmisibles a otras bacterias patógenas.

El Capítulo 2 presenta un método de escrutinio genético basado en ganancia de función, que denominamos "evaluación de enriquecimiento de recombinantes generados mediante transformación" (*transformed recombinant enrichment profiling*, TREP), en el que empleamos transformación natural bacteriana para generar *pools* complejos de recombinantes, selección fenotípica para enriquecer los *pools* en recombinantes específicos que han adquirido un fenotipo de interés, y secuenciación masiva para identificar la variación genetica asociada a la ganancia del function seleccionada. Utilizamos TREP para investigar la arquitectura genetica del fenotipo de invasión epitelial del patógeno oportunista humano *Haemophilus influenzae*. TREP identificó la adhesina HMW1 como un factor clave para la invasión epitelial por esta bacteria, en concreto, el alelo de *hmw1A* presente en la cepa Hi 86-028NP incrementó la tasa de invasión epitelial de forma notable. Las bacterias recombinantes con ganancia de capacidad invasiva aumentaron también su auto-agregación y adhesión a epitelio respiratorio, sugiriendo que la alta invasividad favorecida por $hmw1A_{86\text{-}028NP}$ puede ser consecuencia de ambos fenotipos. Por ultimo, las bacterias intracelulares portadoras de $hmw1A_{86\text{-}028NP}$ mostraron un fenotipo invasivo en grupos, que puede ser consecuencia de la auto-agregación mediada por HMW1A.

El Capítulo 3 describe la evolución pato-adaptativa de *H. influenzae* no tipificable (HiNT) durante la infección de las vías respiratorias bajas en pacientes que sufren Enfermedad Pulmonar Obstructiva Crónica (EPOC). Secuenciamos los genomas completos de 94 aislados previamente recogidos de forma longitudinal a partir de muestras de esputo de 13 pacientes EPOC durante 14 años. El análisis genómico comparado reveló mutaciones recurrentes en varios genes. En concreto, aproximadamente un tercio de los tipos clonales determinados tras análisis filogenómico mostró polimorfismos que causan pérdida de función en el gen *fadL*. El gen *fadL* codifica una proteína de membrane externa con dos funciones bien diferenciadas, por un lado es un ligando bacteriano del receptor eucariota hCEACAM1 y, por otro, es un transportador de ácidos grasos

exógenos. Mediante un panel de experimentos diseñados específicamente para el análisis de esta bifuncionalidad, demostramos que la variación de FadL en el pulmón EPOC puede ser un caso de antagonismo pleiotrópico, en el que su pérdida de función reduce la capacidad de *H. influenzae* para adherirse e invadir celulas epiteliales y, al mismo tiempo, aumenta la resistencia bacteriana al efecto bactericida de los ácidos grasos, cuya presencia se ve aumentada en el pulmón de los pacientes EPOC.

GENERAL INTRODUCTION

1. Evolution by natural selection

All living organisms adapt when challenged by new conditions, and improve their fitness by changing their phenotypes using genetic and non-genetic mechanisms. Adaptation involves physiological acclimation, epigenetic changes, genome structure re-arrangements, and changes in DNA sequences. Physiological responses by the means of gene expression changes are often first to emerge upon environmental changes, but are not actively amplified, memorized or propagated. Epigenetic adaptations have varying degrees of self-perpetuation over time, and may occur at the DNA and chromatin, RNA or protein levels. DNA copy-number adaptations include DNA duplications/deletions ranging from specific genes to whole chromosomes. Genomic mutations represent the ultimate level of adaptation, where specific changes are stored and inherited. Adaptation levels differ by the time needed for the adaptation to be attained at the organism level, the time to become frequent in the population, the duration through which it is sustained beyond the presence of a selective pressure, and the faithfulness/accuracy at which it propagates across generations (Yona et al., 2015).

Following this notion, evolution by natural selection establishes that any population consists of individuals that are all slightly different to each other, and those having variations giving them an advantage to successfully reproduce are the ones passing on their traits more frequently to the next generation (Darwin, 1859). Natural selection is the evolutionary process explaining the match between features of organisms and the environments where they life, and selective pressure is any reason for organisms with certain genotypes/phenotypes to have a survival benefit or a disadvantage. Selective pressures drive natural selection in such a way that some members of the population will not survive and reproduce. Gradually, the population changes, genes improving survival and reproduction become more common, while those disadvantageous become more rare. Ultimately, evolution is this change in the genetic makeup of a population (**Figure 1**).

1.1. Bacterial pathogens adaptive evolution by natural selection

The role of natural selection in shaping morphological, physiological and behavioral adaptations of animals and plants across generations is central to understanding life and diversity. In contrast, understanding the role of natural selection in producing adaptive solutions has historically been contentious when it comes to microorganisms. However, bacteria are unique for their ability to adapt rapidly to different environments. Their generation times are often short, leading to a rapid selection of beneficial genetic traits

within a population (Lenski, 2017). Moreover, bacterial populations large sizes facilitate genetic diversity, in turn likely enhanced by Horizontal Gene Transfer (HGT) (Morjan & Rieseberg, 2004; Vogan & Higgs, 2011). It is currently clear that microorganisms enable us to observe the power of adaptive evolution by natural selection with exceptional rigor and clarity (Lenski, 2017).

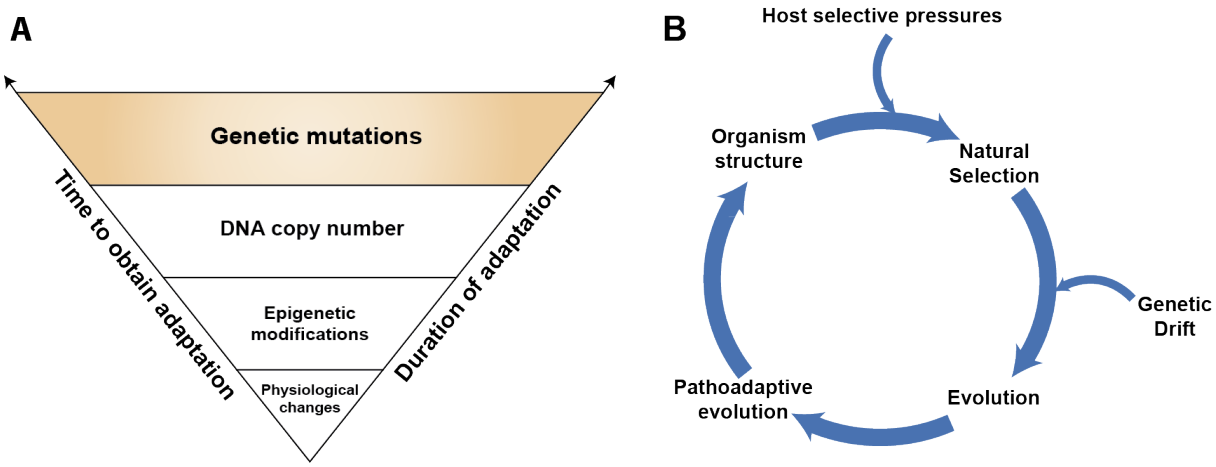


Figure 1. Adaptation toolbox (A) and evolutionary cycle (B) defining the genomic architecture of evolving living organisms. Panel A is adapted from Yoana et al. 2015.

In general, **pathogenic bacteria** can be considered as frank (true or strict) or opportunistic (facultative). **Opportunistic pathogens** cause disease when introduced within a susceptible body site or when hosts are immunologically compromised. The reservoirs of opportunistic pathogens are diverse, including water, soil, animals, and human individuals with active infections. Opportunistic pathogens can be in turn divided into colonizing opportunistic pathogens (COPs) and non-colonizing/simple opportunistic pathogens (SOPs). COPs colonize asymptomatically the human body and, when the conditions are right, cause symptomatic infections. SOPs are present in environmental reservoirs, thereby also named colonizing pathogens of environmental origin (Price et al., 2017). COPs can take up long-term residence in-/on the host body as part of its microbiome. The hologenome concept of evolution postulates that the holobiont (host+symbionts) with its hologenome (host genome+microbiome) is itself a level of selection in evolution. Thus, the hologenome may contribute to its host anatomy, physiology, development, immunity, behavior, genetic variation and evolution. Acquisition of microbes and microbial genes is a powerful mechanism for driving the evolution of complexity, and evolution proceeds via cooperation and competition working in parallel (Rosenberg & Zilber-Rosenberg, 2016). In this context, colonization by COPs is asymptomatic and transmission can occur without detection, which may be an important

risk factor for subsequent disease (Price et al., 2017).

Patho-adaptive evolution refers to any microbial adaptation that confers pathogenicity (Sokurenko et al., 1999). Adaptive evolution of bacterial opportunistic pathogens of environmental origin has been widely studied through gene expression profiling, whole genome sequencing (WGS), systematic phenotyping or genome wide association studies (GWAS), by using both experimental and natural within-host approaches (Jules & Buchrieser, 2007; Lee et al., 2017; Meacci et al., 2005; Vidigal et al., 2014; Wong et al., 2012). Pathogens of environmental origin transit from environmental to host niches. In the environment, these bacteria may have lifestyles associated with other organisms (Jackson et al., 2011), as sheltered by invertebrates, plants or protozoa, which constitute hotspots for genetic exchanges (Moliner et al., 2009). Bacteria overcome their shelter defenses with factors that might be useful for further adaptation to mammalian hosts (Waterfield et al., 2004). Thus, environmental hotspots of emergence and pre-adaptation nurse opportunistic pathogens (Berg et al., 2005; Greub & Raoult, 2004; Scully & Bidochka, 2006). Large genome size and genomic fluidity by intragenomic rearrangement or acquisition of foreign DNA contribute to bacterial versatility during environmental pre-adaptation (Boussau et al., 2004). In contrast, transition to mammalian hosts often involves genome reduction, mainly in genes corresponding to biosynthetic pathways required to synthesize molecules already available within the host, and also in regulatory elements, since living within a host may eliminate extreme environmental fluctuations encountered by free-living bacteria (Moran, 2002). Virulence gene loss-of-function is a frequent trait in this adaptation process, and attenuation may constitute a general mechanism for pathogens ultimately facilitating host colonization (Zdziarski et al., 2008).

During pre-adaptation and/or environment to host transition, bacteria also face physiological changes (i.e. temperature, pH, salinity, O_2, nutrients availability) and host immunity through changes in gene expression (Ledala et al., 2010; Pin et al., 2012; Rothenbacher & Zhu, 2014; Yung et al., 2016). In this context, phase variation is a high frequency, reversible ON/OFF, or graded switching of gene expression mediated through either genetic [e.g., due to variations in the number of simple tandem DNA repeats (Henderson et al., 1999; Moxon et al., 2006)] or epigenetic [e.g., *via* deoxyadenosine methylase (Dam) (Adhikari & Curtis, 2016)] mechanisms at individual genes or promoters. For most phase-variable genes, switching occurs randomly during genome replication and thus gene expression becomes impossible to predict. Phase variation has been characterized in numerous pathogenic bacterial species (Bayliss et al., 2001;

Bayliss et al., 2012; Bergman et al., 2006; Garcia-Pastor et al., 2018; Goldberg et al., 2014; Jennings et al., 1999; van der Woude & Baumler, 2004).

However, free-living bacterial species may accumulate genome insertion/deletion and rearrangements at much higher frequencies than host-restricted bacteria. In the latter, within-host evolution often occurs by nucleotide substitution (Tamas et al., 2002), consisting of substitutions of one nucleotide by another, nucleotide insertion or deletions. Single nucleotide polymorphisms (SNP) lead to synonymous (dS) or non-synonymous (dN) mutations, and nucleotide insertions or deletions result in frame shifts. Within-host bacterial evolution has been mostly tackled for opportunistic pathogens of environmental origin, but barely for host-restricted opportunistic pathogens (for details, see **Figure 2**).

This PhD Thesis work has been focused on the patho-adaptation of two closely related species of host-restricted respiratory colonizing opportunistic bacterial pathogens.

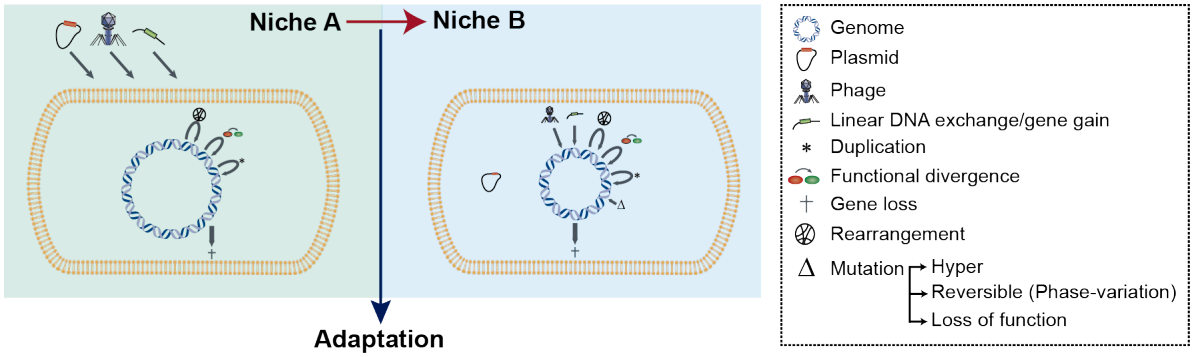


Figure 2. Molecular mechanisms of bacterial genome architecture rearrangement. Arrows pointing toward the bacterial genome indicate possible ways to acquire genetic material by HGT (i.e. by plasmids, phages, natural transformation). Arrows looping back into the bacterial genome indicate changes within the genome, not necessarily involving acquisition of genetic material. Arrows pointing away from the genome indicate loss of genetic material. These changes can occur during bacterial adaptation from niche A (for example, the environment) to niche B (for example, a human host). (Adapted from Toft and Andersson, 2010).

1.2. Outcomes of bacterial patho-adaptive evolution

Adaptations shape bacterial genomes towards increasing their fitness in the host niche. Naturally selecting forces can be measured in terms of positive, neutral or negative selection. **Positive selection** refers to the trend for an allele that confers a survival advantage to increase in frequency and become fixed at a higher rate, being the rapid increase in frequency and fixation of an advantageous allele a selective sweep. Positive selection is achieved when the ratio of dN genetic changes is higher than that of dS ones

(dN/dS>1). **Diversifying selection** is a form of recurrent positive selection that favors the emergence of new alleles in a population (i.e., the selective pressure of host antibodies on pathogen antigenic variation) (Li et al., 2012). In contrast, **purifying selection** refers to the tendency for an allele that incurs a survival disadvantage to decrease in frequency and become lost. Moreover, within-host bacterial diversity is shaped by **random genetic drift**, which refers to variations in allele frequency in a population caused by the random genetic sampling that occurs during birth and death. Finally, when selecting for valuable genetic traits entailing significant fitness cost disadvantages, **compensatory mutations** may arise, i.e. genomic changes that redress the fitness cost of mutations conferring adaptation to specific selection pressures (Schulz zur Wiesch et al., 2010). Without compensatory mutations, adaptations that incur a fitness cost may be lost upon selection pressure removal (Didelot et al., 2016).

Global patterns of patho-adaptive evolution are enriched by identification of recurrent patterns of evolution implicated in pathogenesis, when comparing the genomic adaptation of a single pathogenic bacterial strain in multiple host subjects, i.e. **parallel evolution**. This notion is complemented by the so called **convergent evolution**, the occurrence of mutations resulting in the same phenotype in two or more independently evolving lineages of a bacterial pathogen (Didelot et al., 2016).

In addition, bacterial genes can exhibit **antagonistic pleiotropy**, i.e. a single gene accounts for more than one phenotype, some of which are beneficial to the pathogen and some other deleterious (Bliven & Maurelli, 2016; Williams, 1957). In this context, a gene may confer a selective advantage within one particular environment, but be detrimental within a different environment. Many bacteria avoid the deleterious effect of a gene through its inactivation; loss-of-function mutants can then out-compete their respective wild type (WT) strains and eventually dominate the population (Bliven & Maurelli, 2016). However, certain bacteria simply tolerate deleterious fitness costs if the benefits of expressing the gene outweigh the negative effects. This is the scenario for antibiotic resistance acquisition (Melnyk et al., 2015), which often involve a significant fitness disadvantage, albeit the overall result is strongly beneficial.

1.3. Molecular mechanisms of bacterial genome architecture rearrangement

Bacterial patho-adaptation can occur through **horizontal gene transfer** (HGT), which is the movement of genetic material between organisms other than by the vertical transmission of DNA from parent to offspring, happening across bacterial strain and species boundaries (Baltrus, 2013). Bacterial HGT occurs through transduction,

conjugation or transformation. Transduction involves movement of bacterial DNA from one bacterium to another by a bacteriophage, and conjugation involves DNA transfer by a plasmid from a donor to a recipient cell during cell-to-cell contact (Arber, 2014). Circular and linear DNA can be acquired and/or transmitted through HGT (for details, see **Figure 2**).

Plasmids are extra-chromosomal genetic elements carrying genes that often provide bacteria with genetic advantages. Thus, plasmids play a major role in the acquisition and transmission of antibiotic resistance in bacterial populations, a current major threat to human health (Bennett, 2008).

Bacterial **natural genetic transformation** involves active uptake of free (extracellular) DNA by a cell (Lorenz & Wackernagel, 1994). For natural transformation to happen, bacteria need to reach a physiological state of competence, a temporary state of being able to take up exogenous DNA from the environment, proposed to involve from 20 to 50 proteins, depending on the bacterial species (Chen & Dubnau, 2004). Active excretion of DNA and release of DNA from dead bacteria to the environment are ubiquitous sources of DNA availability. After exposure to Gram-negative competent bacteria, exogenous DNA binds non-covalently to cell surface sites, and double-stranded DNA is converted to single-stranded DNA during translocation across the inner membrane. Most bacterial species can actively take up DNA independently of its sequence, although some bacterial species are selective in the DNA they translocate across the membrane based on short, interspersed nucleotide motifs (Smith et al., 1999; van Belkum et al., 1998). Once reached the cytoplasm, DNA is integrated into the bacterial chromosome by homologous **recombination**, which requires species-specific variable length regions with high similarity between DNA sequences to initiate DNA pairing and strand exchange. Besides DNA replacement, natural transformation can also lead to acquisition of exogenous DNA material with no need to exchange.

Chapter 1 of this PhD Thesis work presents the **identification and characterization of a novel natural plasmid responsible for the acquisition and transmission of resistance to β-lactamic antibiotics**.

Chapter 2 of this PhD Thesis work presents a **genetic screening method based on gain-of-function named *transformed recombinant enrichment profiling* (TREP), where bacterial natural transformation is used to generate complex pools of recombinants, phenotypic selection is used to enrich for specific recombinants, and deep sequencing is used to survey for the genetic variation responsible for**

the gained phenotypic trait.

1.4. Selective pressures and pathogen evolutionary arms race within the host

Upon infection, bacterial pathogens encounter host immunity selective pressures. Evolutionary adaptation allows pathogens overcoming such selective pressures and facilitates colonization and/or infection. Selective pressures at the mammalian host airways are next summarized.

Innate immunity at the human airways defends the air spaces from a whole array of particles and microbial products entering the lungs. The mucociliary barrier is the first factor attempting to avoid pathogens to reach deeper tissues. It consists of a layer of epithelial and endothelial cells (Pohl et al., 2009). Respiratory epithelia synthesize mucin glycoproteins and antimicrobials molecules conforming the mucus, which effectively traps microorganisms (Knowles & Boucher, 2002). Soluble immunity at the human airways includes proteases as lysozyme, iron binding-proteins promoting iron nutritional immunity, defensins, i.e. antimicrobial peptides (AMP) released by leukocytes and respiratory epithelial cells, or lipid signaling mediators with detergent effect (Actor et al., 2009; Desbois & Smith, 2010; Lehrer, 2004; Martin & Frevert, 2005). Alveoli are lined with pulmonary surfactant, a lipoprotein-based lubricating film helping to keep them from collapsing (Saladin, 2003). Complement and surfactant proteins can be bactericidal (Wright, 2003), and act as microbial opsonins facilitating phagocytosis by alveolar macrophages (Martin & Frevert, 2005). Alveolar surfactant lipids and proteins also bind Gram-negative bacterial lipopolysaccharide (LPS) and prevent its interaction with LPS-binding protein (LBP) in alveolar fluids (Borron et al., 2000; Sano et al., 2000).

Moreover, alveolar macrophages are main players of lung cellular innate immunity, avidly phagocytic and ingesting all types of inhaled particulates (Geiser, 2010). Neutrophils are also essential for killing of invading pathogens by phagocytosis and release of a myriad of antimicrobial molecules (Pechous, 2017) and by neutrophil extracellular trap (NET) formation (Mircescu et al., 2009). Dendritic cells send their extended dendritic projections into the airway lumen, where they come in contact with antigens (Jahnsen et al., 2006).

Activation of the airways adaptive immunity includes B cell antibody production and T cell-mediated cytotoxicity. Humoral responses are mediated by B lymphocytes through the release of antibodies specific to the infectious agent (Chen & Kolls, 2013). Cell-mediated responses involve the binding of cytotoxic T lymphocytes to infected cells, followed by their lysis. Antigenic variation is a common strategy employed by pathogens to avoid immune detection (Baxt et al., 2013; Young et al., 2002).

Pathogens have developed strategies to avoid or break these immunity barriers, such as targeting cell-cell junctions (Guttman & Finlay, 2009), using mucinases (Linden et al., 2008; Valeri et al., 2015), iron scavenging mechanisms (Caza & Kronstad, 2013; Miethke & Marahiel, 2007; Parrow et al., 2013), or evasion of complement-mediated killing (Lambris et al., 2008). Regarding professional phagocytes, subversion strategies involve to not only to evade phagocytosis, and also favor surviving within phagocytic cells (Ray et al., 2009; Thi et al., 2012).

Besides host immunity, pathogens also need to overcome exogenous selective pressures to successfully colonize and/or cause disease. These include therapeutic interventions such as antimicrobial administration. Bacteria have evolved sophisticated strategies to counteract them (Blair et al., 2015; Munita & Arias, 2016).

As a whole, selective pressures lead to a **co-evolutionary arms race** between host and pathogens, where an “attack-defense” strategy happens, driving a continuous cycle where the host evolves new or more sophisticated defenses to thwart the pathogen's attack and, therefore, the pathogen adapts a more efficient strategy to overcome the heightened defenses. In turn, the host will again develop novel defenses to cope with the new attack mechanism; ultimately, one host evolutionary defensive step implies a pathogen step towards sharpen attack. This continuous cycle entails the spread of beneficial alleles that will be fixed within the bacterial population. Another complementary model favors frequency-dependent **selection of rare alleles** by preserving allelic diversity within a population, as a strategy of conferring distinct advantages to the pathogen in the presence of different host alleles. Evidence exists within nature for both directional and frequency-dependent selection, and both types probably occur in bacterial populations (Bliven & Maurelli, 2016).

1.5. Bacterial patho-adaptive evolution: from experimental to within-host natural approaches

Pathogen evolutionary studies coupling WGS and systematic phenotyping lead to great progress towards obtaining an integrated view of microbial cell function. Microbial genetics, high-throughput technologies and systems biology provide understanding of the structure and function of bacterial genomes. DNA sequencing allows identifying genetic changes between ancestral and derived organisms on a whole-genome scale for any species (Barrick & Lenski, 2013), and integrated evolutionary perspectives allow relating the dynamics of adaptive changes to the phenotypic and genotypic landscapes of living organisms (Hindre et al., 2012). This can be achieved in laboratory or natural

within-host settings (**Figure 3**).

1.5.1. Understanding bacterial patho-adaptation in experimental settings

Experimental evolution relies on simplified models where adaptation traits are easily tracked, *in vivo* with microorganisms and *in silico* with artificial organisms (Hindre et al., 2012). *In viv*o, populations of organisms are maintained in controlled environments where changes in genotype and phenotype are monitored over timescales spanning many generations. Bringing evolution into the laboratory generates a *fossil* record for later study, and allows testing the predictability of evolution across replicate populations. A long-term experiment with *Escherichia coli* by Lenski and co-workers is a paradigm of bacterial experimental evolution (Lenski, 2017). Started in the late 80´s of the 20th century (Lenski et al., 1991), the original *E. coli* populations have passed over 66,000 generations in a minimal medium with glucose, and the experiment continues (Fox & Lenski, 2015). This evolution experiment highlights that bacterial fitness increases and cultures grow faster across generations. Moreover, some bacterial populations evolve hypermutability or the ability to consume specific carbon sources due to genome rearrangements, improving their fitness in the presence of such carbon source (Good et al., 2017).

Experimental settings also allow addressing evolution and spread of antibiotic resistance in bacterial pathogens. Understanding the evolutionary biology of antibiotic resistance informs effective therapeutic strategies and may mitigate the potential to evolve resistance. Laboratory evolution investigates how the rate and genotypic path to antibiotic resistance varies across controlled drug regimens, by bacteria exposure to fixed drug doses that permit only the growth of resistant mutants, identifying single adaptive steps. Recurrent evolutionary patterns, such as the appearance of mutations in a preferred order, provide some level of predictability to a seemingly stochastic evolutionary process. Multi-step experimental evolution in devices for establishing spatial or temporal gradients of drug concentration allow evolving populations to be continuously challenged by effectively increasing the drug dosage to maintain selective pressure as stronger antibiotic resistance evolves, or to mimic the antibiotic dosing regime experienced within a patient, and spatial drug gradients allow bacteria to expand throughout the device as they evolve increasing levels of antibiotic resistance. These methods to genotype and phenotype drug resistant pathogens can identify resistance genes, and the ways that they might mutate to increase antibiotic resistance (Palmer & Kishony, 2013).

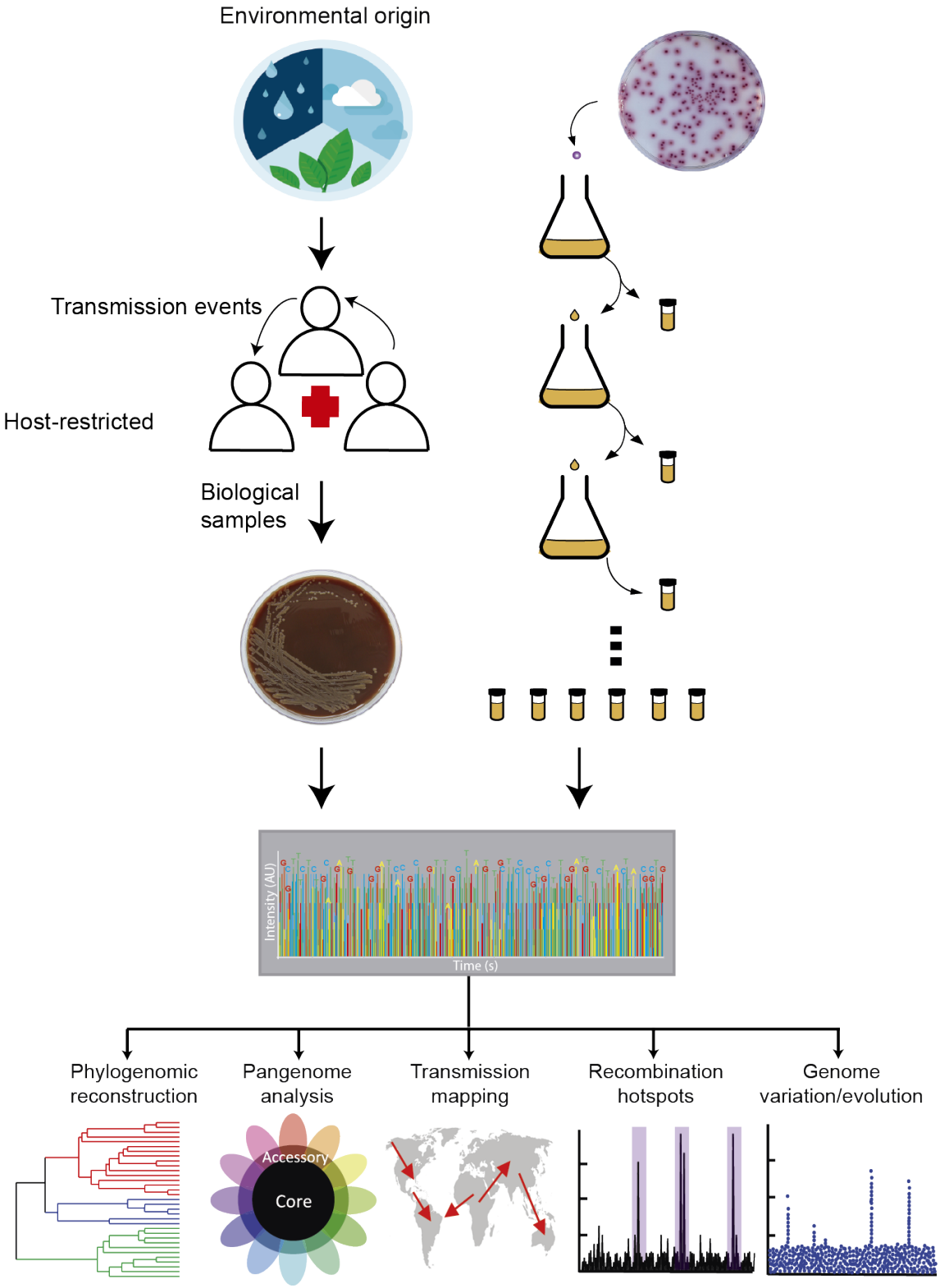


Figure 3. Workflow on WGS-based bacterial adaptation studies. The left flowchart refers to natural or within-host bacterial adaptation studies, where longitudinal microbial collections are isolated from infected patients for further genomic studies. The right flowchart refers to experimental bacterial evolution studies. The bottom panel can be partially common to within-host and experimental adaptation studies, and illustrates several possible outcomes in this type of adaptation studies (adapted from Klemm and Dougan, 2016).

Experimental evolution also allows addressing recombination in action. The dynamics of *Helicobacter pylori* recombination have been tackled by genome sequencing of individual clones and comparative genomics in a short-term transformation assay, with a single *in vitro* transformation cycle with genomic DNA (gDNA) from a donor strain and selection for clones acquiring an antibiotic resistance marker or, alternatively, by repeated transformation cycles by refreshing cultures daily with addition of donor gDNA and platting on non-selective media (Bubendorfer et al., 2016).

As above indicated, **Chapter 2 of this PhD Thesis** work presents a **genetic screening method based on gain-of-function, where bacterial natural transformation generates complex pools of recombinants, phenotypic selection enriches for specific recombinants, and deep sequencing surveys for the genetic variation responsible for the gained phenotypic trait**.

1.5.2. Understanding within-host bacterial patho-adaptation

Despite its advantages, experimental evolution does not inform on how evolutionary mechanisms apply to patho-adaptation in complex natural environments. By contrast, bacterial **within-host natural adaptive evolution** can be tackled at the population (multiple patients (Lieberman et al., 2011; Marvig et al., 2015b)), individual (Lieberman et al., 2014), and organ regional (Chung et al., 2017; Jorth et al., 2015) levels. This type of evolutionary patho-adaptation studies require bacterial clinical isolates longitudinally collected from one or multiple individuals over time, or from different organ regions. Existing within-host adaptation studies are mostly focused on opportunistic pathogens of environmental origin, and are next summarized.

(i) *Pseudomonas aeruginosa* is a major colonizer of the cystic fibrosis (CF) lung whose long-term lower airways infection is linked to the prevalence of dominating and persisting clones (Burns et al., 2001; Johansen & Hoiby, 1992). Within-CF host *P. aeruginosa* adaptive studies encompass wide longitudinally recovered strain collections from multiple patients over years (Bragonzi et al., 2009; Chung et al., 2012; Feliziani et al., 2014; Markussen et al., 2014; Marvig et al., 2013; Marvig et al., 2014; Marvig et al., 2015a), allowing to analyze genomic evolution of single or multiple clonal lineages, identify genes and/or promoters undergoing parallel evolution (Feliziani et al., 2014; Huse et al., 2010; Marvig et al., 2013; Marvig et al., 2014; Sommer et al., 2016; van Mansfeld et al., 2016; Yang et al., 2011), identify genes undergoing convergent evolution (Marvig et al., 2014; Marvig et al., 2015b; Prickett et al., 2017), trace contingency between mutations in signaling pathways (Feliziani et al., 2014; Smith et al., 2006; van Mansfeld et al., 2016),

and analyze regional bacterial diversification associated to environmental heterogeneity (Chung et al., 2012; Feliziani et al., 2014; Jorth et al., 2015; Markussen et al., 2014).

(ii) The *Burkholderia cepacia* complex (Bcc) causes severe respiratory infections in CF individuals (Leitão, 2017). In fact, *B. dolosa* isolates from an epidemic outbreak rendered a longitudinal strain collection recovered from CF patients allowing pioneering genome-based adaptation studies. Identification of recurrent patterns of evolution implicated in pathogenesis by comparing the genetic adaptation of a single strain in multiple human subjects during an epidemic spread revealed for the first time the existence of bacterial genes acquiring dN mutations in multiple individuals, thus indicating parallel adaptive evolution (Lieberman et al., 2011). Moreover, sequencing both individual clones and whole population from single sputum samples revealed extensive intra-sample diversity, suggesting that diversifying lineages coexist for years, supporting a “diverse-community” model where adaptive lineages rise to intermediate frequency and co-exist with other lineages (Lieberman et al., 2014). Strong signals of *B. pseudomallei* and *B. cenocepacia* adaptation including inactivation of virulence and immunogenic factors, and decreased motility and biofilm formation have also been revealed (Lee et al., 2017; Price et al., 2013).

(iii) *H. pylori* is a highly adapted (Ghose et al., 2005; Salama et al., 2013; Salih, 2009) human stomach opportunistic pathogen (Al-Sulami et al., 2010; Atapoor et al., 2014; Blaser, 1997; Perez et al., 2010; Ranjbar et al., 2016a; Ranjbar et al., 2016b; Vinella et al., 2015), able to cause chronic gastritis, stomach and duodenum peptic ulcer, adenocarcinoma, and mucosa associated lymphoid tissue lymphoma (Kusters et al., 2006; Owens & Smith, 2011; Pereira & Medeiros, 2014). Bacterial genome sequencing of longitudinally sampled biopsies revealed polyclonality during adaptation and progression of infection (Cao et al., 2015; Didelot et al., 2013), together with adhesin parallel evolution (Cao et al., 2015; Kennemann et al., 2011), and global and local events of positive selection in DNA repair machinery, outer membrane protein (OMP) encoding genes, metal metabolism and virulence factors (Montano et al., 2015). *H. pylori* has also contributed understanding local inter-patient transmission (Didelot et al., 2013) and global population dynamics (Munoz-Ramirez et al., 2017; Thorell et al., 2017).

(iv) Within-host adaptation and transmission studies highlight *Mycobacterium tuberculosis* genome stability (Merker et al., 2013; Perez-Lago et al., 2014; Sandegren et al., 2011; Saunders et al., 2011). Mutations in genes conferring antibiotic resistance occur during *M. tuberculosis* adaptation (Merker et al., 2013; Saunders et al., 2011;

Senghore et al., 2017; Sun et al., 2012). Co-existence of clonal subpopulations (Black et al., 2015; Lieberman et al., 2016; Merker et al., 2013; Sun et al., 2012), and genotype diversification within patients across human body sites are also features of *M. tuberculosis* persistent infection (Lieberman et al., 2016).

(v) Environmentally found *Stenotrophomonas maltophilia* (Brooke, 2012; Schable et al., 1991) can cause respiratory, urinary, ocular, intra-abdominal and bloodstream infections (Denton & Kerr, 1998). Moreover, *S. maltophilia* CF lung colonization (de Vrankrijker et al., 2010) is associated with increased risk of CF pulmonary exacerbations, lung transplantation and mortality (Waters et al., 2011; Waters et al., 2013). *S. maltophilia* shows intra- and inter-patient genotypic diversity in longitudinally collected strains from CF patients, with appearance of hypermutator phenotypes (Chung et al., 2017; Esposito et al., 2017; Pompilio et al., 2016; Vidigal et al., 2014). *S. maltophilia* has also contributed to further our understanding on pathogen adaptation at the organ level, with positive selection acting in genes related to virulence and antibiotic resistance under local selection (Chung et al., 2017).

(vi) *Legionella pneumophila* is an inhabitant of aquatic and soil environments (Swanson & Hammer, 2000) that causes legionnaire´s disease (McDade et al., 1977). Genome sequencing-based epidemiological studies show that recombination plays a dominant role in the evolution of this pathogen (Coscolla & Gonzalez-Candelas, 2007; Coscolla et al., 2011; Gomez-Valero et al., 2011; McAdam et al., 2014; Sanchez-Buso et al., 2014), among global *L. pneumophila* population and also within an outbreak, where different subpopulations can be found among infected patients (McAdam et al., 2014).

Staphylococcus aureus is, to date, the only member of the human microbiome whose evolution as a colonizing opportunistic pathogen has been studied (Chambers, 2001; Kuehnert et al., 2006; Lowy, 1998; von Eiff et al., 2001; Young et al., 2017). During CF chronic endobronchial infection, *S. aureus* reveals loss of phage content, variation in genes influencing antibiotic resistance and a heterogeneous infecting population evolved from a common ancestor (McAdam et al., 2011). During CF bronchopulmonary colonization, *S. aureus* reveals genome size reduction, involving loss of virulence genes (Liu et al., 2016) and reduced stimulation of host inflammatory responses (Lopez-Collazo et al., 2015). Genome evolutionary dynamics during *S. aureus* transition from carriage to invasive disease report loss-of-function of genes involved in transcriptional regulation (Young et al., 2017), and reduced toxicity by mutations in toxicity-affecting *loci* (Laabei et al., 2015; Young et al., 2012).

Overall, the role of genome evolutionary changes in human host-restricted members of the microbiome causing opportunistic infection has barely been tackled. Moreover, evolutionary studies comprehensively identify genomic changes, but further experimental assessment of the biological significance of such genomic variation is frequently limited, and host selective pressures driving adaptive evolution of those opportunistic pathogens are mostly unknown.

Chapter 3 of this PhD Thesis work describes **patho-adaptive evolution by a human-restricted opportunistic pathogen during chronic lung infection**. Pathogen phylogenomic analysis, identification of bacterial genes under adaptive evolution by tracking recurrent patterns of mutations, and biological significance of such variation are presented.

As a whole, this PhD Thesis work has been focused on three different aspects of **patho-adaptive evolution by the Gram negative bacteria *Haemophilus parasuis* and *Haemophilus influenzae*, two respiratory colonizing opportunistic pathogens,** which are **host-restricted members of the swine and human airway microbiomes**, respectively.

An overview of the mammalian respiratory system and of both bacterial species general features is next presented.

2. The mammalian respiratory system: physiology, patho-physiology and microbiome

The respiratory system allows air to move from the external environment to the inner respiratory surface, where lungs are the primary organs. In mammals, two lungs are located near the backbone on either side of the heart. Each lung, in turn, is divided into lobes. In humans, the right lung is divided into three lobes and the left lung has two. Conversely, swine have four lobes in the right lung and three in the left one. The air enters the nostrils at the upper respiratory tract. Nostrils conduct the air into the hollow nasal passages where it faces the turbinates, which are thin bones responsible for increasing the surface area of these chambers, moistening the air, and warming the incoming air through the high presence of capillaries (Reznik, 1990). The air then passes successively through the pharynx, glottis and larynx and then goes down the trachea. The upper respiratory tract is lined with ciliated cells that secrete mucus, which traps foreign particles such as dust and microorganisms (Reznik, 1990). In the lower

respiratory tract, the trachea branches into two smaller passages called bronchi. One bronchus enters each lung, and each bronchus subsequently branches to produce a network of thinner tubes called bronchioles, also lined with a ciliated mucous membrane (Saladin, 2003). Each bronchiole ends in a grape-like cluster of tiny sacs called alveoli, the place where the exchange of gases occurs (Maina, 2000; Person, 2006). The wall of these alveoli is adjacent to a network of tiny capillaries, where the exchange of oxygen and carbon dioxide takes place. Alveoli are lined with pulmonary surfactant, helping to keep them from collapsing (Saladin, 2003).

2.1. Chronic Obstructive Pulmonary Disease: risk factors and patho-physiology

Chronic Obstructive Pulmonary Disease (COPD) is a common and preventable human respiratory disease characterized by lung function progressive decline and limitation (Buist et al., 2007; Gershon et al., 2011). COPD airway and/or alveolar abnormalities are mostly caused by continuous long-term exposure to noxious particles and gases. COPD patho-physiological features are parenchymal destruction (emphysema), fibrosis, mucociliary dysfunction, mucus hypersecretion, and high local and systemic inflammation (MacNee, 2006). Chronic inflammation causes structural changes, narrowing of the small airways and irreversible destruction of the lung parenchyma, driving the loss of alveolar attachments to the small airways and decreased lung elastic recoil (Barnes, 2016). These changes weaken the ability of the airways to remain open during expiration, ultimately contributing to airflow limitation. COPD affects to more than 5% of the population worldwide. The global burden of this disease was ~250 million cases in 2016, with about 3.17 million deaths in 2015 (WHO, 2017).

2.1.1. Risk factors influencing COPD development and progression

The main risk factor for COPD development is cigarette smoking (Muro, 2011; Silverman & Speizer, 1996), together with other types of tobacco smoking, organic/inorganic dusts, chemical agents and fumes. These noxious particles relate to airflow limitation, emphysema and gas trapping (Vestbo et al., 2013). High levels of urban air pollution are also a factor reducing lung function (Gauderman et al., 2004; Gauderman et al., 2015). Moreover, age is associated to development of COPD, since aging of the airways and parenchyma mimic some structural changes associated to COPD progression (Ito & Barnes, 2009). Current prevalence of COPD is almost equal between genders (Han et al., 2007), but experimental model systems reveal a greater burden of small airway disease in female compared to male mice under similar tobacco smoke exposure conditions (Tam et al., 2016).

About 50% of smokers develop COPD during their lifetime, although there is a controversy about COPD diagnosis since different degree of airway limitation may lead to under-recognition and under-diagnosis. Moreover, COPD is the complex outcome from the interaction between noxious airborne agents and genetic predisposition (Molfino & Coyle, 2008). Several genes have been identified as potential COPD genetic risk factors. A hereditary deficiency of a major inhibitor of serine proteases, α-1 antitrypsin (AATD), and specific SNPs in the genes encoding matrix metalloproteinases 1 and 12 (*MMP1 and MMP12*) relate to early-onset emphysema (Stoller & Aboussouan, 2005) and decline in lung function (Elkington & Cooke, 2010), respectively. Other genes suggested to be linked to COPD are variants in the *family with sequence similarity 13 member A* (*FAM13A*) gene (Cho et al., 2010), and polymorphisms in the gene encoding the Hedgehog-interacting protein (Hip) (Van Durme et al., 2010). Moreover, lung development commences *in utero*, with most marked structural changes occurring during fetal life and soon after birth (Quanjer et al., 2010; Quanjer et al., 2012); thus, lung development is highly susceptible to damage. Maternal smoking during pregnancy, low birth weight, pulmonary sequels in premature infants, postnatal nutrition, postnatal exposure to environmental tobacco smoke and pollution, and childhood respiratory illness are high risk factors for COPD development (Silverman & Speizer, 1996). Other respiratory pathologies are also related to COPD development. Airway hyper-responsiveness, a feature of asthma, is a risk factor for COPD development (de Marco et al., 2011) and thus, asthma is considered as a COPD risk factor itself (Silva et al., 2004). Chronic bronchitis (CB) is also associated with COPD development, exacerbation and severity (Corhay et al., 2013). Besides, tuberculosis and Human Immunodeficiency Virus (HIV) infections accelerate the onset of COPD progression (Byrne et al., 2015; Drummond & Soriano, 2014).

2.1.2. General features of COPD patho-physiology

Noxious agents injure the human airway epithelium and drive key processes into specific airway inflammation and structural changes during COPD development and progression (Ling & van Eeden, 2009). Next to persistent inflammation, increased oxidative stress and a protease:anti-protease imbalance arise during disease (Fischer et al., 2011). COPD airway obstruction is due to changes in lung parenchyma, but also to mucous exudates occluding thickened airways walls (Jeffery, 1999). COPD pathological features are next summarized.

(i) Oxidative stress: Pollutants and noxious particles, mostly cigarette smoke, are a major

external source of oxidative stress, thus oxidative metabolism is over-activated in COPD (Rahman & Adcock, 2006), and increased amounts of hydrogen peroxide and 8-isoprostane in sputum, breath condensate and systemic circulation are found in COPD patient samples (Domej et al., 2014; Lim & Thomas, 2013). Moreover, inflammation heightens internal production of oxidants (Domej et al., 2014). Although the heme-oxygenase and glutathione pathways counteract oxidation, modifications in key regulatory proteins within both pathways lead to an insufficient anti-oxidant activity (Aryal S., 2012). The *nrf2* gene, encoding a master transcriptional regulator of anti-oxidant responses, is under-expressed in lung tissue and alveolar macrophages of COPD patients (Malhotra et al., 2009); opposite, the *bach1* gene, encoding another transcriptional regulator, and *keap1*, a Nrf2 repressor, are over-expressed resulting in an impaired anti-oxidant response (Yamada et al., 2016).

(ii) Protease:anti-protease imbalance: Protease activity is regulated by α-1 antitrypsin, secretory leukoprotease inhibitor and tissue inhibitor of metalloproteinases (TIMPs) (Pandey et al., 2017). α-1 antitrypsin deficiency is a feature of COPD pathology (Eden, 2010; Stoller et al., 1993; Stoller & Aboussouan, 2005). In addition, COPD patients diminish the release of TIMPs by alveolar macrophages (Vignola et al., 2004), facilitating protease-mediated destruction of elastin, which contributes to emphysema and subsequent airflow limitation (Pandey et al., 2017).

(iii) Inflammation: Neutrophil numbers are increased in the COPD airways (Hoenderdos & Condliffe, 2013), which results in a broaden release of oxidants and proteinases, perpetuating the above cited imbalance and favoring lung destruction. Increased numbers of macrophages also contribute to release of inflammatory mediators, including cytokines, chemokines, lipid mediators and growth factors, which attract inflammatory cells from the circulation, amplifying the inflammatory process and inducing structural changes towards airways deleterious remodeling (Barnes, 2016). Likewise, excessive production of growth factors leads to development of repeated airways wall injury, resulting in excessive production of muscle and fibrous tissue (Churg et al., 2006). This fibrotic feature is a contributing factor of small airway limitation, and may eventually precede the development of emphysema (Hogg & Timens, 2009). Airway inflammation is also linked to molecular mediators of disease such as lipids (Petrache & Petrusca, 2013). In this context, arachidonic acid (AA) metabolites play an important role in COPD-related airway inflammation (Malhotra et al., 2012). Thus, phospholipase A_2 (PLA_2) catalyzes the hydrolysis of host cell membrane phospholipids, resulting in the production of free fatty

acids. Among these fatty acids, AA serves as a precursor for inflammatory mediators such as platelet-activating factor (PAF) (Meyer et al., 2005). In turn, lipoprotein-associated PLA_2 mediates PAF hydrolysis and dampens PAF-mediated inflammation (McIntyre et al., 2009).

(iv) Mucus hypersecretion: overproduction and hypersecretion of mucin glycoproteins by globet cells, and decreased elimination of mucus are COPD trademarks (Kim & Criner, 2013; Vestbo et al., 2013). Mucus hypersecretion is a consequence of cigarette smoke exposure (Deshmukh et al., 2005; Ebert & Terracio, 1975), viral infection (Holtzman et al., 2005), bacterial infection (Burgel & Nadel, 2004), or inflammatory cell activation of mucine gene transcription (Burgel & Nadel, 2004). This is also compounded by a difficulty in clearing secretions because of poor ciliary function, distal airway occlusion, and an ineffective cough that is a consequence of respiratory muscle weakness and reduced peak expiratory flow (Burgel & Nadel, 2004; Hogg et al., 2004; Verra et al., 1995).

2.2. The lung microbiome

Mammalian lower airways encompass their own microbiome, including both commensal and opportunistic pathogen microorganisms, whose composition and structure can be modulated by disease or by the exogenous exposome (Shukla et al., 2017). Current knowledge on swine and human lung microbiome is next summarized.

2.2.1. The swine lung microbiome

Despite large economic loses in swine production related to respiratory infections, swine respiratory microbiome and its relationship with respiratory infections are barely known. Metagenomic analysis of healthy swine´s lungs shows *Mycoplasmataceae, Bradyrhizobiaceae* and *Flavobacteriaceae* as common bacteria, whereas in those lungs from pigs with suggestive signals of pneumonia, a less diverse microbial population is found, being *Mycoplasma hyopneumoniae* the most commonly identified bacterium (Siqueira et al., 2017). *M. hyopneumoniae* is the etiologic agent of swine enzootic pneumonia (L'Ecuyer & Boulanger, 1970). During *M. hyopneumoniae* respiratory infection, a displacement within the respiratory microbial population happens, and opportunistic pathogens such as *Pasteurella* spp., *Streptococcus* spp. and ***Haemophilus*** spp. may increase their virulence abilities and cause secondary infections (Siqueira et al., 2017). Likewise, a correlation between gastrointestinal microbiome and development of an effective immune system is related to swine lung pathology outcomes, with decreased disease symptoms and increased weight gain in pigs with increased fecal microbial

diversity, thus suggesting the important role of global microbial diversity as a contributing factor in disease (Niederwerder, 2017).

2.2.2. The human lung microbiome

Until recently, the lower airways of healthy humans were considered to be sterile, opposite to those of respiratory disease patients such as COPD (Erb-Downward et al., 2011). These assumptions relied on culture-based studies where fastidious and non-culturable microorganisms were unable to be detected (Dickson & Huffnagle, 2015). However, molecular studies demonstrated that the lungs are not sterile, and that a wide microbial diversity colonizes this niche. By deep-sequencing of amplicons of the 16S rRNA gene, a core microbiome has been characterized at the human lungs (Einarsson et al., 2016; Erb-Downward et al., 2011) encompassing, at the phylum level, members of Bacteroidetes, Firmicutes, Proteobacteria, Fusobacteria and Actinobacteria. Specifically, *Pseudomonas, Streptococcus, Prevotella, Fusobacterium, Haemophilus, Veillonella* and *Porphyromonas* are the main bacterial species identified within human healthy lungs (Erb-Downward et al., 2011; Pragman et al., 2012).

Differences between healthy and COPD lungs in terms of microbial populations have variable outcomes depending on the sampling methods. The respiratory microbiome has been elucidated on sputum, bronchial aspirates, bronco-alveolar lavage fluid (BALF), bronchial mucosa, protected brush specimens and lung tissue samples (Dy & Sethi, 2016; Erb-Downward et al., 2011; Pragman et al., 2012). Decreased bacterial diversity on COPD compared to healthy lungs (including smokers) is generally observed, together with a shift in community members from Bacteroidetes to Proteobacteria or Firmicutes (Einarsson et al., 2016; Erb-Downward et al., 2011; Sze et al., 2012), suggesting that this disease relates with lower airways microbial changes.

Microbial colonization by opportunistic pathogens contributes to COPD progression (**Figure 4**). The *vicious circle* hypothesis establishes that chronic bacterial colonization of the lower airways perpetuates inflammation and contributes to the progression of smoking-associated diseases (Sethi, 2000; Sethi et al., 2009). Once pathogens gain a foothold in the lower respiratory tract, they persist due to smoking-triggered impairment of the mucociliary clearance (Sethi, 2000; Sethi et al., 2009). Cigarette smoke also upregulates mucus production, impairs epithelial elastic properties, downregulates the levels of IgA and affects the phagocytic activity of professional phagocytes (Barnes, 2004; Marti-Lliteras et al., 2009). These alterations facilitate bacterial colonization, which in turn exacerbates inflammation, produces epithelial damage contributing to host immunity

impairment, further allowing the access of microorganisms to the lower airways in an endless loop named *vicious circle* (Sethi et al., 2009).

COPD acute exacerbations (AECOPD) are episodes of increased respiratory symptoms, particularly dyspnea, cough and sputum (Wedzicha & Donaldson, 2003), associated with increased morbidity and mortality, accelerated decline in lung function, and poor health status. AECOPD relate to bacterial infection, and ~30% of sputum and 50% of bronchial secretion cultures present potential pathogenic bacteria (Fagon et al., 1990; Monso et al., 1995). Though non-capsulated ***Haemophilus influenzae*** is typically a benign commensal of the upper respiratory tract, it is also a common opportunistic pathogen isolated from the lower airways of COPD patients during both exacerbation and clinically stable periods, responsible for ~20-30% of all AECOPD episodes (Bandi et al., 2001; Finney et al., 2014; King & Sharma, 2015; Sethi et al., 2002; Sethi & Murphy, 2008). *H. influenzae* persists within the COPD lung from months to years, contributing to chronic airway inflammation that results in worsening of symptoms and speeds disease progression (Anzueto, 2010; Desai et al., 2014; Finney et al., 2014; Murphy et al., 2004).

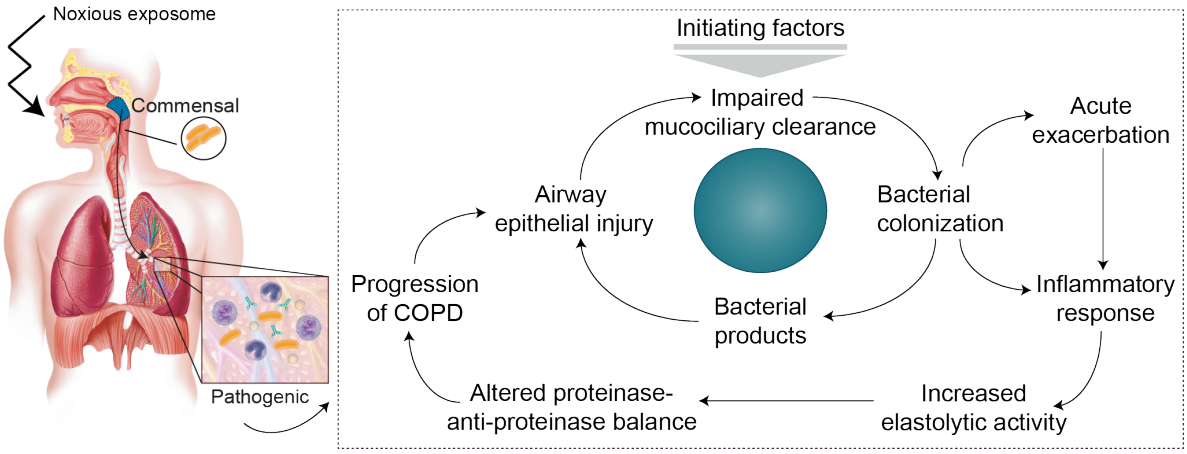


Figure 4. Overview of commensal and opportunistic pathogenic behavior of NTHi within the human respiratory tract. Representation of the *vicious circle* hypothesis contributing to COPD progression (right panel) (adapted from Sethi et al. 2008).

3. *Haemophilus* spp.: general features and pathogenicity

3.1. *Haemophilus parasuis*

Haemophilus parasuis is a Gram negative, non-motile, small pleomorphic bacterium belonging to the *Haemophilus* genus and the *Pasteurellaceae* family (Biberstein & White, 1969). It colonizes de upper airways of healthy pigs, and is commonly isolated from

nasal cavities (Amano et al., 1994), tonsils (Oliveira et al., 2001), and the upper part of the trachea (Segales et al., 1997). Glässer first described it (Glässer, 1910), but it was first isolated by Schermer and Ehrlich (Little, 1970). *H. parasuis* requires nicotinamide adenine dinucleotide (NAD, factor V) for growth. Non-capsulated strains form various structures, from rods to fibers. Capsulated strains are coccobacilli, but also form filaments and fimbriae-like structures (Nedbalcova, 2006).

Besides being a member of the swine respiratory microbiome, *H. parasuis* is a host-restricted colonizing opportunistic pathogen, and infections caused by this pathogen in pigs have become worldwide. *H. parasuis* causes Glässer´s disease, which is characterized by fibrinous polyserositis, polyarthritis and meningitis (Amano et al., 1994), acute pneumonia without polyserositis (Little, 1970), and acute septicemia (Peet et al., 1983).

The consequences of *H. parasuis* infection are important economic loses due to mortality of animals when suffering acute forms of the disease, together with expensive antibiotic treatments (Nedbalcova, 2006). The course of infection by *H. parasuis* is particularly serious in herds with a good health status where outbreaks are accompanied by high morbidity and mortality. Acute infections are occasional, and the clinical disease particularly affects piglets exposed to stress, such as the one occurring at weaning (Correa-Fiz et al., 2016; Quinn, 2011).

In Chapter 1 of this PhD Thesis work, ***H. parasuis*** was employed as a model system for **a host-restricted colonizing opportunistic pathogen, as a means to screen the prevalence of antibiotic resistance in piglets as healthy carrier population, and to identify and characterize genetic determinants accounting for such resistance.**

3.2. *Haemophilus influenzae*

Haemophilus influenzae is a Gram negative facultative anaerobic coccobacillus. *H. influenzae* belongs to the Gamma Proteobacteria class, order Pasteurellales, *Pasteurellaceae* family. It is a member of the human nasopharynx of most healthy humans, from where it can spread to cause both respiratory and systemic infection. Evidence suggests that *H. influenzae* is human host-restricted (Agrawal & Murphy, 2011). It was first identified with the assumption of being the responsible microorganism of influenza pandemic (Pfeiffer, 1892). *H. influenzae* has specific growth factor requirements, hemin (factor X) and NAD. It grows at 37ºC in aerobic conditions with presence of CO_2. Growth on chocolate agar generates small and grey, round, convex colonies, which may be iridescent. Iridescent colonies relate to the presence of a

polysaccharide capsule. Capsulated strains are classified on six types (a-f) (Kirkman & Crawford, 1971) based on their capsular antigens. Serotype b was the most prevalent and invasive serotype (Tudor-Williams et al., 1989) until the introduction of the *H. influenzae* type b (Hib) vaccine (Peltola, 2000). Hib is an important cause of meningitis, epiglotitis and acute pneumonia. Hib vaccine shifted the prevalence of capsular types and the appearance of non-capsulated strains, termed nontypeable *H. influenzae* (NTHi). NTHi strains can cause acute conjunctivitis, otitis media, sinusitis, tracheobronchitis, pneumonia, and are responsible of 20-30% of COPD exacerbations (Finney et al., 2014; King & Sharma, 2015; Sethi & Murphy, 2008).

3.2.1. NTHi genomic heterogeneity

NTHi genomic diversity was initially tackled by genotyping methods including enterobacterial repetitive intergenic consensus (ERIC) typing (Gomez-De-Leon et al., 2000; Pettigrew et al., 2002; Smith-Vaughan et al., 1998), pulse field gel electrophoresis (PFGE) (Pettigrew et al., 2002; Saito et al., 1999), ribotyping (Pettigrew et al., 2002; Wang et al., 2001), and multilocus sequence typing (MLST) (Lacross et al., 2008; Meats et al., 2003). Higher variability among NTHi isolates compared to capsulated strains was reported (Munson et al., 1989; Musser et al., 1988; Smith-Vaughan et al., 1998), strains were clustered pointing to capsule loss as a driver for diversification through recombination (Meats et al., 2003; Smith-Vaughan et al., 1998). A supragenome hybridization technique (SGH) was also developed and employed to characterize *H. influenzae* full genetic content (Eutsey et al., 2013). Moreover, NTHi genetic diversity has been widely tackled through phenotipic analysis (Chin et al., 2005; Maughan & Redfield, 2009; Murphy & Kirkham, 2002) and by assessing the prevalence of genes of interest among NTHi clinical strain collections (Cardines et al., 2012; Gilsdorf et al., 2004).

H. influenzae strain RdKW20 genome was the first free-living organism whose genome was fully sequenced (Fleischmann et al., 1995). The complete nucleotide sequence of its chromosomal DNA is ~1.83 Megabase (Mb), with 38% G+C content and 1,743 predicted coding regions. However, the first report on NTHi comparative genomics came along almost 20 years later. A core and an accessory-genome were defined, polymorphic sites analyzed and evolutionary clades defined, highlighting a great variability among strains, with no correlation between isolation origin and genomic content (De Chiara et al., 2014). Supporting this notion, clear cut correlations between genotypic differences, geographical isolation site and disease origin have not been established, concluding high genomic variability among NTHi strains likely to result from DNA exchanges through

transformation (Power et al., 2009). Genome sequencing, MLST and SNP typing were also jointly used to assess NTHi population diversity (Staples et al., 2017), similarly pointing to a great genomic diversity.

3.2.2. Molecular mechanisms of NTHi genomic variation

Recombination plays an important role in NTHi genomic heterogeneity, as shown by *in vitro* **natural transformation** followed by whole genome sequencing of recombinant variants of this bacterium (Mell et al., 2011). Such heterogeneity relies on NTHi natural competence to be highly transformable. NTHi can actively uptake foreign linear DNA and incorporate it into its chromosome by double homologous recombination, thus providing a potent mechanism of genetic variability, as a means of DNA acquisition and exchange (Goodgal & Mitchell, 1984; Mell et al., 2014; Redfield et al., 2006). DNA uptake by NTHi is 10-100 times more efficient when consensus uptake sequences (uptake signal sequences, USSs) are present in the exogenous DNA (Mell et al., 2012). cAMP receptor protein and the competence-specific activator Sxy activate transcription of competence genes (Redfield et al., 2005), and a set of seventeen genes are needed for DNA uptake and transformation in NTHi, including *pilABCD*, *comNOPQ* and *comABCDEF* operons, and *pilF2*, Hi0659 and *rec2* genes (Sinha et al., 2012).

NTHi variability is also driven by **phase variation**. Switching occurs by slippage of single sequence repeats (SSR) within genes coding for virulence molecules. Phase variation is mediated by short DNA repeats in gene coding or promoter regions. Spontaneous gain or loss of repeats results in translational frame shifts (Gilsdorf et al., 2004; Henderson et al., 1999; Weiser, 2000) and gene expression alteration (Ahmad et al., 2017; Henderson et al., 1999; Power et al., 2009). Phase variation within coding region genes is found in NTHi lipooligosaccharide (LOS)-modifying genes (*lic1A, lic2A* and *lic3A*, 5´-CAAT; *lgtC*, 5´-GACA; *lex2A*, 5´-GCAA) and iron-acquisition genes (*hgpA, hgpB* and *hgpC*, 5´-CCAA). Genes undergoing phase variation due to a variable number of repeats within their promoter region are the pilus gene *hifA* (5´-TA) and adhesin-invasin genes *hmw1A* and *hmw2A* (5´-ATCTTTC) (Cholon et al., 2008; van Ham et al., 1993).

Hypermutation has also been reported in NTHi strains isolated from the respiratory tract of CF patients, and acquisition of antibiotic resistance mutations the main analyzed consequence (Roman et al., 2004; Watson et al., 2004). In this bacterium, hypermutable phenotype refers to an increase in the frequency of recombination and mutation rates that arise into a genome (Jolivet-Gougeon et al., 2011), thus enhancing genomic variation and diversification. The mechanisms underlying such phenotype encompass

polymorphisms in conserved regions of the *mutS* mismatch repair gene (Watson et al., 2004).

In Chapter 3 of this PhD Thesis work, ***H. influenzae*** was employed as a model system for **a human host-restricted colonizing opportunistic pathogen, as a means to elucidate within-host bacterial patho-adaptive evolutionary traits and their biological significance during chronicity.**

3.2.3. Persistent respiratory infection by *Haemophilus influenzae*

NTHi colonizes asymptomatically the nasopharynx of healthy individuals as part of their respiratory microbiome (Kilian, 1991). In patients suffering underlying respiratory pathologies such as COPD, NTHi reaches the lower airways and persists (King, 2012). NTHi encounters a whole range of selective pressures within the various microenvironments of the human host, from asymptomatic nasopharyngeal carriage to disease. Airways epithelial attachment, invasion and intracellular location (as being a facultative intracellular pathogen, **Figure 5**), together with its ability to form biofilms (while extracellular), are NTHi strategies contributing to persistent infection in otitis media, CF or COPD patients (Arce et al., 2009; Clementi et al., 2014; Lopez-Gomez et al., 2012; Morey et al., 2011).

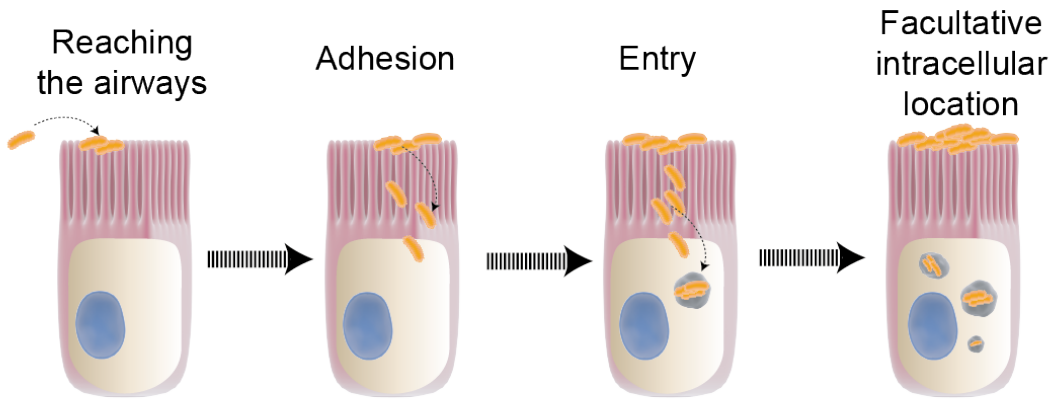


Figure 5. Summary of NTHi stages of colonization at the airways epithelia. NTHi reaches the respiratory airways epithelial cells, it attaches to host cell surface receptors through bacterial ligands, invades the cells and can persist intracellularly.

NTHi forms biofilms in experimental otitis media, in the ear of otitis media, and in the lower airways of CF patients (Hall-Stoodley et al., 2006; Idicula et al., 2016; Jurcisek & Bakaletz, 2007; Post, 2001; Starner et al., 2006), and a role for iron availability in NTHi biofilm architecture has been suggested (Vogel et al., 2012). NTHi biofilm extracellular matrix contains polysaccharides, proteins and nucleic acids (Domenech et al., 2016;

Jurcisek et al., 2017; Wu et al., 2014), and its protein composition has been analyzed (Gallaher et al., 2006; Post et al., 2014; Webster et al., 2006). Current understanding of NTHi biofilm formation mechanisms is mostly focused on otitis media.

NTHi attachment to host factors, including extracellular matrix proteins and host cell surface receptors, is mediated by a panel of bacterial surface-exposed adhesins or ligands allowing NTHi to colonize primary sites and establish infection (Duell et al., 2016; Swords et al., 2000). NTHi proteinaceous adhesins and lipooligosacharide (LOS) moieties contribute to both bacterial attachment to host factors (summarized in **Table 1**) and also overcoming soluble immunity elements such as complement system and AMPs (Hallstrom et al., 2010; Starner et al., 2002), together representing important steps in colonization and infection. Thus, the ability to adhere and adapt to the human respiratory tract mucosa plays a pivotal role in NTHi pathogenic lifestyle. Temporal profiling of host and pathogen mRNA signatures associated with a successful colonization of ciliated human bronchial epithelium shows significant dysregulation of the target cell cytoskeleton elicited by bacterial infection, with a profound effect on the intermediate filament network and junctional complexes; in turn, NTHi downregulates its central metabolism and increases expression of transporters. The oxidative environment generated by infected cells instigates bacterial expression of stress-induced defense mechanisms (Baddal et al., 2015).

As a facultative intracellular pathogen, NTHi enters numerous epithelial cell types (Chang conjunctiva, RPMI 2650 nasal, Detroit-562 pharyngeal, NCI-H292 bronchial, A549 type II pneumocytes) (Euba et al., 2015b; Jiang et al., 1999; Prasadarao et al., 1999; van Schilfgaarde et al., 2000). Host cell subversion upon epithelial cell invasion has been elucidated in A549 cells (Euba et al., 2015a; Lopez-Gomez et al., 2012). Inside epithelial cells, NTHi resides in a late endosome-like compartment in a non-replicative fashion (Morey et al., 2011). Host cells can offer protection against immune clearance mechanisms such as bactericidal antibodies and antibiotics (Clementi & Murphy, 2011; van Schilfgaarde et al., 1999). Expression of immunoglobulin A (IgA) proteases promotes NTHi epithelial invasion and persistence (Clementi et al., 2014; Murphy et al., 2017). Although four different IgA1 proteases have been described for NTHi (Fernaays et al., 2006), a predominant role in persistence is assigned to IgAs B1 and B2 (Murphy et al., 2017). In contrast, NTHi interaction with alveolar macrophages is unlikely to contribute to chronicity due to efficient bacterial phagolysosomal clearance (Marti-Lliteras et al., 2009). To persist within the human airways, NTHi also mitigates toxic levels of oxidative stress. Phase variation of the *modA2* methyltransferase modulates NTHi sensitivity to oxidative

stress and resistance to neutrophil-mediated killing (Brockman et al., 2017). Moreover, NTHi has been found to survive within NETs (Hamaguchi et al., 2012b; Hong et al., 2009; Juneau et al., 2011; King L.B., 2013), being resistant to both extracellular killing within NETs and phagocytic killing by incoming neutrophils (Juneau et al., 2011), which has been proposed to serve as a persistence mechanism (Hamaguchi et al., 2012a; Hong et al., 2009; Juneau et al., 2011).

In Chapter 2 of this PhD Thesis work, ***H. influenzae*** was employed as a model system for **a naturally transformable bacterial pathogen, as a means to develop a novel genetic screening method based on gain-of-function named *transformed recombinant enrichment profiling* (TREP)**. In particular, this methodology was applied to the **search of invasin encoding genes of NTHi in the human respiratory tract**, as a trait likely to be implicated in persistence.

Table 1. List of NTHi adhesin-invasin encoding genes mediating bacterial-host cell interplay.

Gene	Function	Experimental model tested	Phase Variation	Reference
hmw1A	Adhesion and invasion of epithelia	Cell infection	Within promoter region	St.Geme 1994; Mell *et al.* 2016
hmw2A	Adhesion and invasion of epithelia	Cell infection	Within promoter region	St.Geme 1993; Mell *et al.* 2016
ompP2	Adhesion to mucin and laminin	Cell infection	No	Orihuela *et al.* 2009
ompP5	Adhesion to mucin, ICAM-1 and hCEACAM1	Cell infection	No	Hill *et al*. 2001, Euba *et al.* 2015
oapA	Adhesion to Chang epithelial cells	Cell infection	No	Prasadarao *et al*. 1999
pe	Adhesion to laminin and vitronectin	Binding assay	No	Hallström *et al.* 2011
hap	Adherence to fibronectin, laminin and collagen IV	Binding assay	No	Fink *et al.* 2002, Kenjale let al. 2009, Euba *et al.* 2015
hia	Adherence to respiratory epithelia	Cell infection	Within promoter region	Laarman *et al.* 2002
hifABCDE	Adhesion to mucus and to Anton antigen (An-Wj)	Cell infection	Within promoter region	Gilsdorf *et al.* 1997
pilABCD	Adhesion to ICAM-1	Chinchilla infection, cell infection	No	Jurcisek *et al.* 2007*;* Novotny and Bakaletz 2016
hsf	Adhesion to vitronectin	Binding assay	No	Hallström *et al.* 2006
lic1ABCD	Adhesion to PAF-R	Cell infection	Within coding sequence	Gould & Weiser 2002
hpd	Facilitated entry into mononuclear cells	Cell infection	No	Ahren *et al.* 2001
fadL	Adhesion to hCEACAM1	Cell infection, binding assay	No	Tchoupa *et al.* 2015

References

Actor, J. K., Hwang, S. A. & Kruzel, M. L. (2009). Lactoferrin as a natural immune modulator. *Curr Pharm Des*, 15 (17): 1956-73.

Adhikari, S. & Curtis, P. D. (2016). DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiol Rev*, 40 (5): 575-91. doi: 10.1093/femsre/fuw023.

Agrawal, A. & Murphy, T. F. (2011). *Haemophilus influenzae* infections in the *H. influenzae* type b conjugate vaccine era. *J Clin Microbiol*, 49 (11): 3728-32. doi: 10.1128/JCM.05476-11.

Ahmad, S., Ahmad, M., Khan, S., Ahmad, F., Nawaz, S. & Khan, F. U. (2017). An overview on phase variation, mechanisms and roles in bacterial adaptation. *J Pak Med Assoc*, 67 (2): 285-291.

Al-Sulami, A. A., Al-Taee, A. M. & Juma'a, M. G. (2010). Isolation and identification of Helicobacter pylori from drinking water in Basra governorate, Iraq. *East Mediterr Health J*, 16 (9): 920-5.

Amano, H., Shibata, M., Kajio, N. & Morozumi, T. (1994). Pathologic observations of pigs intranasally inoculated with serovar 1, 4 and 5 of *Haemophilus parasuis* using immunoperoxidase method. *J Vet Med Sci*, 56 (4): 639-44.

Anzueto, A. (2010). Impact of exacerbations on COPD. *Eur Respir Rev*, 19 (116): 113-8. doi: 10.1183/09059180.00002610.

Arber, W. (2014). Horizontal gene transfer among bacteria and its role in biological evolution. *Life (Basel)*, 4 (2): 217-24. doi: 10.3390/life4020217.

Arce, F. T., Carlson, R., Monds, J., Veeh, R., Hu, F. Z., Stewart, P. S., Lal, R., Ehrlich, G. D. & Avci, R. (2009). Nanoscale structural and mechanical properties of nontypeable *Haemophilus influenzae* biofilms. *J Bacteriol*, 191 (8): 2512-20. doi: 10.1128/JB.01596-08.

Aryal S., D.-G. E., Mannino D.M. (2012). Epidemiology of comorbidities in chronic obstructive pulmonary disease: clusters, phenotypes and outcomes. *Italian Journal of Medicine*, 6: 276-284.

Atapoor, S., Safarpoor Dehkordi, F. & Rahimi, E. (2014). Detection of *Helicobacter pylori* in various types of vegetables and salads. *Jundishapur J Microbiol*, 7 (5): e10013. doi: 10.5812/jjm.10013.

Baddal, B., Muzzi, A., Censini, S., Calogero, R. A., Torricelli, G., Guidotti, S., Taddei, A. R., Covacci, A., Pizza, M., Rappuoli, R., et al. (2015). Dual RNA-seq of nontypeable *Haemophilus influenzae* and host cell transcriptomes reveals novel insights into host-pathogen cross talk. *MBio*, 6 (6): e01765-15. doi: 10.1128/mBio.01765-15.

Baltrus, D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends Ecol Evol*, 28 (8): 489-95. doi: 10.1016/j.tree.2013.04.002.

Bandi, V., Apicella, M. A., Mason, E., Murphy, T. F., Siddiqi, A., Atmar, R. L. & Greenberg, S. B. (2001). Nontypeable *Haemophilus influenzae* in the lower respiratory tract of patients with chronic bronchitis. *Am J Respir Crit Care Med*, 164 (11): 2114-9. doi: 10.1164/ajrccm.164.11.2104093.

Bari, M. R., Hiron, M. M., Zaman, S. M., Rahman, M. M. & Ganguly, K. C. (2010). Microbes responsible for acute exacerbation of COPD. *Mymensingh Med J*, 19 (4): 576-85.

Barnes, P. J. (2004). Mediators of chronic obstructive pulmonary disease. *Pharmacol Rev*, 56 (4): 515-48. doi: 10.1124/pr.56.4.2.

Barnes, P. J. (2016). Inflammatory mechanisms in patients with chronic obstructive pulmonary disease. *J Allergy Clin Immunol*, 138 (1): 16-27. doi: 10.1016/j.jaci.2016.05.011.

Barrick, J. E. & Lenski, R. E. (2013). Genome dynamics during experimental evolution. *Nat Rev Genet*, 14 (12): 827-39. doi: 10.1038/nrg3564.

Baxt, L. A., Garza-Mayers, A. C. & Goldberg, M. B. (2013). Bacterial subversion of host innate immune pathways. *Science*, 340 (6133): 697-701. doi: 10.1126/science.1235771.

Bayliss, C. D., Field, D. & Moxon, E. R. (2001). The simple sequence contingency loci of *Haemophilus influenzae* and *Neisseria meningitidis*. *J Clin Invest*, 107 (6): 657-62. doi: 10.1172/JCI12557.

Bayliss, C. D., Bidmos, F. A., Anjum, A., Manchev, V. T., Richards, R. L., Grossier, J. P., Wooldridge, K. G., Ketley, J. M., Barrow, P. A., Jones, M. A., et al. (2012). Phase variable genes of *Campylobacter jejuni* exhibit high mutation rates and specific mutational patterns but mutability is not the major

determinant of population structure during host colonization. *Nucleic Acids Res*, 40 (13): 5876-89. doi: 10.1093/nar/gks246.

Beasley, V., Joshi, P. V., Singanayagam, A., Molyneaux, P. L., Johnston, S. L. & Mallia, P. (2012). Lung microbiology and exacerbations in COPD. *Int J Chron Obstruct Pulmon Dis*, 7: 555-69. doi: 10.2147/COPD.S28286.

Bennett, P. M. (2008). Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *Br J Pharmacol*, 153 Suppl 1: S347-57. doi: 10.1038/sj.bjp.0707607.

Berg, G., Eberl, L. & Hartmann, A. (2005). The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. *Environ Microbiol*, 7 (11): 1673-85. doi: 10.1111/j.1462-2920.2005.00891.x.

Bergman, M., Del Prete, G., van Kooyk, Y. & Appelmelk, B. (2006). *Helicobacter pylori* phase variation, immune modulation and gastric autoimmunity. *Nat Rev Microbiol*, 4 (2): 151-9. doi: 10.1038/nrmicro1344.

Berrens, Z. J., Marrs, C. F., Pettigrew, M. M., Sandstedt, S. A., Patel, M. & Gilsdorf, J. R. (2007). Genetic diversity of paired middle-ear and pharyngeal nontypeable *Haemophilus influenzae* isolates from children with acute otitis media. *J Clin Microbiol*, 45 (11): 3764-7. doi: 10.1128/JCM.00964-07.

Biberstein & White. (1969). A proposal for the establishment of two new *Haemophilus* species. *J. Med. Microbiol.*, 2.

Black, P. A., de Vos, M., Louw, G. E., van der Merwe, R. G., Dippenaar, A., Streicher, E. M., Abdallah, A. M., Sampson, S. L., Victor, T. C., Dolby, T., et al. (2015). Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in *Mycobacterium tuberculosis* isolates. *BMC Genomics*, 16: 857. doi: 10.1186/s12864-015-2067-2.

Blair, J. M., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. (2015). Molecular mechanisms of antibiotic resistance. *Nat Rev Microbiol*, 13 (1): 42-51. doi: 10.1038/nrmicro3380.

Blaser, M. J. (1997). The versatility of *Helicobacter pylori* in the adaptation to the human stomach. *J Physiol Pharmacol*, 48 (3): 307-14.

Bliven, K. A. & Maurelli, A. T. (2016). Evolution of bacterial pathogens within the human host. *Microbiol Spectr*, 4 (1). doi: 10.1128/microbiolspec.VMBF-0017-2015.

Borron, P., McIntosh, J. C., Korfhagen, T. R., Whitsett, J. A., Taylor, J. & Wright, J. R. (2000). Surfactant-associated protein A inhibits LPS-induced cytokine and nitric oxide production *in vivo*. *Am J Physiol Lung Cell Mol Physiol*, 278 (4): L840-7. doi: 10.1152/ajplung.2000.278.4.L840.

Boussau, B., Karlberg, E. O., Frank, A. C., Legault, B. A. & Andersson, S. G. (2004). Computational inference of scenarios for alpha-proteobacterial genome evolution. *Proc Natl Acad Sci U S A*, 101 (26): 9722-7. doi: 10.1073/pnas.0400975101.

Bragonzi, A., Paroni, M., Nonis, A., Cramer, N., Montanari, S., Rejman, J., Di Serio, C., Doring, G. & Tummler, B. (2009). *Pseudomonas aeruginosa* microevolution during cystic fibrosis lung infection establishes clones with adapted virulence. *Am J Respir Crit Care Med*, 180 (2): 138-45. doi: 10.1164/rccm.200812-1943OC.

Brockman, K. L., Branstool, M. T., Atack, J. M., Robledo-Avila, F., Partida-Sanchez, S., Jennings, M. P. & Bakaletz, L. O. (2017). The *ModA2* phasevarion of nontypeable *Haemophilus influenzae* regulates resistance to oxidative stress and killing by human neutrophils. *Sci Rep*, 7 (1): 3161. doi: 10.1038/s41598-017-03552-9.

Brooke, J. S. (2012). *Stenotrophomonas maltophilia*: an emerging global opportunistic pathogen. *Clin Microbiol Rev*, 25 (1): 2-41. doi: 10.1128/CMR.00019-11.

Bubendorfer, S., Krebes, J., Yang, I., Hage, E., Schulz, T. F., Bahlawane, C., Didelot, X. & Suerbaum, S. (2016). Genome-wide analysis of chromosomal import patterns after natural transformation of *Helicobacter pylori*. *Nat Commun*, 7: 11995. doi: 10.1038/ncomms11995.

Buist, A. S., McBurnie, M. A., Vollmer, W. M., Gillespie, S., Burney, P., Mannino, D. M., Menezes, A. M., Sullivan, S. D., Lee, T. A., Weiss, K. B., et al. (2007). International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. *Lancet*, 370 (9589): 741-50. doi: 10.1016/S0140-6736(07)61377-4.

Burgel, P. R. & Nadel, J. A. (2004). Roles of epidermal growth factor receptor activation in epithelial

cell repair and mucin production in airway epithelium. *Thorax*, 59 (11): 992-6. doi: 10.1136/thx.2003.018879.

Burns, J. L., Gibson, R. L., McNamara, S., Yim, D., Emerson, J., Rosenfeld, M., Hiatt, P., McCoy, K., Castile, R., Smith, A. L., et al. (2001). Longitudinal assessment of *Pseudomonas aeruginosa* in young children with cystic fibrosis. *J Infect Dis*, 183 (3): 444-52. doi: 10.1086/318075.

Byrne, A. L., Marais, B. J., Mitnick, C. D., Lecca, L. & Marks, G. B. (2015). Risk factors for and origins of COPD. *Lancet*, 385 (9979): 1723-1724. doi: 10.1016/S0140-6736(15)60884-4.

Cao, Q., Didelot, X., Wu, Z., Li, Z., He, L., Li, Y., Ni, M., You, Y., Lin, X., Li, Z., et al. (2015). Progressive genomic convergence of two *Helicobacter pylori* strains during mixed infection of a patient with chronic gastritis. *Gut*, 64 (4): 554-61. doi: 10.1136/gutjnl-2014-307345.

Cardines, R., Giufre, M., Pompilio, A., Fiscarelli, E., Ricciotti, G., Di Bonaventura, G. & Cerquetti, M. (2012). *Haemophilus influenzae* in children with cystic fibrosis: antimicrobial susceptibility, molecular epidemiology, distribution of adhesins and biofilm formation. *Int J Med Microbiol*, 302 (1): 45-52. doi: 10.1016/j.ijmm.2011.08.003.

Caza, M. & Kronstad, J. W. (2013). Shared and distinct mechanisms of iron acquisition by bacterial and fungal pathogens of humans. *Front Cell Infect Microbiol*, 3: 80. doi: 10.3389/fcimb.2013.00080.

Chambers, H. F. (2001). The changing epidemiology of *Staphylococcus aureus*? *Emerg Infect Dis*, 7 (2): 178-82. doi: 10.3201/eid0702.700178.

Chattopadhyay, M. K. (2014). Use of antibiotics as feed additives: a burning question. *Frontiers in Microbiology*, 5: 334. doi: 10.3389/fmicb.2014.00334.

Chen, I. & Dubnau, D. (2004). DNA uptake during bacterial transformation. *Nat Rev Microbiol*, 2 (3): 241-9. doi: 10.1038/nrmicro844.

Chen, K. & Kolls, J. K. (2013). T cell-mediated host immune defenses in the lung. *Annu Rev Immunol*, 31: 605-33. doi: 10.1146/annurev-immunol-032712-100019.

Chin, C. L., Manzel, L. J., Lehman, E. E., Humlicek, A. L., Shi, L., Starner, T. D., Denning, G. M., Murphy, T. F., Sethi, S. & Look, D. C. (2005). *Haemophilus influenzae* from patients with chronic obstructive pulmonary disease exacerbation induce more inflammation than colonizers. *Am J Respir Crit Care Med*, 172 (1): 85-91. doi: 10.1164/rccm.200412-1687OC.

Cho, M. H., Boutaoui, N., Klanderman, B. J., Sylvia, J. S., Ziniti, J. P., Hersh, C. P., DeMeo, D. L., Hunninghake, G. M., Litonjua, A. A., Sparrow, D., et al. (2010). Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet*, 42 (3): 200-2. doi: 10.1038/ng.535.

Cholon, D. M., Cutter, D., Richardson, S. K., Sethi, S., Murphy, T. F., Look, D. C. & St Geme, J. W., 3rd. (2008). Serial isolates of persistent *Haemophilus influenzae* in patients with chronic obstructive pulmonary disease express diminishing quantities of the HMW1 and HMW2 adhesins. *Infect Immun*, 76 (10): 4463-8. doi: 10.1128/IAI.00499-08.

Chung, H., Lieberman, T. D., Vargas, S. O., Flett, K. B., McAdam, A. J., Priebe, G. P. & Kishony, R. (2017). Global and local selection acting on the pathogen *Stenotrophomonas maltophilia* in the human lung. *Nat Commun*, 8: 14078. doi: 10.1038/ncomms14078.

Chung, J. C., Becq, J., Fraser, L., Schulz-Trieglaff, O., Bond, N. J., Foweraker, J., Bruce, K. D., Smith, G. P. & Welch, M. (2012). Genomic variation among contemporary *Pseudomonas aeruginosa* isolates from chronically infected cystic fibrosis patients. *J Bacteriol*, 194 (18): 4857-66. doi: 10.1128/JB.01050-12.

Churg, A., Wang, R. D. & Wright, J. L. (2006). Cigarette smoke causes small airway remodeling by direct growth factor induction and release. *Proc Am Thorac Soc*, 3 (6): 493. doi: 10.1513/pats.200603-066MS.

Clementi, C. F. & Murphy, T. F. (2011). Non-typeable *Haemophilus influenzae* invasion and persistence in the human respiratory tract. *Front Cell Infect Microbiol*, 1: 1. doi: 10.3389/fcimb.2011.00001.

Clementi, C. F., Hakansson, A. P. & Murphy, T. F. (2014). Internalization and trafficking of nontypeable *Haemophilus influenzae* in human respiratory epithelial cells and roles of IgA1 proteases for optimal invasion and persistence. *Infect Immun*, 82 (1): 433-44. doi: 10.1128/IAI.00864-13.

Corhay, J. L., Vincken, W., Schlesser, M., Bossuyt, P. & Imschoot, J. (2013). Chronic bronchitis in COPD patients is associated with increased risk of exacerbations: a cross-sectional multicentre study. *Int J Clin Pract*, 67 (12): 1294-301. doi: 10.1111/ijcp.12248.

Correa-Fiz, F., Fraile, L. & Aragon, V. (2016). Piglet nasal microbiota at weaning may influence the development of Glasser's disease during the rearing period. *BMC Genomics*, 17: 404. doi: 10.1186/s12864-016-2700-8.

Coscolla, M. & Gonzalez-Candelas, F. (2007). Population structure and recombination in environmental isolates of *Legionella pneumophila*. *Environ Microbiol*, 9 (3): 643-56. doi: 10.1111/j.1462-2920.2006.01184.x.

Coscolla, M., Comas, I. & Gonzalez-Candelas, F. (2011). Quantifying nonvertical inheritance in the evolution of *Legionella pneumophila*. *Mol Biol Evol*, 28 (2): 985-1001. doi: 10.1093/molbev/msq278.

Dawid, S., Barenkamp, S. J. & St. Geme, J. W. (1999). Variation in expression of the *Haemophilus influenzae* HMW adhesins: A prokaryotic system reminiscent of eukaryotes. *Proceedings of the National Academy of Sciences*, 96 (3): 1077-1082. doi: 10.1073/pnas.96.3.1077.

De Chiara, M., Hood, D., Muzzi, A., Pickard, D. J., Perkins, T., Pizza, M., Dougan, G., Rappuoli, R., Moxon, E. R., Soriani, M., et al. (2014). Genome sequencing of disease and carriage isolates of nontypeable *Haemophilus influenzae* identifies discrete population structure. *Proc Natl Acad Sci U S A*, 111 (14): 5439-44. doi: 10.1073/pnas.1403353111.

de Groot, R., Sluijter, M., de Bruyn, A., Campos, J., Goessens, W. H., Smith, A. L. & Hermans, P. W. (1996). Genetic characterization of trimethoprim resistance in *Haemophilus influenzae*. *Antimicrobial Agents and Chemotherapy*, 40 (9): 2131-2136.

de Marco, R., Accordini, S., Marcon, A., Cerveri, I., Anto, J. M., Gislason, T., Heinrich, J., Janson, C., Jarvis, D., Kuenzli, N., et al. (2011). Risk factors for chronic obstructive pulmonary disease in a European cohort of young adults. *Am J Respir Crit Care Med*, 183 (7): 891-7. doi: 10.1164/rccm.201007-1125OC.

de Vrankrijker, A. M., Wolfs, T. F. & van der Ent, C. K. (2010). Challenging and emerging pathogens in cystic fibrosis. *Paediatr Respir Rev*, 11 (4): 246-54. doi: 10.1016/j.prrv.2010.07.003.

Denton, M. & Kerr, K. G. (1998). Microbiological and clinical aspects of infection associated with *Stenotrophomonas maltophilia*. *Clin Microbiol Rev*, 11 (1): 57-80.

Desai, H., Eschberger, K., Wrona, C., Grove, L., Agrawal, A., Grant, B., Yin, J., Parameswaran, G. I., Murphy, T. & Sethi, S. (2014). Bacterial colonization increases daily symptoms in patients with chronic obstructive pulmonary disease. *Ann Am Thorac Soc*, 11 (3): 303-9. doi: 10.1513/AnnalsATS.201310-350OC.

Desbois, A. P. & Smith, V. J. (2010). Antibacterial free fatty acids: activities, mechanisms of action and biotechnological potential. *Appl Microbiol Biotechnol*, 85 (6): 1629-42. doi: 10.1007/s00253-009-2355-3.

Deshmukh, H. S., Case, L. M., Wesselkamper, S. C., Borchers, M. T., Martin, L. D., Shertzer, H. G., Nadel, J. A. & Leikauf, G. D. (2005). Metalloproteinases mediate *mucin 5AC* expression by epidermal growth factor receptor activation. *Am J Respir Crit Care Med*, 171 (4): 305-14. doi: 10.1164/rccm.200408-1003OC.

Dickson, R. P. & Huffnagle, G. B. (2015). The lung microbiome: new principles for respiratory bacteriology in health and disease. *PLoS Pathog*, 11 (7): e1004923. doi: 10.1371/journal.ppat.1004923.

Didelot, X., Nell, S., Yang, I., Woltemate, S., van der Merwe, S. & Suerbaum, S. (2013). Genomic evolution and transmission of *Helicobacter pylori* in two South African families. *Proc Natl Acad Sci U S A*, 110 (34): 13880-5. doi: 10.1073/pnas.1304681110.

Didelot, X., Walker, A. S., Peto, T. E., Crook, D. W. & Wilson, D. J. (2016). Within-host evolution of bacterial pathogens. *Nat Rev Microbiol*, 14 (3): 150-62. doi: 10.1038/nrmicro.2015.13.

Domej, W., Oettl, K. & Renner, W. (2014). Oxidative stress and free radicals in COPD--implications and relevance for treatment. *Int J Chron Obstruct Pulmon Dis*, 9: 1207-24. doi: 10.2147/COPD.S51226.

Domenech, M., Pedrero-Vega, E., Prieto, A. & Garcia, E. (2016). Evidence of the presence of nucleic

acids and β-glucan in the matrix of non-typeable *Haemophilus influenzae in vitro* biofilms. *Sci Rep*, 6: 36424. doi: 10.1038/srep36424.

Drummond, M. & Soriano, J. B. (2014). An integrated therapeutic approach: do COPD comorbidities justify it? *Rev Port Pneumol*, 20 (1): 3-4. doi: 10.1016/j.rppneu.2013.12.001.

Duell, B. L., Su, Y. C. & Riesbeck, K. (2016). Host-pathogen interactions of nontypeable *Haemophilus influenzae*: from commensal to pathogen. *FEBS Lett*, 590 (21): 3840-3853. doi: 10.1002/1873-3468.12351.

Dy, R. & Sethi, S. (2016). The lung microbiome and exacerbations of COPD. *Curr Opin Pulm Med*, 22 (3): 196-202. doi: 10.1097/MCP.0000000000000268.

Ebert, R. V. & Terracio, M. J. (1975). The bronchiolar epithelium in cigarette smokers. Observations with the scanning electron microscope. *Am Rev Respir Dis*, 111 (1): 4-11. doi: 10.1164/arrd.1975.111.1.4.

Eden, E. (2010). Asthma and COPD in α-1 antitrypsin deficiency. Evidence for the Dutch hypothesis. *COPD*, 7 (5): 366-74. doi: 10.3109/15412555.2010.510159.

Einarsson, G. G., Comer, D. M., McIlreavey, L., Parkhill, J., Ennis, M., Tunney, M. M. & Elborn, J. S. (2016). Community dynamics and the lower airway microbiota in stable chronic obstructive pulmonary disease, smokers and healthy non-smokers. *Thorax*, 71 (9): 795-803. doi: 10.1136/thoraxjnl-2015-207235.

Elkington, P. T. & Cooke, G. S. (2010). MMP12, lung function, and COPD in high-risk populations. *N Engl J Med*, 362 (13): 1241; author reply 1242. doi: 10.1056/NEJMc1000959.

Enne, V. I., King, A., Livermore, D. M. & Hall, L. M. (2002). Sulfonamide resistance in *Haemophilus influenzae* mediated by acquisition of *sul2* or a short insertion in chromosomal *folP*. *Antimicrob Agents Chemother*, 46 (6): 1934-9.

Erb-Downward, J. R., Thompson, D. L., Han, M. K., Freeman, C. M., McCloskey, L., Schmidt, L. A., Young, V. B., Toews, G. B., Curtis, J. L., Sundaram, B., et al. (2011). Analysis of the lung microbiome in the "healthy" smoker and in COPD. *PLoS One*, 6 (2): e16384. doi: 10.1371/journal.pone.0016384.

Erkan, L., Uzun, O., Findik, S., Katar, D., Sanic, A. & Atici, A. G. (2008). Role of bacteria in acute exacerbations of chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis*, 3 (3): 463-7.

Esposito, A., Pompilio, A., Bettua, C., Crocetta, V., Giacobazzi, E., Fiscarelli, E., Jousson, O. & Di Bonaventura, G. (2017). Evolution of *Stenotrophomonas maltophilia* in cystic fibrosis lung over chronic infection: A genomic and phenotypic population study. *Front Microbiol*, 8: 1590. doi: 10.3389/fmicb.2017.01590.

Euba, B., Moleres, J., Segura, V., Viadas, C., Morey, P., Moranta, D., Leiva, J., de-Torres, J. P., Bengoechea, J. A. & Garmendia, J. (2015a). Genome expression profiling-based identification and administration efficacy of host-directed antimicrobial drugs against respiratory infection by nontypeable *Haemophilus influenzae. Antimicrob Agents Chemother*, 59 (12): 7581-92. doi: 10.1128/AAC.01278-15.

Euba, B., Moleres, J., Viadas, C., Ruiz de los Mozos, I., Valle, J., Bengoechea, J. A. & Garmendia, J. (2015b). Relative contribution of P5 and Hap surface proteins to nontypable *Haemophilus influenzae* interplay with the host upper and lower airways. *PLoS One*, 10 (4): e0123154. doi: 10.1371/journal.pone.0123154.

Eutsey, R. A., Hiller, N. L., Earl, J. P., Janto, B. A., Dahlgren, M. E., Ahmed, A., Powell, E., Schultz, M. P., Gilsdorf, J. R., Zhang, L., et al. (2013). Design and validation of a supragenome array for determination of the genomic content of *Haemophilus influenzae* isolates. *BMC Genomics*, 14: 484. doi: 10.1186/1471-2164-14-484.

Fagon, J. Y., Chastre, J., Trouillet, J. L., Domart, Y., Dombret, M. C., Bornet, M. & Gibert, C. (1990). Characterization of distal bronchial microflora during acute exacerbation of chronic bronchitis. Use of the protected specimen brush technique in 54 mechanically ventilated patients. *Am Rev Respir Dis*, 142 (5): 1004-8. doi: 10.1164/ajrccm/142.5.1004.

Farjo, R. S., Foxman, B., Patel, M. J., Zhang, L., Pettigrew, M. M., McCoy, S. I., Marrs, C. F. & Gilsdorf, J. R. (2004). Diversity and sharing of *Haemophilus influenzae* strains colonizing healthy children attending day-care centers. *Pediatr Infect Dis J*, 23 (1): 41-6. doi:

10.1097/01.inf.0000106981.89572.d1.

Feliziani, S., Marvig, R. L., Lujan, A. M., Moyano, A. J., Di Rienzo, J. A., Krogh Johansen, H., Molin, S. & Smania, A. M. (2014). Coexistence and within-host evolution of diversified lineages of hypermutable *Pseudomonas aeruginosa* in long-term cystic fibrosis infections. *PLoS Genet*, 10 (10): e1004651. doi: 10.1371/journal.pgen.1004651.

Fernaays, M. M., Lesse, A. J., Cai, X. & Murphy, T. F. (2006). Characterization of *igaB*, a second immunoglobulin A1 protease gene in nontypeable *Haemophilus influenzae*. *Infect Immun*, 74 (10): 5860-70. doi: 10.1128/IAI.00796-06.

Finney, L. J., Ritchie, A., Pollard, E., Johnston, S. L. & Mallia, P. (2014). Lower airway colonization and inflammatory response in COPD: a focus on *Haemophilus influenzae*. *Int J Chron Obstruct Pulmon Dis*, 9: 1119-32. doi: 10.2147/COPD.S54477.

Fischer, B. M., Pavlisko, E. & Voynow, J. A. (2011). Pathogenic triad in COPD: oxidative stress, protease-antiprotease imbalance, and inflammation. *Int J Chron Obstruct Pulmon Dis*, 6: 413-21. doi: 10.2147/COPD.S10770.

Fleischmann, R., Adams, M., White, O., Clayton, R., Kirkness, E., Kerlavage, A., Bult, C., Tomb, J., Dougherty, B., Merrick, J., et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 269 (5223): 496-512. doi: 10.1126/science.7542800.

Fox, J. W. & Lenski, R. E. (2015). From Here to Eternity--The theory and practice of a really long experiment. *PLoS Biol*, 13 (6): e1002185. doi: 10.1371/journal.pbio.1002185.

Gallaher, T. K., Wu, S., Webster, P. & Aguilera, R. (2006). Identification of biofilm proteins in non-typeable *Haemophilus Influenzae*. *BMC Microbiol*, 6: 65. doi: 10.1186/1471-2180-6-65.

Gao, R., Hu, Y., Li, Z., Sun, J., Wang, Q., Lin, J., Ye, H., Liu, F., Srinivas, S., Li, D., et al. (2016). Dissemination and mechanism for the MCR-1 colistin resistance. *PLoS Pathog*, 12 (11): e1005957. doi: 10.1371/journal.ppat.1005957.

Garcia-Pastor, L., Puerta-Fernandez, E. & Casadesus, J. (2018). Bistability and phase variation in *Salmonella enterica*. *Biochim Biophys Acta*. doi: 10.1016/j.bbagrm.2018.01.003.

Garmendia, J., Viadas, C., Calatayud, L., Mell, J. C., Marti-Lliteras, P., Euba, B., Llobet, E., Gil, C., Bengoechea, J. A., Redfield, R. J., et al. (2014). Characterization of nontypable *Haemophilus influenzae* isolates recovered from adult patients with underlying chronic lung disease reveals genotypic and phenotypic traits associated with persistent infection. *PLoS One*, 9 (5): e97020. doi: 10.1371/journal.pone.0097020.

Gauderman, W. J., Avol, E., Gilliland, F., Vora, H., Thomas, D., Berhane, K., McConnell, R., Kuenzli, N., Lurmann, F., Rappaport, E., et al. (2004). The effect of air pollution on lung development from 10 to 18 years of age. *N Engl J Med*, 351 (11): 1057-67. doi: 10.1056/NEJMoa040610.

Gauderman, W. J., Urman, R., Avol, E., Berhane, K., McConnell, R., Rappaport, E., Chang, R., Lurmann, F. & Gilliland, F. (2015). Association of improved air quality with lung development in children. *N Engl J Med*, 372 (10): 905-13. doi: 10.1056/NEJMoa1414123.

Geiser, M. (2010). Update on macrophage clearance of inhaled micro- and nanoparticles. *J Aerosol Med Pulm Drug Deliv*, 23 (4): 207-17. doi: 10.1089/jamp.2009.0797.

Gershon, A. S., Warner, L., Cascagnette, P., Victor, J. C. & To, T. (2011). Lifetime risk of developing chronic obstructive pulmonary disease: a longitudinal population study. *Lancet*, 378 (9795): 991-6. doi: 10.1016/S0140-6736(11)60990-2.

Ghose, C., Perez-Perez, G. I., van Doorn, L. J., Dominguez-Bello, M. G. & Blaser, M. J. (2005). High frequency of gastric colonization with multiple *Helicobacter pylori* strains in Venezuelan subjects. *J Clin Microbiol*, 43 (6): 2635-41. doi: 10.1128/JCM.43.6.2635-2641.2005.

Gilsdorf, J. R., Marrs, C. F. & Foxman, B. (2004). *Haemophilus influenzae*: genetic variability and natural selection to identify virulence factors. *Infect Immun*, 72 (5): 2457-61.

Glässer. (1910). Die fibrinöse Serosen- und Gelenk- entzündung der Ferkel. *Die Krankheiten des Sch.*

weines: 122–125.

Goldberg, A., Fridman, O., Ronin, I. & Balaban, N. Q. (2014). Systematic identification and

quantification of phase variation in commensal and pathogenic *Escherichia coli. Genome Med*, 6 (11): 112. doi: 10.1186/s13073-014-0112-4.

Gomez-De-Leon, P., Santos, J. I., Caballero, J., Gomez, D., Espinosa, L. E., Moreno, I., Pinero, D. & Cravioto, A. (2000). Genomic variability of *Haemophilus influenzae* isolated from Mexican children determined by using enterobacterial repetitive intergenic consensus sequences and PCR. *J Clin Microbiol*, 38 (7): 2504-11.

Gomez-Valero, L., Rusniok, C., Jarraud, S., Vacherie, B., Rouy, Z., Barbe, V., Medigue, C., Etienne, J. & Buchrieser, C. (2011). Extensive recombination events and horizontal gene transfer shaped the *Legionella pneumophila* genomes. *BMC Genomics*, 12: 536. doi: 10.1186/1471-2164-12-536.

Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. (2017). The dynamics of molecular evolution over 60,000 generations. *Nature*, 551 (7678): 45-50. doi: 10.1038/nature24287.

Goodgal, S. H. & Mitchell, M. (1984). Uptake of heterologous DNA by *Haemophilus influenzae*. *J Bacteriol*, 157 (3): 785-8.

Greub, G. & Raoult, D. (2004). Microorganisms resistant to free-living amoebae. *Clin Microbiol Rev*, 17 (2): 413-33.

Guttman, J. A. & Finlay, B. B. (2009). Tight junctions as targets of infectious agents. *Biochim Biophys Acta*, 1788 (4): 832-41. doi: 10.1016/j.bbamem.2008.10.028.

Hall-Stoodley, L., Hu, F. Z., Gieseke, A., Nistico, L., Nguyen, D., Hayes, J., Forbes, M., Greenberg, D. P., Dice, B., Burrows, A., et al. (2006). Direct detection of bacterial biofilms on the middle-ear mucosa of children with chronic otitis media. *JAMA*, 296 (2): 202-11. doi: 10.1001/jama.296.2.202.

Hallstrom, T., Resman, F., Ristovski, M. & Riesbeck, K. (2010). Binding of complement regulators to invasive nontypeable *Haemophilus influenzae* isolates is not increased compared to nasopharyngeal isolates, but serum resistance is linked to disease severity. *J Clin Microbiol*, 48 (3): 921-7. doi: 10.1128/JCM.01654-09.

Hamaguchi, S., Seki, M. & Tomono, K. (2012a). The role of neutrophil extracellular traps -a translational research between infection and allergy. *Arerugi*, 61 (12): 1729-35.

Hamaguchi, S., Seki, M., Yamamoto, N., Hirose, T., Matsumoto, N., Irisawa, T., Takegawa, R., Shimazu, T. & Tomono, K. (2012b). Case of invasive nontypeable *Haemophilus influenzae* respiratory tract infection with a large quantity of neutrophil extracellular traps in sputum. *J Inflamm Res*, 5: 137-40. doi: 10.2147/JIR.S39497.

Han, M. K., Postma, D., Mannino, D. M., Giardino, N. D., Buist, S., Curtis, J. L. & Martinez, F. J. (2007). Gender and chronic obstructive pulmonary disease: why it matters. *Am J Respir Crit Care Med*, 176 (12): 1179-84. doi: 10.1164/rccm.200704-553CC.

Henderson, I. R., Owen, P. & Nataro, J. P. (1999). Molecular switches--the ON and OFF of bacterial phase variation. *Mol Microbiol*, 33 (5): 919-32.

Herath, S. C. & Poole, P. (2013). Prophylactic antibiotic therapy for chronic obstructive pulmonary disease (COPD). *Cochrane Database Syst Rev* (11): CD009764. doi: 10.1002/14651858.CD009764.pub2.

Hindre, T., Knibbe, C., Beslon, G. & Schneider, D. (2012). New insights into bacterial adaptation through *in vivo* and *in silico* experimental evolution. *Nat Rev Microbiol*, 10 (5): 352-65. doi: 10.1038/nrmicro2750.

Hoenderdos, K. & Condliffe, A. (2013). The neutrophil in chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol*, 48 (5): 531-9. doi: 10.1165/rcmb.2012-0492TR.

Hogg, J. C., Chu, F., Utokaparch, S., Woods, R., Elliott, W. M., Buzatu, L., Cherniack, R. M., Rogers, R. M., Sciurba, F. C., Coxson, H. O., et al. (2004). The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N Engl J Med*, 350 (26): 2645-53. doi: 10.1056/NEJMoa032158.

Hogg, J. C. & Timens, W. (2009). The pathology of chronic obstructive pulmonary disease. *Annu Rev Pathol*, 4: 435-59. doi: 10.1146/annurev.pathol.4.110807.092145.

Holtzman, M. J., Tyner, J. W., Kim, E. Y., Lo, M. S., Patel, A. C., Shornick, L. P., Agapov, E. & Zhang, Y. (2005). Acute and chronic airway responses to viral infection: implications for asthma and chronic obstructive pulmonary disease. *Proc Am Thorac Soc*, 2 (2): 132-40. doi: 10.1513/pats.200502-015AW.

Hong, W., Juneau, R. A., Pang, B. & Swords, W. E. (2009). Survival of bacterial biofilms within neutrophil extracellular traps promotes nontypeable *Haemophilus influenzae* persistence in the chinchilla model for otitis media. *J Innate Immun*, 1 (3): 215-24. doi: 10.1159/000205937.

Howell, K. J., Weinert, L. A., Chaudhuri, R. R., Luan, S. L., Peters, S. E., Corander, J., Harris, D., Angen, O., Aragon, V., Bensaid, A., et al. (2014). The use of genome wide association methods to investigate pathogenicity, population structure and serovar in *Haemophilus parasuis*. *BMC Genomics*, 15: 1179. doi: 10.1186/1471-2164-15-1179.

Huse, H. K., Kwon, T., Zlosnik, J. E., Speert, D. P., Marcotte, E. M. & Whiteley, M. (2010). Parallel evolution in *Pseudomonas aeruginosa* over 39,000 generations *in vivo*. *MBio*, 1 (4). doi: 10.1128/mBio.00199-10.

Idicula, W. K., Jurcisek, J. A., Cass, N. D., Ali, S., Goodman, S. D., Elmaraghy, C. A., Jatana, K. R. & Bakaletz, L. O. (2016). Identification of biofilms in post-tympanostomy tube otorrhea. *Laryngoscope*, 126 (8): 1946-51. doi: 10.1002/lary.25826.

Ito, K. & Barnes, P. J. (2009). COPD as a disease of accelerated lung aging(a). *Rev Port Pneumol*, 15 (4): 743-6. doi: 10.1016/S0873-2159(15)30173-2.

Jackson, R. W., Johnson, L. J., Clarke, S. R. & Arnold, D. L. (2011). Bacterial pathogen evolution: breaking news. *Trends Genet*, 27 (1): 32-40. doi: 10.1016/j.tig.2010.10.001.

Jahnsen, F. L., Strickland, D. H., Thomas, J. A., Tobagus, I. T., Napoli, S., Zosky, G. R., Turner, D. J., Sly, P. D., Stumbles, P. A. & Holt, P. G. (2006). Accelerated antigen sampling and transport by airway mucosal dendritic cells following inhalation of a bacterial stimulus. *J Immunol*, 177 (9): 5861-7.

Jeffery, P. K. (1999). Differences and similarities between chronic obstructive pulmonary disease and asthma. *Clin Exp Allergy*, 29 Suppl 2: 14-26.

Jennings, M. P., Srikhanta, Y. N., Moxon, E. R., Kramer, M., Poolman, J. T., Kuipers, B. & van der Ley, P. (1999). The genetic basis of the phase variation repertoire of lipopolysaccharide immunotypes in *Neisseria meningitidis*. *Microbiology*, 145 (Pt 11): 3013-21. doi: 10.1099/00221287-145-11-3013.

Jiang, J., Thoren, P., Caligiuri, G., Hansson, G. K. & Pernow, J. (1999). Enhanced phenylephrine-induced rhythmic activity in the atherosclerotic mouse aorta via an increase in opening of KCa channels: relation to Kv channels and nitric oxide. *Br J Pharmacol*, 128 (3): 637-46. doi: 10.1038/sj.bjp.0702855.

Johansen, H. K. & Hoiby, N. (1992). Seasonal onset of initial colonisation and chronic infection with *Pseudomonas aeruginosa* in patients with cystic fibrosis in Denmark. *Thorax*, 47 (2): 109-11.

Jolivet-Gougeon, A., Kovacs, B., Le Gall-David, S., Le Bars, H., Bousarghin, L., Bonnaure-Mallet, M., Lobel, B., Guille, F., Soussy, C. J. & Tenke, P. (2011). Bacterial hypermutation: clinical implications. *J Med Microbiol*, 60 (Pt 5): 563-73. doi: 10.1099/jmm.0.024083-0.

Jorth, P., Staudinger, B. J., Wu, X., Hisert, K. B., Hayden, H., Garudathri, J., Harding, C. L., Radey, M. C., Rezayat, A., Bautista, G., et al. (2015). Regional isolation drives bacterial diversification within cystic fibrosis lungs. *Cell Host Microbe*, 18 (3): 307-19. doi: 10.1016/j.chom.2015.07.006.

Jules, M. & Buchrieser, C. (2007). *Legionella pneumophila* adaptation to intracellular life and the host response: clues from genomics and transcriptomics. *FEBS Lett*, 581 (15): 2829-38. doi: 10.1016/j.febslet.2007.05.026.

Juneau, R. A., Pang, B., Weimer, K. E., Armbruster, C. E. & Swords, W. E. (2011). Nontypeable *Haemophilus influenzae* initiates formation of neutrophil extracellular traps. *Infect Immun*, 79 (1): 431-8. doi: 10.1128/IAI.00660-10.

Jurcisek, J. A. & Bakaletz, L. O. (2007). Biofilms formed by nontypeable *Haemophilus influenzae in vivo* contain both double-stranded DNA and type IV pilin protein. *J Bacteriol*, 189 (10): 3868-75. doi: 10.1128/JB.01935-06.

Jurcisek, J. A., Brockman, K. L., Novotny, L. A., Goodman, S. D. & Bakaletz, L. O. (2017). Nontypeable *Haemophilus influenzae* releases DNA and DNABII proteins via a T4SS-like complex and ComE of the type IV pilus machinery. *Proc Natl Acad Sci U S A*, 114 (32): E6632-E6641. doi: 10.1073/pnas.1705508114.

Kennemann, L., Didelot, X., Aebischer, T., Kuhn, S., Drescher, B., Droege, M., Reinhardt, R., Correa,

P., Meyer, T. F., Josenhans, C., et al. (2011). *Helicobacter pylori* genome evolution during human infection. *Proc Natl Acad Sci U S A*, 108 (12): 5033-8. doi: 10.1073/pnas.1018444108.

Kilian, M. (1991). *Haemophilus*. *Manual of clinical microbiology. American Society for Microbiology, Washington, DC*: 463-470.

Kim, V. & Criner, G. J. (2013). Chronic bronchitis and chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*, 187 (3): 228-37. doi: 10.1164/rccm.201210-1843CI.

King L.B., P. B., Perez A.C., Reimche J.L., Kirse D.J., Whigham A.S., Evans A.K., and Swords W.E. (2013). Observation of viable nontypeable *Haemophilus Influenzae* bacteria within neutrophil extracellular traps in clinical samples from chronic otitis media. *Otolaryngology*, 3 (145).

King, P. (2012). *Haemophilus influenzae* and the lung (*Haemophilus* and the lung). *Clin Transl Med*, 1 (1): 10. doi: 10.1186/2001-1326-1-10.

King, P. T. & Sharma, R. (2015). The lung immune response to nontypeable *Haemophilus influenzae* (Lung immunity to NTHi). *J Immunol Res*, 2015: 706376. doi: 10.1155/2015/706376.

Kirkman, J. B. & Crawford, J. J. (1971). Serotyping of noncapsular *Haemophilus influenzae*. *Applied Microbiology*, 22 (1): 133-134.

Knapp, H. R. & Melly, M. A. (1986). Bactericidal effects of polyunsaturated fatty acids. *J Infect Dis*, 154 (1): 84-94.

Knowles, M. R. & Boucher, R. C. (2002). Mucus clearance as a primary innate defense mechanism for mammalian airways. *J Clin Invest*, 109 (5): 571-7. doi: 10.1172/JCI15217.

Kostyanev, T. S. & Sechanova, L. P. (2012). Virulence factors and mechanisms of antibiotic resistance of *Haemophilus influenzae*. *Folia Med (Plovdiv)*, 54 (1): 19-23.

Krasan, G. P., Cutter, D., Block, S. L. & St Geme, J. W., 3rd. (1999). Adhesin expression in matched nasopharyngeal and middle ear isolates of nontypeable *Haemophilus influenzae* from children with acute otitis media. *Infect Immun*, 67 (1): 449-54.

Kruse, H. & Sorum, H. (1994). Transfer of multiple drug resistance plasmids between bacteria of diverse origins in natural microenvironments. *Appl Environ Microbiol*, 60 (11): 4015-21.

Kuehnert, M. J., Kruszon-Moran, D., Hill, H. A., McQuillan, G., McAllister, S. K., Fosheim, G., McDougal, L. K., Chaitram, J., Jensen, B., Fridkin, S. K., et al. (2006). Prevalence of *Staphylococcus aureus* nasal colonization in the United States, 2001-2002. *J Infect Dis*, 193 (2): 172-9. doi: 10.1086/499632.

Kusters, J. G., van Vliet, A. H. & Kuipers, E. J. (2006). Pathogenesis of *Helicobacter pylori* infection. *Clin Microbiol Rev*, 19 (3): 449-90. doi: 10.1128/CMR.00054-05.

L'Ecuyer, C. & Boulanger, P. (1970). Enzootic pneumonia in pigs: identification of a causative mycoplasma in infected pigs and in cultures by immunofluorescent staining. *Can J Comp Med*, 34 (1): 38-46.

Laabei, M., Uhlemann, A. C., Lowy, F. D., Austin, E. D., Yokoyama, M., Ouadi, K., Feil, E., Thorpe, H. A., Williams, B., Perkins, M., et al. (2015). Evolutionary trade-offs underlie the multi-faceted virulence of *Staphylococcus aureus*. *PLoS Biol*, 13 (9): e1002229. doi: 10.1371/journal.pbio.1002229.

Lacross, N. C., Marrs, C. F., Patel, M., Sandstedt, S. A. & Gilsdorf, J. R. (2008). High genetic diversity of nontypeable *Haemophilus influenzae* isolates from two children attending a day care center. *J Clin Microbiol*, 46 (11): 3817-21. doi: 10.1128/JCM.00940-08.

Lambris, J. D., Ricklin, D. & Geisbrecht, B. V. (2008). Complement evasion by human pathogens. *Nat Rev Microbiol*, 6 (2): 132-42. doi: 10.1038/nrmicro1824.

Langereis, J. D. & de Jonge, M. I. (2015). Invasive disease caused by nontypeable *Haemophilus influenzae*. *Emerg Infect Dis*, 21 (10): 1711-8. doi: 10.3201/eid2110.150004.

Ledala, N., Sengupta, M., Muthaiyan, A., Wilkinson, B. J. & Jayaswal, R. K. (2010). Transcriptomic response of *Listeria monocytogenes* to iron limitation and *fur* mutation. *Appl Environ Microbiol*, 76 (2): 406-16. doi: 10.1128/AEM.01389-09.

Lee, A. H., Flibotte, S., Sinha, S., Paiero, A., Ehrlich, R. L., Balashov, S., Ehrlich, G. D., Zlosnik, J. E.,

Mell, J. C. & Nislow, C. (2017). Phenotypic diversity and genotypic flexibility of *Burkholderia cenocepacia* during long-term chronic infection of cystic fibrosis lungs. *Genome Res*, 27 (4): 650-662. doi: 10.1101/gr.213363.116.

Lehrer, R. I. (2004). Primate defensins. *Nat Rev Microbiol*, 2 (9): 727-38. doi: 10.1038/nrmicro976.

Leitão, J. H. F., J.R.; Sousa, S.A.; Pita, T. and Guerreiro, S.I. (2017). *Burkholderia cepacia* complex infections among cystic fibrosis patients: Perspectives and challenges. progress in understanding Cystic Fibrosis.

Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. (1991). Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 Generations. *The American Naturalist*, 138 (6): 1315-1341. doi: 10.1086/285289.

Lenski, R. E. (2017). Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J*, 11 (10): 2181-2194. doi: 10.1038/ismej.2017.69.

Li, Y., Gierahn, T., Thompson, C. M., Trzcinski, K., Ford, C. B., Croucher, N., Gouveia, P., Flechtner, J. B., Malley, R. & Lipsitch, M. (2012). Distinct effects on diversifying selection by two mechanisms of immunity against *Streptococcus pneumoniae*. *PLoS Pathog*, 8 (11): e1002989. doi: 10.1371/journal.ppat.1002989.

Lieberman, T. D., Michel, J. B., Aingaran, M., Potter-Bynoe, G., Roux, D., Davis, M. R., Jr., Skurnik, D., Leiby, N., LiPuma, J. J., Goldberg, J. B., et al. (2011). Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet*, 43 (12): 1275-80. doi: 10.1038/ng.997.

Lieberman, T. D., Flett, K. B., Yelin, I., Martin, T. R., McAdam, A. J., Priebe, G. P. & Kishony, R. (2014). Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nat Genet*, 46 (1): 82-7. doi: 10.1038/ng.2848.

Lieberman, T. D., Wilson, D., Misra, R., Xiong, L. L., Moodley, P., Cohen, T. & Kishony, R. (2016). Genomic diversity in autopsy samples reveals within-host dissemination of HIV-associated *Mycobacterium tuberculosis*. *Nat Med*, 22 (12): 1470-1474. doi: 10.1038/nm.4205.

Lim, M. Y. & Thomas, P. S. (2013). Biomarkers in exhaled breath condensate and serum of chronic obstructive pulmonary disease and non-small-cell lung cancer. *Int J Chronic Dis*, 2013: 578613. doi: 10.1155/2013/578613.

Linden, S. K., Sutton, P., Karlsson, N. G., Korolik, V. & McGuckin, M. A. (2008). Mucins in the mucosal barrier to infection. *Mucosal Immunol*, 1 (3): 183-97. doi: 10.1038/mi.2008.5.

Ling, S. H. & van Eeden, S. F. (2009). Particulate matter air pollution exposure: role in the development and exacerbation of chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis*, 4: 233-43.

Little, T. W. (1970). *Haemophilus* infection in pigs. *Vet Rec*, 87 (14): 399-402.

Liu, Q., Yeo, W. S. & Bae, T. (2016). The SaeRS two-component system of *Staphylococcus aureus*. *Genes (Basel)*, 7 (10). doi: 10.3390/genes7100081.

Lopez-Collazo, E., Jurado, T., de Dios Caballero, J., Perez-Vazquez, M., Vindel, A., Hernandez-Jimenez, E., Tamames, J., Cubillos-Zapata, C., Manrique, M., Tobes, R., et al. (2015). *In vivo* attenuation and genetic evolution of a ST247-SCCmecI MRSA clone after 13 years of pathogenic bronchopulmonary colonization in a patient with cystic fibrosis: implications of the innate immune response. *Mucosal Immunol*, 8 (2): 362-71. doi: 10.1038/mi.2014.73.

Lopez-Gomez, A., Cano, V., Moranta, D., Morey, P., Garcia del Portillo, F., Bengoechea, J. A. & Garmendia, J. (2012). Host cell kinases, α5 and β1 integrins, and Rac1 signalling on the microtubule cytoskeleton are important for non-typable *Haemophilus influenzae* invasion of respiratory epithelial cells. *Microbiology*, 158 (Pt 9): 2384-98. doi: 10.1099/mic.0.059972-0.

Lorenz, M. G. & Wackernagel, W. (1994). Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev*, 58 (3): 563-602.

Lowy, F. D. (1998). *Staphylococcus aureus* infections. *N Engl J Med*, 339 (8): 520-32. doi: 10.1056/NEJM199808203390806.

MacNee, W. (2006). Pathology, pathogenesis, and pathophysiology. *BMJ*, 332 (7551): 1202-1204. doi: 10.1136/bmj.332.7551.1202.

Maina, J. N. (2000). Comparative respiratory morphology: themes and principles in the design and construction of the gas exchangers. *Anat Rec*, 261 (1): 25-44.

Malhotra, D., Thimmulappa, R., Vij, N., Navas-Acien, A., Sussan, T., Merali, S., Zhang, L., Kelsen, S. G., Myers, A., Wise, R., et al. (2009). Heightened endoplasmic reticulum stress in the lungs of patients with chronic obstructive pulmonary disease: the role of *Nrf2*-regulated proteasomal activity. *Am J Respir Crit Care Med*, 180 (12): 1196-207. doi: 10.1164/rccm.200903-0324OC.

Malhotra, S., Deshmukh, S. S. & Dastidar, S. G. (2012). COX inhibitors for airway inflammation. *Expert Opin Ther Targets*, 16 (2): 195-207. doi: 10.1517/14728222.2012.661416.

Markussen, T., Marvig, R. L., Gomez-Lozano, M., Aanaes, K., Burleigh, A. E., Hoiby, N., Johansen, H. K., Molin, S. & Jelsbak, L. (2014). Environmental heterogeneity drives within-host diversification and evolution of *Pseudomonas aeruginosa*. *MBio*, 5 (5): e01592-14. doi: 10.1128/mBio.01592-14.

Marti-Lliteras, P., Regueiro, V., Morey, P., Hood, D. W., Saus, C., Sauleda, J., Agusti, A. G., Bengoechea, J. A. & Garmendia, J. (2009). Nontypeable *Haemophilus influenzae* clearance by alveolar macrophages is impaired by exposure to cigarette smoke. *Infect Immun*, 77 (10): 4232-42. doi: 10.1128/IAI.00305-09.

Martin, T. R. & Frevert, C. W. (2005). Innate immunity in the lungs. *Proc Am Thorac Soc*, 2 (5): 403-11. doi: 10.1513/pats.200508-090JS.

Marvig, R. L., Johansen, H. K., Molin, S. & Jelsbak, L. (2013). Genome analysis of a transmissible lineage of *Pseudomonas aeruginosa* reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genet*, 9 (9): e1003741. doi: 10.1371/journal.pgen.1003741.

Marvig, R. L., Damkiaer, S., Khademi, S. M., Markussen, T. M., Molin, S. & Jelsbak, L. (2014). Within-host evolution of *Pseudomonas aeruginosa* reveals adaptation toward iron acquisition from hemoglobin. *MBio*, 5 (3): e00966-14. doi: 10.1128/mBio.00966-14.

Marvig, R. L., Sommer, L. M., Jelsbak, L., Molin, S. & Johansen, H. K. (2015a). Evolutionary insight from whole-genome sequencing of *Pseudomonas aeruginosa* from cystic fibrosis patients. *Future Microbiol*, 10 (4): 599-611. doi: 10.2217/fmb.15.3.

Marvig, R. L., Sommer, L. M., Molin, S. & Johansen, H. K. (2015b). Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. *Nat Genet*, 47 (1): 57-64. doi: 10.1038/ng.3148.

Maughan, H. & Redfield, R. J. (2009). Extensive variation in natural competence in *Haemophilus influenzae*. *Evolution*, 63 (7): 1852-66. doi: 10.1111/j.1558-5646.2009.00658.x.

Maughan, H. S., S.; Wilson, L. and Redfield, R. (2018). Competence, DNA uptake and transformation in *Pasteurellaceae*. Pasteurellaceae: Biology, Genomics and Molecular Aspects: Caister Academic Press.

McAdam, P. R., Holmes, A., Templeton, K. E. & Fitzgerald, J. R. (2011). Adaptive evolution of *Staphylococcus aureus* during chronic endobronchial infection of a cystic fibrosis patient. *PLoS One*, 6 (9): e24301. doi: 10.1371/journal.pone.0024301.

McAdam, P. R., Vander Broek, C. W., Lindsay, D. S., Ward, M. J., Hanson, M. F., Gillies, M., Watson, M., Stevens, J. M., Edwards, G. F. & Fitzgerald, J. R. (2014). Gene flow in environmental *Legionella pneumophila* leads to genetic and pathogenic heterogeneity within a Legionnaires' disease outbreak. *Genome Biol*, 15 (11): 504. doi: 10.1186/PREACCEPT-1675723368141690.

McDade, J. E., Shepard, C. C., Fraser, D. W., Tsai, T. R., Redus, M. A. & Dowdle, W. R. (1977). Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease. *N Engl J Med*, 297 (22): 1197-203. doi: 10.1056/NEJM197712012972202.

McIntyre, T. M., Prescott, S. M. & Stafforini, D. M. (2009). The emerging roles of PAF acetylhydrolase. *J Lipid Res*, 50 Suppl: S255-9. doi: 10.1194/jlr.R800024-JLR200.

Meacci, F., Orru, G., Iona, E., Giannoni, F., Piersimoni, C., Pozzi, G., Fattorini, L. & Oggioni, M. R. (2005). Drug resistance evolution of a *Mycobacterium tuberculosis* strain from a noncompliant patient. *J Clin Microbiol*, 43 (7): 3114-20. doi: 10.1128/JCM.43.7.3114-3120.2005.

Meats, E., Feil, E. J., Stringer, S., Cody, A. J., Goldstein, R., Kroll, J. S., Popovic, T. & Spratt, B. G. (2003). Characterization of encapsulated and noncapsulated *Haemophilus influenzae* and

determination of phylogenetic relationships by multilocus sequence typing. *J Clin Microbiol*, 41 (4): 1623-36.

Mell, J. C., Shumilina, S., Hall, I. M. & Redfield, R. J. (2011). Transformation of natural genetic variation into *Haemophilus influenzae* genomes. *PLoS Pathog*, 7 (7): e1002151. doi: 10.1371/journal.ppat.1002151.

Mell, J. C., Hall, I. M. & Redfield, R. J. (2012). Defining the DNA uptake specificity of naturally competent *Haemophilus influenzae* cells. *Nucleic Acids Res*, 40 (17): 8536-49. doi: 10.1093/nar/gks640.

Mell, J. C., Lee, J. Y., Firme, M., Sinha, S. & Redfield, R. J. (2014). Extensive cotransformation of natural variation into chromosomes of naturally competent *Haemophilus influenzae*. *G3 (Bethesda)*, 4 (4): 717-31. doi: 10.1534/g3.113.009597.

Mell, J. C., Viadas, C., Moleres, J., Sinha, S., Fernandez-Calvet, A., Porsch, E. A., St Geme, J. W., 3rd, Nislow, C., Redfield, R. J. & Garmendia, J. (2016). Transformed recombinant enrichment profiling rapidly identifies HMW1 as an intracellular invasion locus in *Haemophilus influenzae*. *PLoS Pathog*, 12 (4): e1005576. doi: 10.1371/journal.ppat.1005576.

Melnyk, A. H., Wong, A. & Kassen, R. (2015). The fitness costs of antibiotic resistance mutations. *Evol Appl*, 8 (3): 273-83. doi: 10.1111/eva.12196.

Merker, M., Kohl, T. A., Roetzer, A., Truebe, L., Richter, E., Rusch-Gerdes, S., Fattorini, L., Oggioni, M. R., Cox, H., Varaine, F., et al. (2013). Whole genome sequencing reveals complex evolution patterns of multidrug-resistant *Mycobacterium tuberculosis* Beijing strains in patients. *PLoS One*, 8 (12): e82551. doi: 10.1371/journal.pone.0082551.

Meyer, M. C., Rastogi, P., Beckett, C. S. & McHowat, J. (2005). Phospholipase A2 inhibitors as potential anti-inflammatory agents. *Curr Pharm Des*, 11 (10): 1301-12.

Miethke, M. & Marahiel, M. A. (2007). Siderophore-based iron acquisition and pathogen control. *Microbiol Mol Biol Rev*, 71 (3): 413-51. doi: 10.1128/MMBR.00012-07.

Mircescu, M. M., Lipuma, L., van Rooijen, N., Pamer, E. G. & Hohl, T. M. (2009). Essential role for neutrophils but not alveolar macrophages at early time points following *Aspergillus fumigatus* infection. *J Infect Dis*, 200 (4): 647-56. doi: 10.1086/600380.

Molfino, N. A. & Coyle, A. J. (2008). Gene-environment interactions in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis*, 3 (3): 491-7.

Moliner, C., Raoult, D. & Fournier, P. E. (2009). Evidence of horizontal gene transfer between amoeba and bacteria. *Clin Microbiol Infect*, 15 Suppl 2: 178-80. doi: 10.1111/j.1469-0691.2008.02216.x.

Monso, E., Ruiz, J., Rosell, A., Manterola, J., Fiz, J., Morera, J. & Ausina, V. (1995). Bacterial infection in chronic obstructive pulmonary disease. A study of stable and exacerbated outpatients using the protected specimen brush. *Am J Respir Crit Care Med*, 152 (4 Pt 1): 1316-20. doi: 10.1164/ajrccm.152.4.7551388.

Montano, V., Didelot, X., Foll, M., Linz, B., Reinhardt, R., Suerbaum, S., Moodley, Y. & Jensen, J. D. (2015). Worldwide population structure, long-term demography, and local adaptation of *Helicobacter pylori*. *Genetics*, 200 (3): 947-63. doi: 10.1534/genetics.115.176404.

Moran, N. A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell*, 108 (5): 583-6.

Morey, P., Cano, V., Marti-Lliteras, P., Lopez-Gomez, A., Regueiro, V., Saus, C., Bengoechea, J. A. & Garmendia, J. (2011). Evidence for a non-replicative intracellular stage of nontypeable *Haemophilus influenzae* in epithelial cells. *Microbiology*, 157 (Pt 1): 234-50. doi: 10.1099/mic.0.040451-0.

Morjan, C. L. & Rieseberg, L. H. (2004). How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Mol Ecol*, 13 (6): 1341-56. doi: 10.1111/j.1365-294X.2004.02164.x.

Moxon, R., Bayliss, C. & Hood, D. (2006). Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet*, 40: 307-33. doi: 10.1146/annurev.genet.40.110405.090442.

Munita, J. M. & Arias, C. A. (2016). Mechanisms of antibiotic resistance. *Microbiol Spectr*, 4 (2). doi:

10.1128/microbiolspec.VMBF-0016-2015.

Munoz-Ramirez, Z. Y., Mendez-Tenorio, A., Kato, I., Bravo, M. M., Rizzato, C., Thorell, K., Torres, R., Aviles-Jimenez, F., Camorlinga, M., Canzian, F., et al. (2017). Whole genome sequence and phylogenetic analysis show *Helicobacter pylori* strains from latin america have followed a unique evolution pathway. *Front Cell Infect Microbiol*, 7: 50. doi: 10.3389/fcimb.2017.00050.

Munson, R., Jr., Bailey, C. & Grass, S. (1989). Diversity of the outer membrane protein P2 gene from major clones of *Haemophilus influenzae* type b. *Mol Microbiol*, 3 (12): 1797-803.

Muro, S. (2011). Cigarette smoking is the most important causal factor for developing chronic obstructive pulmonary disease (COPD). *Nihon Rinsho*, 69 (10): 1735-40.

Murphy, T. F., Sethi, S., Klingman, K. L., Brueggemann, A. B. & Doern, G. V. (1999). Simultaneous respiratory tract colonization by multiple strains of nontypeable *Haemophilus influenzae* in chronic obstructive pulmonary disease: implications for antibiotic therapy. *J Infect Dis*, 180 (2): 404-9. doi: 10.1086/314870.

Murphy, T. F. & Kirkham, C. (2002). Biofilm formation by nontypeable *Haemophilus influenzae*: strain variability, outer membrane antigen expression and role of pili. *BMC Microbiol*, 2: 7.

Murphy, T. F., Brauer, A. L., Schiffmacher, A. T. & Sethi, S. (2004). Persistent colonization by *Haemophilus influenzae* in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*, 170 (3): 266-72. doi: 10.1164/rccm.200403-354OC.

Murphy, T. F., Kirkham, C., Gallo, M. C., Yang, Y., Wilding, G. E. & Pettigrew, M. M. (2017). Immunoglobulin A protease variants facilitate intracellular survival in epithelial cellsb by Nontypeable *Haemophilus influenzae* that persist in the human respiratory tract in chronic obstructive pulmonary disease. *J Infect Dis*, 216 (10): 1295-1302. doi: 10.1093/infdis/jix471.

Musser, J. M., Kroll, J. S., Moxon, E. R. & Selander, R. K. (1988). Clonal population structure of encapsulated *Haemophilus influenzae*. *Infect Immun*, 56 (8): 1837-45.

Nedbalcova, K. S., P; Jaglic, Z.; Ondriasova, R. and Kucerova, Z. (2006). *Haemophilus parasuis* and Glässer's disease in pigs: A review. *Veterinarni Medicina*, 51 (5): 168-179.

Niederwerder, M. C. (2017). Role of the microbiome in swine respiratory disease. *Vet Microbiol*, 209: 97-106. doi: 10.1016/j.vetmic.2017.02.017.

Oliveira, S., Batista, L., Torremorell, M. & Pijoan, C. (2001). Experimental colonization of piglets and gilts with systemic strains of *Haemophilus parasuis* and *Streptococcus suis* to prevent disease. *Can J Vet Res*, 65 (3): 161-7.

Owens, S. R. & Smith, L. B. (2011). Molecular aspects of *H. pylori*-Related MALT lymphoma. *Patholog Res Int*, 2011: 193149. doi: 10.4061/2011/193149.

Palmer, A. C. & Kishony, R. (2013). Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nat Rev Genet*, 14 (4): 243-8. doi: 10.1038/nrg3351.

Pandey, K. C., De, S. & Mishra, P. K. (2017). Role of proteases in chronic obstructive pulmonary disease. *Front Pharmacol*, 8: 512. doi: 10.3389/fphar.2017.00512.

Parrow, N. L., Fleming, R. E. & Minnick, M. F. (2013). Sequestration and scavenging of iron in infection. *Infect Immun*, 81 (10): 3503-14. doi: 10.1128/IAI.00602-13.

Pechous, R. D. (2017). With friends like these: The complex role of neutrophils in the progression of severe pneumonia. *Front Cell Infect Microbiol*, 7: 160. doi: 10.3389/fcimb.2017.00160.

Peet, R. L., Fry, J., Lloyd, J., Henderson, J., Curran, J. & Moir, D. (1983). *Haemophilus parasuis* septicaemia in pigs. *Australian Veterinary Journal*, 60 (6): 187-187. doi: 10.1111/j.1751-0813.1983.tb05960.x.

Peltola, H. (2000). Worldwide *Haemophilus influenzae* type b disease at the beginning of the 21st century: Global analysis of the disease burden 25 years after the use of the polysaccharide vaccine and a decade after the advent of conjugates. *Clinical Microbiology Reviews*, 13 (2): 302-317.

Pereira, M. I. & Medeiros, J. A. (2014). Role of *Helicobacter pylori* in gastric mucosa-associated lymphoid tissue lymphomas. *World J Gastroenterol*, 20 (3): 684-98. doi: 10.3748/wjg.v20.i3.684.

Perez, L. M., Codony, F., Lopez Leyton, D., Fittipaldi, M., Adrados, B. & Morato, J. (2010).

Quantification of *Helicobacter pylori* levels in soil samples from public playgrounds in Spain. *J Zhejiang Univ Sci B*, 11 (1): 27-9. doi: 10.1631/jzus.B0900238.

Perez-Lago, L., Comas, I., Navarro, Y., Gonzalez-Candelas, F., Herranz, M., Bouza, E. & Garcia-de-Viedma, D. (2014). Whole genome sequencing analysis of intrapatient microevolution in *Mycobacterium tuberculosis*: potential impact on the inference of tuberculosis transmission. *J Infect Dis*, 209 (1): 98-108. doi: 10.1093/infdis/jit439.

Person, A., and Mintz, M. (2006). Anatomy and physiology of the respiratory tract. *Current Clinical Practice: Disorders of the Respiratory Tract: Common Challenges in Primary Care*.

Petrache, I. & Petrusca, D. N. (2013). The involvement of sphingolipids in chronic obstructive pulmonary diseases. *Handb Exp Pharmacol* (216): 247-64. doi: 10.1007/978-3-7091-1511-4_12.

Pettigrew, M. M., Foxman, B., Ecevit, Z., Marrs, C. F. & Gilsdorf, J. (2002). Use of pulsed-field gel electrophoresis, enterobacterial repetitive intergenic consensus typing, and automated ribotyping to assess genomic variability among strains of nontypeable *Haemophilus influenzae*. *J Clin Microbiol*, 40 (2): 660-2.

Pfeiffer. (1892). Preliminary communication on the exciting causes of influenza. *British Medical Journal 1*, I (1620): 128.

Pin, C., Hansen, T., Munoz-Cuevas, M., de Jonge, R., Rosenkrantz, J. T., Lofstrom, C., Aarts, H. & Olsen, J. E. (2012). The transcriptional heat shock response of *Salmonella typhimurium* shows hysteresis and heated cells show increased resistance to heat and acid stress. *PLoS One*, 7 (12): e51196. doi: 10.1371/journal.pone.0051196.

Pohl, C., Hermanns, M. I., Uboldi, C., Bock, M., Fuchs, S., Dei-Anang, J., Mayer, E., Kehe, K., Kummer, W. & Kirkpatrick, C. J. (2009). Barrier functions and paracellular integrity in human cell culture models of the proximal respiratory unit. *Eur J Pharm Biopharm*, 72 (2): 339-49. doi: 10.1016/j.ejpb.2008.07.012.

Pompilio, A., Crocetta, V., Ghosh, D., Chakrabarti, M., Gherardi, G., Vitali, L. A., Fiscarelli, E. & Di Bonaventura, G. (2016). *Stenotrophomonas maltophilia* phenotypic and genotypic diversity during a 10-year colonization in the lungs of a cystic fibrosis patient. *Front Microbiol*, 7: 1551. doi: 10.3389/fmicb.2016.01551.

Post, D. M., Held, J. M., Ketterer, M. R., Phillips, N. J., Sahu, A., Apicella, M. A. & Gibson, B. W. (2014). Comparative analyses of proteins from *Haemophilus influenzae* biofilm and planktonic populations using metabolic labeling and mass spectrometry. *BMC Microbiol*, 14: 329. doi: 10.1186/s12866-014-0329-9.

Post, J. C. (2001). Direct evidence of bacterial biofilms in otitis media. *Laryngoscope*, 111 (12): 2083-94. doi: 10.1097/00005537-200112000-00001.

Power, P. M., Sweetman, W. A., Gallacher, N. J., Woodhall, M. R., Kumar, G. A., Moxon, E. R. & Hood, D. W. (2009). Simple sequence repeats in *Haemophilus influenzae*. *Infect Genet Evol*, 9 (2): 216-28. doi: 10.1016/j.meegid.2008.11.006.

Pragman, A. A., Kim, H. B., Reilly, C. S., Wendt, C. & Isaacson, R. E. (2012). The lung microbiome in moderate and severe chronic obstructive pulmonary disease. *PLoS One*, 7 (10): e47305. doi: 10.1371/journal.pone.0047305.

Prasadarao, N. V., Lysenko, E., Wass, C. A., Kim, K. S. & Weiser, J. N. (1999). Opacity-associated protein A contributes to the binding of *Haemophilus influenzae* to Chang epithelial cells. *Infect Immun*, 67 (8): 4153-60.

Price, E. P., Sarovich, D. S., Mayo, M., Tuanyok, A., Drees, K. P., Kaestli, M., Beckstrom-Sternberg, S. M., Babic-Sternberg, J. S., Kidd, T. J., Bell, S. C., et al. (2013). Within-host evolution of *Burkholderia pseudomallei* over a twelve-year chronic carriage infection. *MBio*, 4 (4). doi: 10.1128/mBio.00388-13.

Price, L. B., Hungate, B. A., Koch, B. J., Davis, G. S. & Liu, C. M. (2017). Colonizing opportunistic pathogens (COPs): The beasts in all of us. *PLoS Pathog*, 13 (8): e1006369. doi: 10.1371/journal.ppat.1006369.

Prickett, M. H., Hauser, A. R., McColley, S. A., Cullina, J., Potter, E., Powers, C. & Jain, M. (2017). Aminoglycoside resistance of *Pseudomonas aeruginosa* in cystic fibrosis results from convergent evolution in the *mexZ* gene. *Thorax*, 72 (1): 40-47. doi: 10.1136/thoraxjnl-2015-208027.

Quanjer, P. H., Stanojevic, S., Stocks, J., Hall, G. L., Prasad, K. V., Cole, T. J., Rosenthal, M., Perez-Padilla, R., Hankinson, J. L., Falaschetti, E., et al. (2010). Changes in the FEV(1)/FVC ratio during childhood and adolescence: an intercontinental study. *Eur Respir J*, 36 (6): 1391-9. doi: 10.1183/09031936.00164109.

Quanjer, P. H., Hall, G. L., Stanojevic, S., Cole, T. J., Stocks, J. & Global Lungs, I. (2012). Age- and height-based prediction bias in spirometry reference equations. *Eur Respir J*, 40 (1): 190-7. doi: 10.1183/09031936.00161011.

Quinn, P. J. M., B.K.; Leonard, F.C.; Hartigan, P.; Fanning, S. and Fitzpatrick, E.S. (2011). *Veterinary Microbiology and Microbial Disease*: Wiley.

Rahman, I. & Adcock, I. M. (2006). Oxidative stress and redox regulation of lung inflammation in COPD. *Eur Respir J*, 28 (1): 219-42. doi: 10.1183/09031936.06.00053805.

Ranjbar, R., Khamesipour, F., Jonaidi-Jafari, N. & Rahimi, E. (2016a). *Helicobacter pylori* in bottled mineral water: genotyping and antimicrobial resistance properties. *BMC Microbiol*, 16: 40. doi: 10.1186/s12866-016-0647-1.

Ranjbar, R., Khamesipour, F., Jonaidi-Jafari, N. & Rahimi, E. (2016b). *Helicobacter pylori* isolated from Iranian drinking water: *vacA*, *cagA*, *iceA*, *oipA* and *babA2* genotype status and antimicrobial resistance properties. *FEBS Open Bio*, 6 (5): 433-41. doi: 10.1002/2211-5463.12054.

Ray, K., Marteyn, B., Sansonetti, P. J. & Tang, C. M. (2009). Life on the inside: the intracellular lifestyle of cytosolic bacteria. *Nat Rev Microbiol*, 7 (5): 333-40. doi: 10.1038/nrmicro2112.

Redfield, R. J., Cameron, A. D., Qian, Q., Hinds, J., Ali, T. R., Kroll, J. S. & Langford, P. R. (2005). A novel CRP-dependent regulon controls expression of competence genes in *Haemophilus influenzae*. *J Mol Biol*, 347 (4): 735-47. doi: 10.1016/j.jmb.2005.01.012.

Redfield, R. J., Findlay, W. A., Bosse, J., Kroll, J. S., Cameron, A. D. & Nash, J. H. (2006). Evolution of competence and DNA uptake specificity in the *Pasteurellaceae*. *BMC Evol Biol*, 6: 82. doi: 10.1186/1471-2148-6-82.

Reznik, G. K. (1990). Comparative anatomy, physiology, and function of the upper respiratory tract. *Environ Health Perspect*, 85: 171-6.

Rodriguez, C. A., Avadhanula, V., Buscher, A., Smith, A. L., St Geme, J. W., 3rd & Adderson, E. E. (2003). Prevalence and distribution of adhesins in invasive non-type b encapsulated *Haemophilus influenzae*. *Infect Immun*, 71 (4): 1635-42.

Roman, F., Canton, R., Perez-Vazquez, M., Baquero, F. & Campos, J. (2004). Dynamics of long-term colonization of respiratory tract by *Haemophilus influenzae* in cystic fibrosis patients shows a marked increase in hypermutable strains. *J Clin Microbiol*, 42 (4): 1450-9.

Rosenberg, E. & Zilber-Rosenberg, I. (2016). Microbes drive evolution of animals and plants: the hologenome concept. *MBio*, 7 (2): e01395. doi: 10.1128/mBio.01395-15.

Rothenbacher, F. P. & Zhu, J. (2014). Efficient responses to host and bacterial signals during *Vibrio cholerae* colonization. *Gut Microbes*, 5 (1): 120-8. doi: 10.4161/gmic.26944.

Saito, M., Umeda, A. & Yoshida, S. (1999). Subtyping of *Haemophilus influenzae* strains by pulsed-field gel electrophoresis. *J Clin Microbiol*, 37 (7): 2142-7.

Saladin, K. (2003). *Anatomy & Physiology: The Unity of Form and Function, Third Edition*.

Salama, N. R., Hartung, M. L. & Muller, A. (2013). Life in the human stomach: persistence strategies of the bacterial pathogen *Helicobacter pylori*. *Nat Rev Microbiol*, 11 (6): 385-99. doi: 10.1038/nrmicro3016.

Salih, B. A. (2009). *Helicobacter pylori* infection in developing countries: the burden for how long? *Saudi J Gastroenterol*, 15 (3): 201-7. doi: 10.4103/1319-3767.54743.

Sanchez-Buso, L., Comas, I., Jorques, G. & Gonzalez-Candelas, F. (2014). Recombination drives genome evolution in outbreak-related *Legionella pneumophila* isolates. *Nat Genet*, 46 (11): 1205-11. doi: 10.1038/ng.3114.

Sandegren, L., Groenheit, R., Koivula, T., Ghebremichael, S., Advani, A., Castro, E., Pennhag, A., Hoffner, S., Mazurek, J., Pawlowski, A., et al. (2011). Genomic stability over 9 years of an isoniazid

resistant *Mycobacterium tuberculosis* outbreak strain in Sweden. *PLoS One*, 6 (1): e16647. doi: 10.1371/journal.pone.0016647.

Sano, H., Chiba, H., Iwaki, D., Sohma, H., Voelker, D. R. & Kuroki, Y. (2000). Surfactant proteins A and D bind CD14 by different mechanisms. *J Biol Chem*, 275 (29): 22442-51. doi: 10.1074/jbc.M001107200.

San Millan, A., Garcia-Cobos, S., Escudero, J. A., Hidalgo, L., Gutierrez, B., Carrilero, L., Campos, J. & Gonzalez-Zorn, B. (2010). *Haemophilus influenzae* clinical isolates with plasmid pB1000 bearing bla_{ROB-1}: fitness cost and interspecies dissemination. *Antimicrob Agents Chemother*, 54 (4): 1506-11. doi: 10.1128/AAC.01489-09.

Saunders, N. J., Trivedi, U. H., Thomson, M. L., Doig, C., Laurenson, I. F. & Blaxter, M. L. (2011). Deep resequencing of serial sputum isolates of *Mycobacterium tuberculosis* during therapeutic failure due to poor compliance reveals stepwise mutation of key resistance genes on an otherwise stable genetic background. *J Infect*, 62 (3): 212-7. doi: 10.1016/j.jinf.2011.01.003.

Schable, B., Villarino, M. E., Favero, M. S. & Miller, J. M. (1991). Application of multilocus enzyme electrophoresis to epidemiologic investigations of *Xanthomonas maltophilia*. *Infect Control Hosp Epidemiol*, 12 (3): 163-7.

Schulz zur Wiesch, P., Engelstadter, J. & Bonhoeffer, S. (2010). Compensation of fitness costs and reversibility of antibiotic resistance mutations. *Antimicrob Agents Chemother*, 54 (5): 2085-95. doi: 10.1128/AAC.01460-09.

Scully, L. R. & Bidochka, M. J. (2006). Developing insect models for the study of current and emerging human pathogens. *FEMS Microbiol Lett*, 263 (1): 1-9. doi: 10.1111/j.1574-6968.2006.00388.x.

Segales, J., Domingo, M., Solano, G. I. & Pijoan, C. (1997). Immunohistochemical detection of *Haemophilus parasuis* serovar 5 in formalin-fixed, paraffin-embedded tissues of experimentally infected swine. *J Vet Diagn Invest*, 9 (3): 237-43. doi: 10.1177/104063879700900303.

Senghore, M., Otu, J., Witney, A., Gehre, F., Doughty, E. L., Kay, G. L., Butcher, P., Salako, K., Kehinde, A., Onyejepu, N., et al. (2017). Whole-genome sequencing illuminates the evolution and spread of multidrug-resistant tuberculosis in Southwest Nigeria. *PLoS One*, 12 (9): e0184510. doi: 10.1371/journal.pone.0184510.

Sethi, S. (2000). Bacterial infection and the pathogenesis of COPD. *Chest*, 117 (5 Suppl 1): 286S-91S.

Sethi, S., Evans, N., Grant, B. J. & Murphy, T. F. (2002). New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. *N Engl J Med*, 347 (7): 465-71. doi: 10.1056/NEJMoa012561.

Sethi, S. (2004). Bacteria in exacerbations of chronic obstructive pulmonary disease: phenomenon or epiphenomenon? *Proc Am Thorac Soc*, 1 (2): 109-14. doi: 10.1513/pats.2306029.

Sethi, S. & Murphy, T. F. (2008). Infection in the pathogenesis and course of chronic obstructive pulmonary disease. *N Engl J Med*, 359 (22): 2355-65. doi: 10.1056/NEJMra0800353.

Sethi, S., Mallia, P. & Johnston, S. L. (2009). New paradigms in the pathogenesis of chronic obstructive pulmonary disease II. *Proc Am Thorac Soc*, 6 (6): 532-4. doi: 10.1513/pats.200905-025DS.

Shukla, S. D., Budden, K. F., Neal, R. & Hansbro, P. M. (2017). Microbiome effects on immunity, health and disease in the lung. *Clin Transl Immunology*, 6 (3): e133. doi: 10.1038/cti.2017.6.

Siegel, S. J. & Weiser, J. N. (2015). Mechanisms of bacterial colonization of the respiratory tract. *Annual review of microbiology*, 69: 425-444. doi: 10.1146/annurev-micro-091014-104209.

Silva, G. E., Sherrill, D. L., Guerra, S. & Barbee, R. A. (2004). Asthma as a risk factor for COPD in a longitudinal study. *Chest*, 126 (1): 59-65. doi: 10.1378/chest.126.1.59.

Silverman, E. K. & Speizer, F. E. (1996). Risk factors for the development of chronic obstructive pulmonary disease. *Med Clin North Am*, 80 (3): 501-22.

Sinha, S., Mell, J. C. & Redfield, R. J. (2012). Seventeen Sxy-dependent cyclic AMP receptor protein site-regulated genes are needed for natural transformation in *Haemophilus influenzae*. *J Bacteriol*, 194 (19): 5245-54. doi: 10.1128/JB.00671-12.

Siqueira, F. M., Perez-Wohlfeil, E., Carvalho, F. M., Trelles, O., Schrank, I. S., Vasconcelos, A. T. R. & Zaha, A. (2017). Microbiome overview in swine lungs. *PLoS One*, 12 (7): e0181503. doi:

10.1371/journal.pone.0181503.

Smith, E. E., Buckley, D. G., Wu, Z., Saenphimmachak, C., Hoffman, L. R., D'Argenio, D. A., Miller, S. I., Ramsey, B. W., Speert, D. P., Moskowitz, S. M., et al. (2006). Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc Natl Acad Sci U S A*, 103 (22): 8487-92. doi: 10.1073/pnas.0602138103.

Smith, H. O., Gwinn, M. L. & Salzberg, S. L. (1999). DNA uptake signal sequences in naturally transformable bacteria. *Res Microbiol*, 150 (9-10): 603-16.

Smith-Vaughan, H. C., Sriprakash, K. S., Leach, A. J., Mathews, J. D. & Kemp, D. J. (1998). Low genetic diversity of *Haemophilus influenzae* type b compared to nonencapsulated *H. influenzae* in a population in which *H. influenzae* is highly endemic. *Infection and Immunity*, 66 (7): 3403-3409.

Sokurenko, E. V., Hasty, D. L. & Dykhuizen, D. E. (1999). Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol*, 7 (5): 191-5.

Sommer, L. M., Alanin, M. C., Marvig, R. L., Nielsen, K. G., Hoiby, N., von Buchwald, C., Molin, S. & Johansen, H. K. (2016). Bacterial evolution in PCD and CF patients follows the same mutational steps. *Sci Rep*, 6: 28732. doi: 10.1038/srep28732.

St Geme, J. W. (1994). The HMW1 adhesin of nontypeable *Haemophilus influenzae* recognizes sialylated glycoprotein receptors on cultured human epithelial cells. *Infection and Immunity*, 62 (9): 3881-3889.

St Geme, J. W., 3rd & Falkow, S. (1990). *Haemophilus influenzae* adheres to and enters cultured human epithelial cells. *Infect Immun*, 58 (12): 4036-44.

St Geme, J. W., 3rd, Falkow, S. & Barenkamp, S. J. (1993). High-molecular-weight proteins of nontypeable *Haemophilus influenzae* mediate attachment to human epithelial cells. *Proc Natl Acad Sci U S A*, 90 (7): 2875-9.

St Geme, J. W., 3rd & Grass, S. (1998). Secretion of the *Haemophilus influenzae* HMW1 and HMW2 adhesins involves a periplasmic intermediate and requires the HMWB and HMWC proteins. *Mol Microbiol*, 27 (3): 617-30.

St Geme, J. W., 3rd, Kumar, V. V., Cutter, D. & Barenkamp, S. J. (1998). Prevalence and distribution of the *hmw* and *hia* genes and the HMW and Hia adhesins among genetically diverse strains of nontypeable *Haemophilus influenzae*. *Infect Immun*, 66 (1): 364-8.

St Geme, J. W., 3rd & Yeo, H. J. (2009). A prototype two-partner secretion pathway: the *Haemophilus influenzae* HMW1 and HMW2 adhesin systems. *Trends Microbiol*, 17 (8): 355-60. doi: 10.1016/j.tim.2009.06.002.

Staples, M., Graham, R. M. A. & Jennison, A. V. (2017). Characterisation of invasive clinical *Haemophilus influenzae* isolates in Queensland, Australia using whole-genome sequencing. *Epidemiol Infect*, 145 (8): 1727-1736. doi: 10.1017/S0950268817000450.

Starner, T. D., Swords, W. E., Apicella, M. A. & McCray, P. B., Jr. (2002). Susceptibility of nontypeable *Haemophilus influenzae* to human β-defensins is influenced by lipooligosaccharide acylation. *Infect Immun*, 70 (9): 5287-9.

Starner, T. D., Zhang, N., Kim, G., Apicella, M. A. & McCray, P. B., Jr. (2006). *Haemophilus influenzae* forms biofilms on airway epithelia: implications in cystic fibrosis. *Am J Respir Crit Care Med*, 174 (2): 213-20. doi: 10.1164/rccm.200509-1459OC.

Stoller, J. K., Lacbawan, F. L. & Aboussouan, L. S. (1993). Alpha-1 antitrypsin deficiency. I: Adam, M. P., Ardinger, H. H., Pagon, R. A., Wallace, S. E., Bean, L. J. H., Mefford, H. C., Stephens, K., Amemiya, A. & Ledbetter, N. (red.) *GeneReviews(R)*. Seattle (WA).

Stoller, J. K. & Aboussouan, L. S. (2005). α1-antitrypsin deficiency. *Lancet*, 365 (9478): 2225-36. doi: 10.1016/S0140-6736(05)66781-5.

Sun, G., Luo, T., Yang, C., Dong, X., Li, J., Zhu, Y., Zheng, H., Tian, W., Wang, S., Barry, C. E., 3rd, et al. (2012). Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. *J Infect Dis*, 206 (11): 1724-33. doi: 10.1093/infdis/jis601.

Swanson, M. S. & Hammer, B. K. (2000). *Legionella pneumophila* pathogenesis: a fateful journey from amoebae to macrophages. *Annu Rev Microbiol*, 54: 567-613. doi: 10.1146/annurev.micro.54.1.567.

Swords, W. E., Buscher, B. A., Ver Steeg Ii, K., Preston, A., Nichols, W. A., Weiser, J. N., Gibson, B. W. & Apicella, M. A. (2000). Non-typeable *Haemophilus influenzae* adhere to and invade human bronchial epithelial cells via an interaction of lipooligosaccharide with the PAF receptor. *Mol Microbiol*, 37 (1): 13-27.

Sze, M. A., Dimitriu, P. A., Hayashi, S., Elliott, W. M., McDonough, J. E., Gosselink, J. V., Cooper, J., Sin, D. D., Mohn, W. W. & Hogg, J. C. (2012). The lung tissue microbiome in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*, 185 (10): 1073-80. doi: 10.1164/rccm.201111-2075OC.

Tam, A., Bates, J. H., Churg, A., Wright, J. L., Man, S. F. & Sin, D. D. (2016). Sex-related differences in pulmonary function following 6 months of cigarette exposure: implications for sexual dimorphism in mild COPD. *PLoS One*, 11 (10): e0164835. doi: 10.1371/journal.pone.0164835.

Tamas, I., Klasson, L., Canback, B., Naslund, A. K., Eriksson, A. S., Wernegreen, J. J., Sandstrom, J. P., Moran, N. A. & Andersson, S. G. (2002). 50 million years of genomic stasis in endosymbiotic bacteria. *Science*, 296 (5577): 2376-9. doi: 10.1126/science.1071278.

Tchoupa, A. K., Lichtenegger, S., Reidl, J. & Hauck, C. R. (2015). Outer membrane protein P1 is the CEACAM-binding adhesin of *Haemophilus influenzae*. *Mol Microbiol*, 98 (3): 440-55. doi: 10.1111/mmi.13134.

Thi, E. P., Lambertz, U. & Reiner, N. E. (2012). Sleeping with the enemy: how intracellular pathogens cope with a macrophage lifestyle. *PLoS Pathog*, 8 (3): e1002551. doi: 10.1371/journal.ppat.1002551.

Thorell, K., Yahara, K., Berthenet, E., Lawson, D. J., Mikhail, J., Kato, I., Mendez, A., Rizzato, C., Bravo, M. M., Suzuki, R., et al. (2017). Rapid evolution of distinct *Helicobacter pylori* subpopulations in the Americas. *PLoS Genet*, 13 (2): e1006546. doi: 10.1371/journal.pgen.1006546.

Tudor-Williams, G., Frankland, J., Isaacs, D., Mayon-White, R. T., MacFarlane, J. A., Slack, M. P., Anderson, E., Rees, D. G. & Moxon, E. R. (1989). *Haemophilus influenzae* type b disease in the Oxford region. *Archives of Disease in Childhood*, 64 (4): 517-519.

Valeri, M., Rossi Paccani, S., Kasendra, M., Nesta, B., Serino, L., Pizza, M. & Soriani, M. (2015). Pathogenic *E. coli* exploits *SslE* mucinase activity to translocate through the mucosal barrier and get access to host cells. *PLoS One*, 10 (3): e0117486. doi: 10.1371/journal.pone.0117486.

van Belkum, A., Scherer, S., van Alphen, L. & Verbrugh, H. (1998). Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev*, 62 (2): 275-93.

van der Woude, M. W. & Baumler, A. J. (2004). Phase and antigenic variation in bacteria. *Clin Microbiol Rev*, 17 (3): 581-611, table of contents. doi: 10.1128/CMR.17.3.581-611.2004.

Van Durme, Y. M., Eijgelsheim, M., Joos, G. F., Hofman, A., Uitterlinden, A. G., Brusselle, G. G. & Stricker, B. H. (2010). Hedgehog-interacting protein is a COPD susceptibility gene: the Rotterdam Study. *Eur Respir J*, 36 (1): 89-95. doi: 10.1183/09031936.00129509.

van Ham, S. M., van Alphen, L., Mooi, F. R. & van Putten, J. P. (1993). Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. *Cell*, 73 (6): 1187-96.

van Mansfeld, R., de Been, M., Paganelli, F., Yang, L., Bonten, M. & Willems, R. (2016). Within-host evolution of the dutch high-prevalent *Pseudomonas aeruginosa* clone ST406 during chronic colonization of a patient with cystic fibrosis. *PLoS One*, 11 (6): e0158106. doi: 10.1371/journal.pone.0158106.

van Schilfgaarde, M., Eijk, P., Regelink, A., van Ulsen, P., Everts, V., Dankert, J. & van Alphen, L. (1999). *Haemophilus influenzae* localized in epithelial cell layers is shielded from antibiotics and antibody-mediated bactericidal activity. *Microb Pathog*, 26 (5): 249-62. doi: 10.1006/mpat.1998.0269.

van Schilfgaarde, M., van Ulsen, P., Eijk, P., Brand, M., Stam, M., Kouame, J., van Alphen, L. & Dankert, J. (2000). Characterization of adherence of nontypeable *Haemophilus influenzae* to human epithelial cells. *Infect Immun*, 68 (8): 4658-65.

Verra, F., Escudier, E., Lebargy, F., Bernaudin, J. F., De Cremoux, H. & Bignon, J. (1995). Ciliary abnormalities in bronchial epithelium of smokers, ex-smokers, and nonsmokers. *Am J Respir Crit Care Med*, 151 (3 Pt 1): 630-4. doi: 10.1164/ajrccm/151.3_Pt_1.630.

Vestbo, J., Hurd, S. S., Agusti, A. G., Jones, P. W., Vogelmeier, C., Anzueto, A., Barnes, P. J., Fabbri,

L. M., Martinez, F. J., Nishimura, M., et al. (2013). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*, 187 (4): 347-65. doi: 10.1164/rccm.201204-0596PP.

Vidigal, P. G., Dittmer, S., Steinmann, E., Buer, J., Rath, P. M. & Steinmann, J. (2014). Adaptation of *Stenotrophomonas maltophilia* in cystic fibrosis: molecular diversity, mutation frequency and antibiotic resistance. *Int J Med Microbiol*, 304 (5-6): 613-9. doi: 10.1016/j.ijmm.2014.04.002.

Vignola, A. M., Paganin, F., Capieu, L., Scichilone, N., Bellia, M., Maakel, L., Bellia, V., Godard, P., Bousquet, J. & Chanez, P. (2004). Airway remodelling assessed by sputum and high-resolution computed tomography in asthma and COPD. *Eur Respir J*, 24 (6): 910-7. doi: 10.1183/09031936.04.00032603.

Vinella, D., Fischer, F., Vorontsov, E., Gallaud, J., Malosse, C., Michel, V., Cavazza, C., Robbe-Saule, M., Richaud, P., Chamot-Rooke, J., et al. (2015). Evolution of *Helicobacter*: Acquisition by gastric species of two histidine-rich proteins essential for colonization. *PLoS Pathog*, 11 (12): e1005312. doi: 10.1371/journal.ppat.1005312.

Vogan, A. A. & Higgs, P. G. (2011). The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. *Biol Direct*, 6: 1. doi: 10.1186/1745-6150-6-1.

Vogel, A. R., Szelestey, B. R., Raffel, F. K., Sharpe, S. W., Gearinger, R. L., Justice, S. S. & Mason, K. M. (2012). SapF-mediated heme-iron utilization enhances persistence and coordinates biofilm architecture of *Haemophilus*. *Front Cell Infect Microbiol*, 2: 42. doi: 10.3389/fcimb.2012.00042.

von Eiff, C., Becker, K., Machka, K., Stammer, H. & Peters, G. (2001). Nasal carriage as a source of *Staphylococcus aureus* bacteremia. Study Group. *N Engl J Med*, 344 (1): 11-6. doi: 10.1056/NEJM200101043440102.

Wang, C. C., Siu, L. K., Chen, M. K., Yu, Y. L., Lin, F. M., Ho, M. & Chu, M. L. (2001). Use of automated riboprinter and pulsed-field gel electrophoresis for epidemiological studies of invasive *Haemophilus influenzae* in Taiwan. *J Med Microbiol*, 50 (3): 277-83. doi: 10.1099/0022-1317-50-3-277.

Waterfield, N. R., Wren, B. W. & Ffrench-Constant, R. H. (2004). Invertebrates as a source of emerging human pathogens. *Nat Rev Microbiol*, 2 (10): 833-41. doi: 10.1038/nrmicro1008.

Waters, V., Yau, Y., Prasad, S., Lu, A., Atenafu, E., Crandall, I., Tom, S., Tullis, E. & Ratjen, F. (2011). *Stenotrophomonas maltophilia* in cystic fibrosis: serologic response and effect on lung disease. *Am J Respir Crit Care Med*, 183 (5): 635-40. doi: 10.1164/rccm.201009-1392OC.

Waters, V., Atenafu, E. G., Lu, A., Yau, Y., Tullis, E. & Ratjen, F. (2013). Chronic *Stenotrophomonas maltophilia* infection and mortality or lung transplantation in cystic fibrosis patients. *J Cyst Fibros*, 12 (5): 482-6. doi: 10.1016/j.jcf.2012.12.006.

Watson, M. E., Jr., Burns, J. L. & Smith, A. L. (2004). Hypermutable *Haemophilus influenzae* with mutations in *mutS* are found in cystic fibrosis sputum. *Microbiology*, 150 (Pt 9): 2947-58. doi: 10.1099/mic.0.27230-0.

Webster, P., Wu, S., Gomez, G., Apicella, M., Plaut, A. G. & St Geme, J. W., 3rd. (2006). Distribution of bacterial proteins in biofilms formed by non-typeable *Haemophilus influenzae*. *J Histochem Cytochem*, 54 (7): 829-42. doi: 10.1369/jhc.6A6922.2006.

Wedzicha, J. A. & Donaldson, G. C. (2003). Exacerbations of chronic obstructive pulmonary disease. *Respir Care*, 48 (12): 1204-13; discussion 1213-5.

Weiser, J. N. (2000). The generation of diversity by *Haemophilus influenzae*. *Trends Microbiol*, 8 (10): 433-5.

WHO. (2017). Fact sheet on Chronic Obstructive Pulmonary Disease.

Williams. (1957). Pleiotropy, natural selection, and the evolution of senescence. *International journal of organic evolution*, 11 (4): 398-411.

Wong, A., Rodrigue, N. & Kassen, R. (2012). Genomics of adaptation during experimental evolution of the opportunistic pathogen *Pseudomonas aeruginosa*. *PLoS Genet*, 8 (9): e1002928. doi: 10.1371/journal.pgen.1002928.

Wright, J. R. (2003). Pulmonary surfactant: a front line of lung host defense. *J Clin Invest*, 111 (10): 1453-5. doi: 10.1172/JCI18650.

Wu, S., Baum, M. M., Kerwin, J., Guerrero, D., Webster, S., Schaudinn, C., VanderVelde, D. & Webster, P. (2014). Biofilm-specific extracellular matrix proteins of nontypeable *Haemophilus influenzae*. *Pathog Dis*, 72 (3): 143-60. doi: 10.1111/2049-632X.12195.

Yamada, K., Asai, K., Nagayasu, F., Sato, K., Ijiri, N., Yoshii, N., Imahashi, Y., Watanabe, T., Tochino, Y., Kanazawa, H., et al. (2016). Impaired nuclear factor erythroid 2-related factor 2 expression increases apoptosis of airway epithelial cells in patients with chronic obstructive pulmonary disease due to cigarette smoking. *BMC Pulm Med*, 16: 27. doi: 10.1186/s12890-016-0189-1.

Yang, L., Jelsbak, L. & Molin, S. (2011). Microbial ecology and adaptation in cystic fibrosis airways. *Environ Microbiol*, 13 (7): 1682-9. doi: 10.1111/j.1462-2920.2011.02459.x.

Yin, W., Li, H., Shen, Y., Liu, Z., Wang, S., Shen, Z., Zhang, R., Walsh, T. R., Shen, J. & Wang, Y. (2017). Novel Plasmid-mediated colistin resistance gene *mcr-3* in *Escherichia coli*. *MBio*, 8 (3). doi: 10.1128/mBio.00543-17.

Yona, A. H., Frumkin, I. & Pilpel, Y. (2015). A relay race on the evolutionary adaptation spectrum. *Cell*, 163 (3): 549-59. doi: 10.1016/j.cell.2015.10.005.

Young, B. C., Golubchik, T., Batty, E. M., Fung, R., Larner-Svensson, H., Votintseva, A. A., Miller, R. R., Godwin, H., Knox, K., Everitt, R. G., et al. (2012). Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. *Proc Natl Acad Sci U S A*, 109 (12): 4550-5. doi: 10.1073/pnas.1113219109.

Young, D., Hussell, T. & Dougan, G. (2002). Chronic bacterial infections: living with unwanted guests. *Nat Immunol*, 3 (11): 1026-32. doi: 10.1038/ni1102-1026.

Young, B. C., Wu, C. H., Gordon, N. C., Cole, K., Price, J. R., Liu, E., Sheppard, A. E., Perera, S., Charlesworth, J., Golubchik, T., et al. (2017). Severe infections emerge from commensal bacteria by adaptive evolution. *Elife*, 6. doi: 10.7554/eLife.30637.

Yung, P. Y., Grasso, L. L., Mohidin, A. F., Acerbi, E., Hinks, J., Seviour, T., Marsili, E. & Lauro, F. M. (2016). Global transcriptomic responses of *Escherichia coli* K-12 to volatile organic compounds. *Sci Rep*, 6: 19899. doi: 10.1038/srep19899.

Zdziarski, J., Svanborg, C., Wullt, B., Hacker, J. & Dobrindt, U. (2008). Molecular basis of commensalism in the urinary tract: low virulence or virulence attenuation? *Infect Immun*, 76 (2): 695-703. doi: 10.1128/IAI.01215-07.

Zhang, B., Tang, C., Liao, M. & Yue, H. (2014). Update on the pathogenesis of *Haemophilus parasuis* infection and virulence factors. *Vet Microbiol*, 168 (1): 1-7. doi: 10.1016/j.vetmic.2013.07.027.

HYPOTHESIS & OBJECTIVES

Following the requirements imposed by the Escuela de Doctorado de Navarra (EDONA) for the deposit of a PhD Thesis document following a “collection of publications” format at the Universidad Pública de Navarra (UPNa), I next summarize the Hypotheses and Objectives of my research during this PhD Thesis work, and my contributions to each Results Chapter.

All the Methods employed during my PhD Thesis work are detailed in each Results Chapter.

A general discussion for the entire content of this PhD Thesis work, and a list of specific conclusions, can be found in the last two sections of this document.

HYPOTHESIS:

Haemophilus spp. are host-restricted commensal bacteria residing in the upper airways of healthy individuals that, under adequate circumstances, reach the lower airways of immune-compromised subjects causing respiratory infection and persistent colonization. Antibiotic administration is a selective pressure driving bacterial adaptive evolution and natural selection of resistance determinants. The effects of these interventions within colonizing *Haemophilus* spp. bacteria remain unknown. Based on this notion, we state the following **hypothesis: "Commensal bacteria with potential to behave as colonizing opportunistic pathogens can be reservoirs of transmissible antibiotic resistance genes"**.

Epithelial invasion is a well-described mechanism of NTHi respiratory infection, and NTHi is a natural competent bacterium that actively takes and incorporates exogenous DNA in its chromosome through homologous recombination. Likewise, experimental whole genome evolution studies allow analyzing selective pressures and identifying genetic traits responsible for phenotypic adaptation to such pressures by using relatively simple study model systems. Based on these evidences, we state the following **hypothesis: "Co-occurrence and infection of diverse NTHi strains within the same niche leads to interchange of beneficial genetic traits adapted to the phenotype under study".**

NTHi causes persistent infection within the lung of COPD patients, being a colonizing opportunistic pathogen frequently isolated in COPD sputum samples. Within-host natural patho-adaptive evolution has mostly focused on opportunistic pathogens of environmental origin. However, the role of genome evolutionary changes in host-restricted members of the human microbiome causing opportunistic infection is mostly unknown. Moreover, evolutionary studies comprehensively identify genomic changes, but further experimental assessment of the biological significance of such genomic variation is frequently limited. Based on these evidences, we state the following **hypothesis: "NTHi undergoes patho-adaptation within the lower airways of COPD patients through genomic mechanisms overcoming host selective pressures, and likely facilitating long-term persistent infection"**.

OBJECTIVES:

The proposed hypotheses have been addressed through the following **Objectives**:

Objective 1. Use of *Haemophilus parasuis* as a model system for a host-restricted colonizing opportunistic pathogen, to screen the prevalence of antibiotic resistance in piglets as healthy carrier population, identify and characterize genetic determinants accounting for such resistance.

Objective 2. Use of *Haemophilus influenzae* as a model system for a naturally transformable bacterial pathogen to develop a genetic screening method based on gain-of-function, and its application to the search of *H. influenzae* invasin encoding genes.

Objective 3. Analysis of the patho-adaptive evolution by the human-restricted colonizing opportunistic pathogen *Haemophilus influenzae* during chronic lung infection within patients suffering Chronic Obstructive Pulmonary Disease (COPD).

The **Results** obtained from the proposed Objectives are presented in three **Chapters**:

Chapter 1. Novel *bla*$_{ROB-1}$-bearing plasmid conferring resistance to β-lactams in *Haemophilus parasuis* isolates from healthy weaning pigs.

Chapter 2. Transformed recombinant enrichment profiling rapidly identifies HMW1 as an intracellular invasion locus in *Haemophilus influenzae*.

Chapter 3. Antagonistic pleiotropy in a bifunctional fatty acid transporter during bacterial adaptation to chronic lung infection.

Javier Moleres (**JM**) contribution to this research work:

Chapter 1:

- Conceived and designed the experiments: JG, VA, BGZ
- Performed the experiments: **JM**, ASL
- Analyzed the data: JG, BGZ, VA, **JM**, ASL, VA

Experimental work performed by JM:

- Participation in nasal sampling of healthy piglets
- Isolation of *Haemophilus parasuis* strains from nasal swab samples
- Identification of *H. parasuis* by biochemical and molecular approaches
- Participation in molecular typing of *H. parasuis* isolates by ERIC-PCR
- Detection of the *lsgB* and group 1 *vtaA* genes in the *H. parasuis* strain collection generated
- Determination of extra-chromosomal genetic elements in *H. parasuis* β-lactam-resistant strains
- Antimicrobial susceptibility testing, adapted to *H. parasuis*
- Biofilm formation assays: setting up, quantification and determination of their matrix nature
- Culture of swine epithelial cells (PK-15 cells), infection by *H. parasuis* clinical isolates, quantification of IL-8 secretion by swine cells upon *H. parasuis* infection
- Fully sequencing of pJMA-1
- Stability assays for pJMA-1 in heterologous hosts

Chapter 2:

- Conceived and designed the experiments: JCM, JG, RJR
- Performed the experiments: JCM, CV, JG, SS, **JM,** AFC, EAP
- Analyzed the data: JCM, CV, JG, **JM,** AFC, EAP, SS, RJR, JWSG, CN

Experimental work performed by JM:

- Culture of human epithelial cells (A549 cells), infection by *H. influenzae* isolates.
- Quantification of adhesion and intracelular invasion frequencies for recombinant clones and strains engineered in the laboratory for TREP validation
- Immunodetection of HMW proteins
- Bacterial self-aggregation setting up and determination
- Immunofluorescence microscopy on NTHi epithelial infected cells
- PCR validation assays for all strains engineered in the laboratory for TREP validation

Chapter 3:

- Conceived and designed the experiments: JCM, JG, **JM**, AFC, JL, SMS,
- Performed the experiments: **JM,** AFC, SM, IRA, LPR, SB
- Analyzed the data: JCM, JG, **JM,** AFC, RLE, SM

Experimental work performed by JM:

- Handling the entire NTHi COPD strain collection generated at HUB, involving strain growth and preservation, DNA extraction and preparation for sequencing. Illumina and PacBio WGS.
- Variant calling
- Antimicrobial susceptibility testing. Assignment of genotypic traits related to intra-clonal variation on antimicrobial resistance
- Deep analysis of FadL allelic variation, sequence manual retrieval; complete and truncated variants detailed organization
- Analysis of FadL variation at the protein level (visualization on SDS-PAGE gels)
- Analysis of the *fadL* gene expression
- Bacterial infection assays: culture of human epithelial cells (HeLa and HeLa-BGP), infection by COPD isolates
- Free fatty acid susceptibility testing by COPD isolates
- Bacterial mouse lung infection
- Determination of phase variation at the *hmw1A* promoter region. Sanger sequencing of the *hmw1A* gene in clonal strains P651, P653, P654 and their and predicted promoter regions

RESULTS

Chapter 1. Novel *bla*$_{ROB-1}$-bearing plasmid conferring resistance to β-lactams in *Haemophilus parasuis* isolates from healthy weaning pigs.

Moleres J, Santos-López A, Lázaro I, Labairu J, Prat C, Ardanuy C, González-Zorn B, Aragon V, Garmendia J. Appl Environ Microbiol. 2015 May 1;81(9):3255-67. doi: 10.1128/AEM.03865-14. Epub 2015 Mar 6.

AEM
Journals.ASM.org

Novel bla_{ROB-1}-Bearing Plasmid Conferring Resistance to β-Lactams in *Haemophilus parasuis* Isolates from Healthy Weaning Pigs

Javier Moleres,[a] Alfonso Santos-López,[b] Isidro Lázaro,[c] Javier Labairu,[c] Cristina Prat,[d,e] Carmen Ardanuy,[e,f] Bruno González-Zorn,[b] Virginia Aragon,[g] Junkal Garmendia[a,e]

Instituto de Agrobiotecnología, CSIC-Universidad Pública de Navarra-Gobierno de Navarra, Mutilva, Spain[a]; Departamento de Sanidad Animal, Facultad de Veterinaria y VISAVET, Universidad Complutense, Madrid, Spain[b]; Instituto Navarro de Tecnologías e Infraestructuras Agroalimentarias-INTIA, Navarra, Spain[c]; Hospital Universitari Germans Trias i Pujol, Badalona, Spain[d]; Centro de Investigación Biomédica en Red Enfermedades Respiratorias (CIBERES), Madrid, Spain[e]; Hospital Universitari Bellvitge, Barcelona, Spain[f]; Centre de Recerca en Sanitat Animal (CReSA), Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Campus de la Universitat Autònoma de Barcelona, Bellaterra, Spain[g]

Haemophilus parasuis*, the causative agent of Glässer's disease, is one of the early colonizers of the nasal mucosa of piglets. It is prevalent in swine herds, and lesions associated with disease are fibrinous polyserositis and bronchopneumonia. Antibiotics are commonly used in disease control, and resistance to several antibiotics has been described in *H. parasuis*. Prediction of *H. parasuis* virulence is currently limited by our scarce understanding of its pathogenicity. Some genes have been associated with *H. parasuis* virulence, such as *lsgB* and group 1 *vtaA*, while biofilm growth has been associated with nonvirulent strains. In this study, 86 *H. parasuis* nasal isolates from farms that had not had a case of disease for more than 10 years were obtained by sampling piglets at weaning. Isolates were studied by enterobacterial repetitive intergenic consensus PCR and determination of the presence of *lsgB* and group 1 *vtaA*, biofilm formation, inflammatory cell response, and resistance to antibiotics. As part of the diversity encountered, a novel 2,661-bp plasmid, named pJMA-1, bearing the bla_{ROB-1} β-lactamase was detected in eight colonizing strains. pJMA-1 was shown to share a backbone with other small plasmids described in the *Pasteurellaceae*, to be 100% stable, and to have a lower biological cost than the previously described plasmid pB1000. pJMA-1 was also found in nine *H. parasuis* nasal strains from a separate collection, but it was not detected in isolates from the lesions of animals with Glässer's disease or in nontypeable *Haemophilus influenzae* isolates. Altogether, we show that commensal *H. parasuis* isolates represent a reservoir of β-lactam resistance genes which can be transferred to pathogens or other bacteria.

Haemophilus parasuis is an early colonizer and a member of the normal microbiota of the upper respiratory tract of piglets. *H. parasuis* initial acquisition occurs through direct contact with the sow after birth, and the bacterium establishes colonization in the upper respiratory tract, with a maximum level of colonization occurring at about 2 months of age (1). Under certain circumstances, some strains spread to the lungs to cause pneumonia or invade systemic sites. Systemic invasion produces fibrinous polyserositis and arthritis, which are the characteristic lesions of Glässer's disease (2, 3).

Although *H. parasuis* is an important swine pathogen, its host-pathogen interactions remain to be well understood. Different *H. parasuis* strains can be isolated from the nasal cavity of a given animal, and these nasal strains are not particularly stable, since they experience turnover during the life of the pigs (1, 4–6). Strains of *H. parasuis* are heterogeneous and include virulent and nonvirulent strains. It is common to find nonvirulent strains in the upper respiratory tract of healthy animals, but virulent strains can also be found (6). Different methods have been developed to differentiate *H. parasuis* strains. Genotyping methods include multilocus sequence typing (MLST) (7, 8), partial sequence of the 60-kDa heat shock protein-encoding gene *hsp60* (9), enterobacterial repetitive intergenic consensus PCR (ERIC-PCR) (10), or pulsed-field gel electrophoresis (11). A search for genetic markers to identify putative virulent *H. parasuis* isolates revealed the potential of the virulence-associated trimeric autotransporter-encoding (*vtaA*) genes to be a diagnostic tool (12) and suggested an association of the *lsgB* gene, which encodes a sialyltransferase involved in the sialylation of the lipooligosaccharide (LOS) molecule, with virulent strains (13). A number of other bacterial factors have been associated with *H. parasuis* virulence, including 6-phosphogluconate dehydrogenase (14), a complete LOS molecule (15), cytolethal distending toxin (16, 17), immunoglobulin A protease (18), capsule (19), and the outer membrane protein OmpP2 (20–22).

Serovar diversity and the high number of nonserotypeable isolates that have been reported have negatively affected the development of effective cross-protective vaccines, due to limited cross protection among *H. parasuis* strains (23). Since no definite vaccine is available, antimicrobial treatment continues to be the strategy used to control the disease. Tetracyclines are the major antimicrobials used against this bacterium, but resistance has been

Received 26 November 2014 **Accepted** 25 February 2015

Accepted manuscript posted online 6 March 2015

Citation Moleres J, Santos-López A, Lázaro I, Labairu J, Prat C, Ardanuy C, González-Zorn B, Aragon V, Garmendia J. 2015. Novel bla_{ROB-1}-bearing plasmid conferring resistance to β-lactams in *Haemophilus parasuis* isolates from healthy weaning pigs. Appl Environ Microbiol 81:3255–3267. doi:10.1128/AEM.03865-14.

Editor: M. A. Elliot

Address correspondence to Virginia Aragon, Virginia.Aragon@creas.uab.cat, or Junkal Garmendia, juncal.garmendia@unavarra.es.

J.M. and A.S.-L. contributed equally to this article.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.03865-14.

doi:10.1128/AEM.03865-14

found in many instances (24–26), suggesting the need for more effective molecules to treat infected animals (27). Penicillins and aminopenicillins are being used as alternative treatments for infections due to *H. parasuis*. Of note, clinical isolates resistant to β-lactams have been found in large numbers in several countries (26, 27), and biofilm formation by *H. parasuis* strains may have a positive correlation with resistance to β-lactam antibiotics (28). β-Lactam resistance in *H. parasuis* clinical strains isolated from Glässer's disease lesions has been related to plasmid pB1000, which bears the ROB-1 β-lactamase, belongs to the ColE1 superfamily, is mobilized into *Escherichia coli* using the conjugation machinery of an IncP plasmid, and has also been found in *Pasteurella multocida* and *Haemophilus influenzae* clinical isolates (11, 29, 30). However, antibiotic resistance in the colonizer population of *H. parasuis* has not been previously examined, and its potential as a reservoir of antibiotic resistance genes has not been explored. Here, we describe the novel small plasmid pJMA-1, which was isolated from strains of *H. parasuis* found in the nasal cavities of healthy animals and which carries ROB-1-mediated β-lactam resistance. pJMA-1 sequence features and its relationship with previously described mobile elements, functionality, transmissibility, and stability were analyzed. The implications of antibiotic resistance carried by a plasmid in *H. parasuis* nasal strains are further discussed.

MATERIALS AND METHODS

Bacterial strains and culture conditions. *H. parasuis* strains were grown overnight at 37°C with 5% CO_2 on chocolate agar plates (bioMérieux, France). The *H. parasuis* strains belonged to three isolate collections: (i) 86 strains isolated from nasal swabs taken from healthy pigs at weaning in Navarra, Spain (generated in this study), (ii) 20 strains isolated from nasal swab samples taken from healthy pigs at weaning in Catalonia and Mallorca, Spain (8, 9, 31), and (iii) 24 strains isolated from Glässer's disease lesions (8, 9, 31). *H. parasuis* Nagasaki (virulent), SW114 (nonvirulent), BB1033 (10), and BB1023 (10) were used as reference strains, when necessary.

Nontypeable (NT) *Haemophilus influenzae* strains were grown overnight at 37°C with 5% CO_2 on chocolate agar plates or on brain heart infusion (BHI) agar plates supplemented with 10 μg/ml hemin and 10 μg/ml β-NAD, referred to as sBHI agar. We employed 20 NT *H. influenzae* isolates recovered from the nasopharynges of healthy children in day care centers and schools in Oviedo, Spain (32), 16 pediatric ear isolates (Hospital Universitario Germans Trias i Pujol [HUGTiP], Badalona, Spain) (this study), and 41 respiratory strains from pulmonary patients (HUGTiP and Hospital Universitario Bellvitge, Barcelona, Spain) (this study). *H. influenzae* Rd KW20 (33) was used as a reference strain for plasmid transmissibility and stability analysis. When necessary, sBHI agar containing ampicillin (AMP) at 25 μg/ml was used.

Isolation of *H. parasuis* strains from nasal swab samples from healthy piglets. From April to September 2012, nasal swab samples were taken at weaning (3 to 4 weeks of age) from healthy pigs on six farrow-to-finish farms and two farrow-to-weaning farms that had not had a case of Glässer's disease since 2000. The farms were located in an area of 9,500 km^2 in Navarra, northern Spain. The farms were included on the basis of the veterinarians' knowledge of the farm and were selected because of their different patterns regarding the respiratory health status of the pigs and because they used the most common herd health patterns, housing practices, and herd management practices encountered in the field at the time. The sampling procedure was part of a periodic veterinarian farm surveillance routine. The number of sows, production type, and weaning age for each farm are summarized in Table 1. The nares of four animals per farm were sampled. Swab (Deltalab, Spain) specimens from the nasal cavities of all piglets were placed in Amies transport medium and kept on ice until inoculation on chocolate agar plates. The plates were incubated at 37°C with 5% CO_2, and suspect colonies of *H. parasuis* were selected and subcultured on chocolate agar and blood agar (bioMérieux) plates. Colonies growing on chocolate agar but not on blood agar were identified by conventional biochemical methods, by mass spectrometry using a matrix-assisted laser desorption ionization biotyper (version 3.0; Bruker), and by colony PCR based on species-specific amplification of the 16S RNA gene with primers HPS-F and HPS-R (Table 2), rendering an 821-bp product (34). For species-specific PCR, crude DNA templates derived directly from colonies on agar were prepared by harvesting a loopful of bacteria from a culture that had been grown overnight on a chocolate agar plate and placing it into 0.5 ml of sterile water of the purity required for DNA analysis; Tween 20 was added to a final concentration of 0.5%, and three cycles of boiling (for 5 min) and cooling on ice (for 2 min) were performed. The samples were centrifuged for 10 s at 15,000 × *g* and immediately placed on ice. An aliquot of 1 μl of the supernatant was used in the PCR. When a nasal swab sample rendered a positive *H. parasuis* identification, independent colonies from each swab (i.e., per animal) were separately tested and, if positive, stored in tryptic soy broth–20% glycerol at −80°C.

Molecular typing by ERIC-PCR. Crude DNA was prepared following the protocol described above and quantified. In each case, 100 ng was used as the template for ERIC-PCR with primers ERIC1R and ERIC2 (Table 2). We followed a previously published protocol (10) and included an extra final extension step of 20 min. PCR products were analyzed by electrophoresis (70 V, 3 h) in a 2% agarose gel. Band patterns were visualized by staining with ethidium bromide, and comparison of the patterns was performed visually.

Serotyping. Serotype determination was performed by indirect hemagglutination at the Animal Health Department of the Veterinary School at the University of León (León, Spain) following a previously published protocol (35). Soluble antigen was obtained after boiling of a bacterial suspension and subsequent centrifugation to eliminate insoluble debris.

Detection of the *lsgB* and group 1 *vtaA* genes. Crude DNA templates derived directly from colonies on agar were prepared as described above. The *lsgB* gene was PCR amplified using *Taq* polymerase (Biotools, Spain) and the four primer pairs lsgB-F1 and lsgB-R1, lsgB-F1 and lsgB-R2-2, lsgB-F2-2 and lsgB-R1, and lsgB-F2-2 and lsgB-R2-2 (Table 2). The *vtaA* translocator domains from groups 1 and 3 were amplified in a multiplex format following a previously published protocol (12). Multiplex PCR tubes contained GoTaq buffer, which consisted of 2 mM $MgCl_2$, 0.4 mM each deoxynucleoside triphosphate dNTP, 800 nM primers YADAF1 and PADHR1 (each), 400 nM primers YADAF3 and PADHR3 (each), 1 U GoTaq polymerase (Promega, USA) and 10 ng genomic DNA in a final volume of 25 μl. Cycling conditions were 5 min at 94°C, followed by 25 cycles of 45 s at 94°C, 45 s at 64°C, and 1 min at 72°C, and then a final incubation was carried out at 72°C for 7 min. In all cases, the virulent reference strain *H. parasuis* Nagasaki and the nonvirulent reference strain *H. parasuis* SW114 were used as controls. PCR for *vtaA* group 3 (with primers YADAF3 and PADHR3), previously shown to be present in all *H. parasuis* strains (12), was used as a control.

Plasmid analysis of β-lactam-resistant strains. Plasmid DNA was extracted from bacteria grown on chocolate agar plates with a QIAprep spin miniprep kit (Qiagen, Germany). PCRs were performed using *Taq* polymerase (Biotools) with crude DNA (see above) or the extracted plasmid as the template. Identification of plasmids bearing bla_{ROB-1} in β-lactam-resistant *H. parasuis* strains (strains for which the MICs of β-lactam antibiotics were elevated) was performed using primers rob-1F and rob-1R, rendering a 457-bp amplicon, and with divergent bla_{ROB-1} primers rob-1D and rob-1U, which rendered a 2,238-bp or a 4,189-bp amplicon, depending on the strain (Table 2). Purification of the PCR amplicons was performed using a NucleoSpin gel and PCR cleanup kit (Macherey-Nagel, Germany). Plasmid pJMA-1 was Sanger sequenced by PCR walking with primers pl-SEQ-rob-1F, rob-1D, pB1000-F1, plSeqF2, plSeqF3, rob-1F, rob-1R, and pl-SEQ-rob-1R (Table 2). Plasmid pB1000 was Sanger se-

TABLE 1 Farm, number of sows, type of production, piglet weaning age, animal sampled, isolates, ERIC-PCR profiles, and genetic features of *H. parasuis* isolates collected in this study

Farm	No. of sows	Production type/weaning age (days)	Animal	*H. parasuis* isolate	ERIC-PCR profile	Presence of the following:			
						lsgB	*vtaA1*	bla_{ROB-1}	bla_{ROB-1}-containing plasmid (plasmid size [kb])
A	250	Farrow to finish/28	A1	A1.1	NA-1	−	−	−	−
			A1	A1.2	NA-2	−	−	−	−
			A1	A1.3	NA-1	−	−	−	−
			A2	A2.1	NA-1	−	−	−	−
			A2	A2.2	NA-1	−	−	−	−
			A2	A2.3	NA-1	−	−	−	−
			A3	A3.1	NA-7	−	−	+	+ (2.6)
			A3	A3.2	NA-7	−	−	+	+ (2.6)
			A3	A3.3	NA-7	−	−	+	+ (2.6)
			A4	A4.1	NA-10	−	−	−	−
			A4	A4.2	NA-10	−	−	−	−
			A4	A4.3	NA-10	−	−	−	−
B	475	Farrow to weaning/21	B1	B1.1	NA-2	−	−	−	−
			B1	B1.2	NA-2	−	−	−	−
			B1	B1.3	NA-15	+	+	−	−
			B2	B2.1	NA-16	+	+	−	−
			B2	B2.2	NA-17	+	+	−	−
			B2	B2.3	NA-18	−	+	−	−
			B3	B3.1	NA-17	+	+	−	−
			B3	B3.2	NA-21	+	+	−	−
			B3	B3.3	NA-21	+	+	−	−
			B4	B4.1	NA-21	+	+	−	−
			B4	B4.2	NA-21	−	+	−	−
			B4	B4.3	NA-21	+	+	−	−
C	95	Farrow to finish/28	C1	C1.1	NA-25	−	−	−	−
			C1	C1.3	NA-27	−	−	−	−
			C2	C2.1	NA-27	−	−	−	−
			C2	C2.2	NA-29	−	−	−	−
			C2	C2.3	NA-29	−	−	−	−
			C3	C3.1	NA-29	−	−	−	−
			C3	C3.2	NA-29	−	−	−	−
			C3	C3.3	NA-29	−	−	−	−
			C4	C4.1	NA-29	−	−	−	−
			C4	C4.2	NA-29	−	−	−	−
			C4	C4.3	NA-29	−	−	−	−
D	180	Farrow to finish/28	D1	D1.1	NA-37	−	−	−	−
			D1	D1.2	NA-37	−	−	−	−
			D1	D1.3	NA-39	−	−	−	−
			D2	D2.1	NA-40	−	−	−	−
			D2	D2.2	NA-41	−	−	−	−
			D3	D3.1	NA-37	−	−	−	−
			D3	D3.2	NA-37	−	−	−	−
			D3	D3.3	NA-37	−	−	−	−
			D4	D4.1	NA-46	−	−	+	+ (2.6)
			D4	D4.2	NA-46	−	−	+	+ (2.6)
			D4	D4.3	NA-48	−	−	+	+ (4.6)
E	500	Farrow to finish/28	E1	E1.1	NA-67	−	−	−	−
			E1	E1.2	NA-67	−	−	−	−
			E1	E1.3	NA-67	−	−	−	−
			E2	E2.1	NA-52	−	−	−	−
			E2	E2.2	NA-52	−	−	−	−
			E2	E2.3	NA-54	−	−	−	−
			E3[a]						
			E4[a]						

(Continued on following page)

TABLE 1 (Continued)

Farm	No. of sows	Production type/weaning age (days)	Animal	*H. parasuis* isolate	ERIC-PCR profile	Presence of the following:			
						lsgB	*vtaA1*	bla_{ROB-1}	bla_{ROB-1}-containing plasmid (plasmid size [kb])
F	260	Farrow to finish/28	F1	F1.1	NA-29	−	−	−	−
			F1	F1.2	NA-29	−	−	−	−
			F1	F1.3	NA-29	−	−	−	−
			F2	F2.1	NA-29	−	−	−	−
			F2	F2.2	NA-29	−	−	−	−
			F2	F2.3	NA-29	−	−	−	−
			F3	F3.1	NA-2	−	−	−	−
			F3	F3.2	NA-2	−	−	−	−
			F3	F3.3	NA-2	−	−	−	−
			F4	F4.1	NA-64	+	−	+	+ (2.6)
			F4	F4.2	NA-64	+	−	+	+ (2.6)
			F4	F4.3	NA-64	−	−	+	+ (2.6)
G	160	Farrow to finish/21	G1	G1.1	NA-67	−	−	−	−
			G1	G1.2	NA-67	−	−	−	−
			G1	G1.3	NA-67	−	−	−	−
			G2	G2.1	NA-70	−	−	−	−
			G2	G2.2	NA-70	−	−	−	−
			G2	G2.3	NA-29	−	−	−	−
			G3	G3.1	NA-73	−	−	−	−
			G3	G3.2	NA-73	−	−	−	−
			G3	G3.3	NA-73	−	−	−	−
			G4	G4.1	NA-76	−	+	−	−
			G4	G4.2	NA-73	−	−	−	−
			G4	G4.3	NA-73	−	−	−	−
H	800	Farrow to weaning/28	H1	H1.1	NA-79	−	+	−	−
			H1	H1.2	NA-80	−	−	−	−
			H1	H1.3	NA-79	−	+	−	−
			H2	H2.1	NA-82	−	−	−	−
			H3	H3.1	NA-79	−	+	−	−
			H3	H3.2	NA-79	−	+	−	−
			H3	H3.3	NA-79	−	+	−	−
			H4	H4.1	NA-88	−	−	−	−
			H4	H4.2	NA-88	−	−	−	−
			H4	H4.3	NA-88	−	−	−	−

[a] *H. parasuis* was not isolated from animals E3 and E4.

quenced by PCR walking with primers rob-1D, rob-1U, plSeqF3, Pl298-SEQ-rob-1F, Pl298-SEQ-F, Pl298-SEQ-R, Pl298-SEQ2-F, and Pl157-SEQ-R (Table 2).

Antimicrobial susceptibility testing. *In vitro* antimicrobial susceptibility was determined by microdilution methods following Clinical and Laboratory Standards Institute (CLSI) guidelines (36, 37) and the methods described in previous studies (11, 29). Briefly, bacterial inocula were prepared from colonies freshly grown for 24 h on chocolate agar plates, and the inoculum was adjusted to a 0.5 McFarland standard in haemophilus test medium (HTM; Becton Dickinson, USA); 50 μl of each bacterial suspension was used per well. The antimicrobials used for microdilution were AMP and amoxicillin (AMX). Stock solutions were prepared from the pure powder form according to the CLSI standard method (36, 37), and the antimicrobials were used over a concentration range from 1 to 2,048 μg/ml. Alternatively, we used commercially available dehydrated Sensititre plate panels (EUMVS2; Trek Diagnostics Inc., USA) for cefotaxime (CTX), ceftazidime (CAZ), sulfamethoxazole (SXT), gentamicin (GEN), ciprofloxacin (CIP), tetracycline (TET), streptomycin (STR), trimethoprim (TMP), chloramphenicol (CHL), colistin (CST), florfenicol (FFN), kanamycin (KAN), and nalidixic acid (NAL). For both the microdilution and Sensititre assays, panel quality controls were performed with strains *H. parasuis* BB1023 (β-lactam resistant) and BB1033 (β-lactam sensitive) (11). The MIC was defined as the lowest antimicrobial concentration that inhibited bacterial growth. Given that a standard technique for determination of the antimicrobial susceptibility of *H. parasuis* and agreed-upon interpretation criteria do not currently exist (38), the MIC results were reviewed and the distribution of strains over the MIC range was considered.

Biofilm formation. *H. parasuis* strains were grown on chocolate agar plates for 16 h at 37°C in 5% CO_2. Four to five freshly grown colonies were used to inoculate 10 ml BHI with 10 μg/ml β-NAD and grown at 37°C with 5% CO_2 to an optical density at 600 nm (OD_{600}) of ~0.7. Ten microliters of the grown culture was diluted in 100 μl sBHI on 96-well polystyrene plates (Iwaki, Japan); alternatively, 100 μl of the grown culture was diluted in 586 μl sBHI on 24-well polystyrene plates (Costar; Corning, USA). The plates were incubated statically for 36 h at 37°C in 5% CO_2. The culture medium was then carefully removed, the plates were washed in distilled water, and the attached bacteria were fixed for 15 min with 100 μl (96-well plates) or 586 μl (24-well plates) methanol. The methanol was removed, and the plates were air dried for 15 min. Once the

TABLE 2 Primers used in this study

Primer	Sequence (5′–3′)	Reference or source
HPS-F	GTGATGAGGAAGGGTGGTGT	34
HPS-R	GGCTTCGTCACCCTCTGT	34
ERIC1R	ATGTAAGCTCCTGGGGATTCAC	10
ERIC2	AAGTAAGTGACTGGGGTGAGCG	10
lsgB-F1	ATGAATTTGATTATTTGTATGACTCCATTT	13
lsgB-F2-2	ATTCGGTCTGGAATGATAAATATCAGTATT	This study
lsgB-R1	CTATTGGCATGTGTAGTCAATTACTTC	13
lsgB-R2-2	AAGAGGAGCTCCACTATAGAACGTATAAAT	This study
YADAF1	TTTAGGTAAAGATAAGCAAGGAAATCC	12
PADHR1	CCACACAAAACCTACCCCTCCTCC	12
YADAF3	AATGGTAGCCAGTTGTATAATGTT	12
PADHR3	CCACTG TAATGCAATACCTGCACC	12
rob-1F	TGTTGCAATCGCTGCC	11
rob-1R	TTATCGTACACTTTCCA	11
rob-1U	ATCGTCATGCCTTTGCCAACG	11
rob-1D	AATTGGTTGGACAATAACGCA	11
pl-SEQ-rob-1F	CCCGACCGCTTTCAGCGGTCAAAA	This study
pl-SEQ-rob-1R	TAGGAAGTACTCATCATTTGGAAG	This study
pB1000-F1	TCCATCTAAAGAATGTGAAGTATTACT	This study
pB1000-R2	CTGCTAACGCCCTGCGGTTGTTAGACT	This study
plSeqF2	TTAGCACTATAAGGCTAAGCGACTAGG	This study
plSeqR2	AGGGAAGTGGTTATTGTGGGCGGTAGG	This study
plSeqF3	AGTGATAACGTTCTTAATACG	This study
Pl298-SEQ-F	GGCTCACGATGTTCGCCTAAT	This study
Pl298-SEQ-R	ACGGCTAGGAAGGGAATAGGG	This study
Pl298-SEQ-rob-1F	CGAATTTCCGCGCAATGTACG	This study
Pl298-SEQ2-F	GGCAGGCTTGAGCTGAATTTC	This study
Pl157-SEQ-R	TGTCCACTGCTACACAAGGCT	This study
Tet(B)-F	ACGTTACTCGATGCCAT	This study
Tet(B)-R	AGCACTTGTCTCCTGTT	This study
Tet(H)-F	ATACTGCTGATCACCGT	This study
Tet(H)-R	TCCCAATAAGCGACGCT	This study
Tet(M/O/S)-F	ATAGAYACGCCAGGMCATAT	This study
Tet(M/O/S)-R	GAAGCCCAGAAAGGATTYGGY	This study
Tet(B)invF	GGTTAGTTTTCCCTGTTTTA	29
Tet(B)invR	ACCAACCGAACCACTTCACG	29
Tet(H)invF	GGGTCATCTTACCAGCATTA	29
Tet(H)invR	AGAAACCAAAATAGCGAGTT	29

plates were dry, 100 μl (96-well plates) or 586 μl (24-well plates) of 5% crystal violet was added to each well for 5 min and the plates were incubated at room temperature. The plates were washed in distilled water and then dried for 30 min at 37°C. Stained biofilms were resolubilized with 100 or 586 μl (96-well and 24-well plates, respectively) 80% ethanol–20% acetic acid. Biofilm formation was quantified by measuring the absorbance at 595 nm. To assess the nature of the biofilm matrix, biofilm assays were performed on 24-well plates. Once biofilms formed, the wells were washed in distilled water and treated with 100 μl (96-well plates) or 586 μl (24-well plates) 10 mM sodium *meta*-periodate dissolved in 50 mM sodium acetate buffer (pH 4.5) or with 100 μg/ml proteinase K dissolved in a buffer containing 100 mM NaCl, 20 mM Tris (pH 7.5), as previously described (39). Controls with adequate buffers (50 mM sodium acetate buffer [pH 4.5] or 100 mM NaCl and 20 mM Tris [pH 7.5], respectively) were run in parallel. After treatment, the remaining material on the wells was fixed, and the plates were stained and monitored as described above. In all cases, assays were carried out in triplicate and on at least three independent occasions ($n \geq 9$). Means and standard deviation (SDs) of the absorbance values were calculated.

Cell culture, bacterial infection, and quantification of IL-8 secretion. PK-15 immortalized pig kidney epithelial cells (ATCC CCL-33) were maintained in RPMI 1640 medium supplemented with 10 mM HEPES, 10% fetal calf serum (FCS), 100 units/ml penicillin, and 0.1 mg/ml streptomycin (complete medium). Cells were seeded at 8×10^4 cells per well in 24-well tissue culture plates for 24 h. The cells were then serum starved for 16 h before infection by replacement of the complete medium with medium lacking FCS. A confluence of 90% was reached at the time of infection. *H. parasuis* cells grown to stationary phase were recovered with 1 ml phosphate-buffered saline (PBS) from a chocolate agar plate and adjusted with PBS to an OD_{600} of 1 ($\sim 10^9$ CFU/ml). Cells in 1 ml Earle's balanced salt solution (EBSS) were infected with the freshly obtained bacterial suspension at a multiplicity of infection of approximately 100:1. The plates were centrifuged at $400 \times g$ for 5 min and incubated at 37°C in 5% CO_2 for 2 h. Infected cells were washed 3 times with PBS and incubated with fresh RPMI 1640 medium containing 10 mM HEPES, 10% FCS, 100 μg/ml gentamicin, and 5 μg/ml penicillin for 6 h. Supernatants were removed from the wells, the cell debris was removed by centrifugation, and samples were frozen at −80°C. The interleukin-8 (IL-8) levels in the supernatants were measured using an enzyme-linked immunosorbent assay (ELISA) kit (Kingfisher Biotech, USA). Infection experiments were carried out in duplicate and on at least two independent occasions ($n \geq 4$). Means and SDs were calculated, and comparison of means for statistically significant differences was performed using the

two-tailed *t* test (Prism software, version 4, for the personal computer). A *P* value of <0.05 was considered statistically significant.

Stability of pJMA-1. Plasmid pJMA-1 was transformed into *H. influenzae* Rd KW20 electrocompetent cells (40). Transformants were selected on sBHI agar plates containing 25 μg/ml ampicillin. The rate of curing of plasmid pJMA-1 in *H. influenzae* Rd KW20 transformants was determined as previously described (29). Briefly, 4 to 5 colonies of bacteria harboring pJMA-1 grown on chocolate agar containing 25 μg/ml ampicillin were inoculated in 10 ml sBHI and grown at 37°C in 5% CO_2 for 12 h. This culture was serially passaged eight times by serial 1:100 dilution in 10 ml sBHI and growth for 12 h. In each subculture step, bacteria were plated on chocolate agar, and the proportion of resistant colonies harboring pJMA-1 was deduced by replica plating of at least 100 colonies on sBHI agar and sBHI agar plates containing 25 μg/ml ampicillin. The rate of plasmid curing was calculated as the percentage of ampicillin-resistant colonies at cycle 8/total number of colonies replicated at cycle 8.

Fitness cost determination. The fitness cost of plasmid pJMA-1 was determined by eight independent competition experiments between *H. influenzae* Rd KW20 and *H. influenzae* Rd KW20(pJMA-1), as previously described (30, 41). Strains were grown in HTM broth for 16 h at 37°C with 5% CO_2. Then, 10^6 CFU of *H. influenzae* Rd KW20 was mixed with 10^6 CFU of *H. influenzae* Rd KW20(pJMA-1) in 2 ml of antibiotic-free HTM. The mixture was grown at 37°C in 5% CO_2 at 100 rpm. A total of 2×10^6 CFU was transferred to 2 ml of fresh HTM (1:1,000 dilution) every 24 h. Samples were taken every 24 h for 5 days. For each sample, aliquots were plated on nonselective chocolate agar, and the proportion of resistant colonies was deduced by replica plating of at least 50 colonies on chocolate agar plates containing ampicillin at 128 μg/ml. Relative fitness is expressed as the competition index (CI), calculated as the ratio of the mean number of CFU from eight independent competition experiments between the resistant and susceptible strains at a given time point (t_1) divided by the same ratio at time zero. The selection coefficient (s) was calculated as the slope of the linear regression model: $s = \ln(\mathrm{CI})/t$, where t is time, which was the number of bacterial generations calculated as the $\log_2$ value of the dilution factor. The selection coefficient estimates the difference between the relative fitness of the two competitors over the entire competition experiment (42).

Nucleotide sequence accession numbers. The plasmid nucleotide sequences obtained in this study have been deposited in GenBank under the following accession numbers: KP164835 for pJMA-1 from strain F4.1, KP164833 for pB1000 from strain D4.3, KP164834 for pJMA-1 from strain PM5-4, and KP164832 for pB1000 from strain NU5-3.

RESULTS

Genetic features of *H. parasuis* isolates collected from the nasal cavities of healthy pigs at weaning. Eighty-six *H. parasuis* strains were isolated in 2012 from nasal swab samples collected from healthy piglets on eight swine farms located in Navarra, Spain, that had not had a case of Glässer's disease since 2000. Table 1 summarizes the farms, the animals sampled, the isolate ERIC-PCR profiles, and the genetic features of the *H. parasuis* isolates. Four animals per farm were sampled ($n = 32$ animals); all animals were positive for the presence of *H. parasuis*, except for two of the animals sampled on farm E, and 86 isolates were confirmed to be *H. parasuis* by biochemical, mass spectrometry, and molecular methods. Isolates were genotyped by ERIC-PCR, rendering 29 profiles. Most ERIC-PCR profiles were found on a single farm (farm specific); the exceptions were 3 profiles (NA-2, -29, and -67) for isolates recovered from animals sampled on more than one farm. Nasal isolates with the same ERIC-PCR profile could be recovered from the same animal, from animals sampled on the same farm, and from animals sampled on different farms.

Given that putative virulent strains can be found in the upper respiratory tract of healthy animals (6) and that the presence of the *lsgB* and group 1 *vtaA* genes has previously been found to be associated with *H. parasuis* virulence (12, 13), we assessed their distribution among the collected *H. parasuis* nasal isolates. The *lsgB* gene was present in 10 isolates displaying 5 different ERIC-PCR profiles (Table 1). Most *lsgB*-positive strains were isolated from animals sampled on farm B. The *lsgB* gene was not found in isolates B4.2 and F4.3, even though they presented the same ERIC-PCR profile as other isolates rendering a positive amplification for this gene. The group 1 *vtaA* gene was present in 16 isolates from farms B, G, and H displaying 6 different ERIC-PCR profiles (Table 1). Isolates with the same ERIC-PCR profile did not differ regarding the distribution of the group 1 *vtaA* gene.

In summary, upon nasal swab sampling of 32 healthy animals on 8 farms, *H. parasuis* was detected in 30, and a total of 86 isolates were obtained. ERIC-PCR-based genotyping showed diversity among the isolates from the different farms. A subset of isolates contained the *lsgB* and group 1 *vtaA* genes, previously found to be associated with *H. parasuis* virulence potential.

Biofilm formation and epithelial inflammation caused by *H. parasuis* isolates collected from the nasal cavities of healthy pigs at weaning. Given that biofilm formation has previously been related to a lack of virulence (43), we next asked if *H. parasuis* isolates collected from the nasal cavities of healthy piglets could form biofilms. Assessment of biofilm growth requires inoculation of comparably grown bacterial cultures. Thus, we first screened the ability of *H. parasuis* isolates with profiles representative of the previously established ERIC-PCR profiles to grow aerobically in BHI medium supplemented with β-NAD. We found that a significant proportion of isolates displayed aggregative growth or, alternatively, low turbidity after 16 h of growth at 37°C (data not shown). Based on reliable growth in sBHI, we selected isolates A1.2, B2.2, B3.2, F4.1, G4.1, and H1.1 and included the reference strains *H. parasuis* SW114 (nonvirulent) and Nagasaki (virulent). Under the conditions tested, isolates A1.2, B2.2, B3.2, F4.1, and G4.1 displayed variable biofilm growth on 96-well plates. Strains H1.1, SW114, and Nagasaki did not form biofilms (Table 3). To compare the nature of the biofilm structures, a biofilm detachment assay was performed on 24-well plates. The biofilm structures were sensitive to proteinase K and resistant to sodium *meta*-periodate (Table 3). These findings were independent of the amount of biofilm formed, suggesting that *H. parasuis* biofilm formation, although heterogeneous among isolates, may depend on the presence of proteins but not on the presence of sugar components in the extracellular matrix.

H. parasuis has also been shown to induce inflammation, which may occur to different extents depending on the strain (22, 44, 45). We analyzed the ability of the previously selected *H. parasuis* isolates to trigger IL-8 secretion by PK-15 porcine epithelial cells at 8 h postinfection. All strains stimulated IL-8 secretion by PK-15 cells at comparable levels, independently of their origin or virulence potential (Fig. 1).

These results show protein-dependent biofilm growth for *H. parasuis* nasal isolates, which, moreover, also triggered secretion of IL-8 by PK-15 cells.

Identification of a novel small plasmid bearing $bla_{\mathrm{ROB\text{-}1}}$ in colonizing *H. parasuis* isolates. Clinical isolates of *H. parasuis* and *P. multocida* recovered from animals in Spain have previously been reported to carry the 4,613-bp plasmid pB1000 bearing the $bla_{\mathrm{ROB\text{-}1}}$ β-lactamase and conferring resistance to β-lactams (11, 29). Animals play important roles as reservoirs and sources of

TABLE 3 Absorbance (OD_{595}) of bacterial biofilm detached from polystyrene wells for representative *H. parasuis* strains

H. parasuis isolate	OD_{595} determined on:				
		24-well plates			
		Sugar dependence		Protein dependence	
	96-well plates	Buffer	Sodium *meta*-periodate	Buffer	Proteinase K
A1.2	0.24 ± 0.08	ND[a]	ND	ND	ND
B2.2	0.43 ± 0.07	1.43 ± 0.2	1.48 ± 0.22	1.04 ± 0.1	0.11 ± 0.04
B3.2	0.29 ± 0.05	0.9 ± 0.21	0.97 ± 0.11	0.57 ± 0.05	0.14 ± 0.07
F4.1	0.23 ± 0.13	ND	ND	ND	ND
G4.1	0.6 ± 0.2	1.77 ± 0.29	1.63 ± 0.25	1.16 ± 0.29	0.22 ± 0.06
H1.1	0.02 ± 0.02	ND	ND	ND	ND
Nagasaki	0.07 ± 0.02	ND	ND	ND	ND
SW114	0.01 ± 0.01	ND	ND	ND	ND

[a] ND, not determined.

antimicrobial-resistant pathogens (46), due to the zoonotic circulation of resistant bacteria and the spread of antimicrobial resistance determinants between animal and human pathogens (47). In fact, pB1000 has also been found in *H. influenzae* clinical isolates in Spain, Italy, Australia, and North America (30, 48, 49). Moreover, biofilm formation has been associated with resistance to β-lactams in *H. parasuis* (28). Considering both observations, the collection of *H. parasuis* isolates recovered from the nasal cavities of healthy pigs at weaning generated in this study was surveyed for bla_{ROB-1} by PCR. Nine bla_{ROB-1}-positive isolates with 4 different ERIC-PCR profiles were identified (Table 1). The β-lactam resistance of one bla_{ROB-1}-positive strain per ERIC-PCR profile was determined (Table 4). bla_{ROB-1} is a 918-bp β-lactamase that has been found to confer resistance to ampicillin and that has been found to be carried by different plasmids in *Pasteurellaceae* species, such as *H. influenzae* (30, 48, 49), *P. multocida* (29), *H. parasuis* (11), *Actinobacillus porcitonsillarum* (50), and *Actinobacillus pleuropneumoniae* (51). To characterize the genetic platform bearing bla_{ROB-1} in the nasal isolates under study, we performed a bla_{ROB-1}-specific inverted PCR with primers rob-1D and rob-1U, using as the template crude DNA and plasmid purified from the nine bla_{ROB-1}-positive isolates. Inverted PCR rendered an approximately 2.2-kb amplicon for isolates A3.1, A3.2, A3.3, D4.1, D4.2, F4.1, F4.2, and F4.3 and an approximately 4.2-kb amplicon for strain D4.3. Purified plasmid extracted from isolate F4.1, used as a representative strain, was completely sequenced by walking PCR using the primers listed in Table 2 (see the Materials and Methods section), rendering a novel plasmid named pJMA-1 (GenBank accession number KP164835). Purified plasmid extracted from strain D4.3 was also completely sequenced and shown to correspond to the previously described pB1000 (GenBank accession number KP164833).

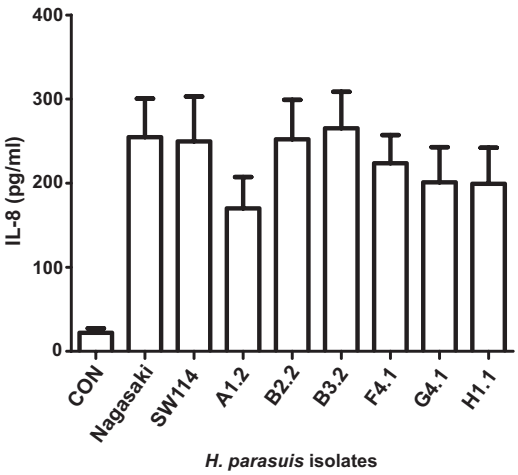


FIG 1 Stimulation of IL-8 secretion by *H. parasuis* infection of cultured porcine cells. The concentration of IL-8 secreted into the supernatant by PK-15 cells upon infection with *H. parasuis* strains Nagasaki, SW1174, A1.2, B2.2, B3.2, F4.1, G4.1, and H1.1 was quantified by ELISA at 8 h postinfection. CON, control (uninfected) cells.

TABLE 4 MICs of four β-lactams for 16 *H. parasuis* and 2 *H. influenzae* strains used in this study

Strain	MIC[a] (μg/ml)			
	Ampicillin	Amoxicillin	Cefotaxime	Ceftazidime
H. parasuis A3.1	64	2,048	0.12	≤0.25
H. parasuis D4.1	256	512	≤0.06	≤0.25
H. parasuis D4.3	16	2,048	≤0.06	≤0.25
H. parasuis F4.1	>32	>2,048	≤0.06	≤0.25
H. parasuis F9	128	>2,048	≤0.06	≤0.25
H. parasuis MU21-2	32	>2,048	≤0.06	≤0.25
H. parasuis V56-2	32	>2,048	≤0.06	2
H. parasuis SC14-1	>32	1,024	≤0.06	≤0.25
H. parasuis 416-1	128	>2,048	≤0.06	≤0.25
H. parasuis MU25-5	32	1,024	≤0.06	≤0.25
H. parasuis NU5-3	>32	>2,048	≤0.06	≤0.25
H. parasuis CD7-3	256	>2,048	≤0.06	1
H. parasuis CD9-1	256	>2,048	≤0.06	≤0.25
H. parasuis PM5-4	>32	512	≤0.06	≤0.25
H. influenzae Rd KW20	<2	<4	ND	ND
H. influenzae Rd KW20(pJMA-1)	512	512	ND	ND
H. parasuis BB1033 (Amp^s)	≤2	≤4	0.25	2
H. parasuis BB1023 (Amp^r)	>1,024	1,024	≤0.06	≤0.25

[a] Ampicillin and amoxicillin MICs were determined by microdilution. Cefotaxime and ceftazidime MICs were determined by the Sensititre assay. ND, not determined.

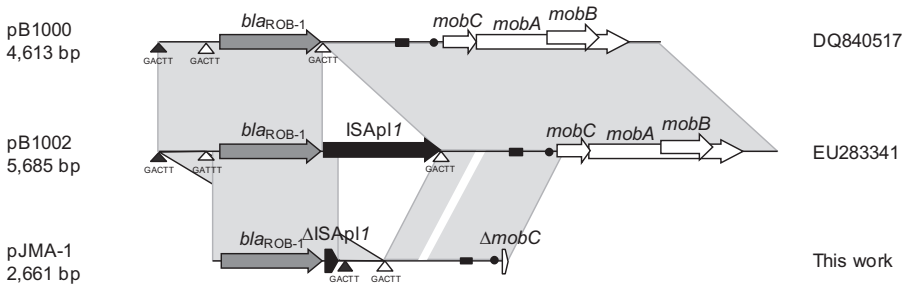


FIG 2 Comparison of the genetic structures of pB1000, pB1002, and pJMA-1. The reading frames for genes are shown as arrows, and the direction of transcription is indicated by arrowheads. Regions with more than 99% identity are shaded gray. Triangles, positions of the CTGAA/GACTT sequence; black triangles, sites used by ISApl*1* to integrate bla_{ROB-1} in pJMA-1; white triangles, probable insertion sites used to generate pB1000 and pB1002; black rectangles, putative origins of replication (*oriV*); black dots, putative transfer origin (*oriT*). The designations DQ840517 and EU283341 are the GenBank accession numbers for pB1000 and pB1002, respectively.

The 2,661-bp plasmid pJMA-1 bears bla_{ROB-1} and shares a common backbone with other small plasmids described in the *Pasteurellaceae*; this common backbone is virtually identical in pJMA-1, pB1000 (30), and pB1002 (29). Furthermore, pJMA-1 is highly similar to pHS-Tet (52), which bears *tet*(B) instead of bla_{ROB-1}. The pJMA-1 sequence, excluding the ca. 90 nucleotides near the origin of replication, is 99% identical to the sequence of pB1002 (Fig. 2), a homologue of pB1000 which bears the transposase ISApl*1* (29). ISApl*1* is a 1,072-bp insertion sequence (IS) belonging to the IS*30* family found in the *Pasteurellaceae* family (53, 54). pJMA-1 contains the first 119 nucleotides of ISApl*1*, including the left inverted repeat downstream of the bla_{ROB-1} gene. In pB1000 and pB1002, ISApl*1* has been suggested to mobilize bla_{ROB-1} through integration in and duplication of 5 nucleotides (CTGAA/GACTT), caused by the insertion via ISApl*1* (29). The same 5 nucleotides are present in pJMA-1 downstream of the ISApl*1* deletion site. Interestingly, bla_{ROB-1} from pJMA-1 is inserted in the 5-bp repeat (nucleotides 2126 to 2131 in the pB1000 sequence), whereas bla_{ROB-1} from pB1000 is inserted between nucleotides 2528 and 2532 (Fig. 2), showing that the target sequence of ISApl*1* is present in several locations in these plasmids. These observations suggest that the integration/excision of resistance genes through this IS is likely to be a plastic process and may elicit different combinations. pJMA-1 belongs to the ColE1 superfamily. Replication is likely to be mediated by two antisense RNAs, which occurs in all ColE1 plasmids (55). Putative DnaA boxes are encoded between positions 2248 and 2257 and positions 2421 and 2430 (TGATGGATAA and TTTATATATCA, respectively). The putative origin of replication (*oriV*) is encoded between nucleotides 2156 and 2174 (TTAGTAAAAAGCGTTCTTT). Additionally, plasmid pJMA-1 carries 49 nucleotides of the 5′ end of the *mobC* relaxase, as well as the putative origin of transfer (*oriT*; CAAAAAGGGTACTCTTCGAGTTTCCTTTTTGACCTTG), required for plasmid mobilization.

The presence of pJMA-1 in *H. parasuis* strains A3.1, A3.2, A3.3, D4.1, D4.2, F4.2, and F4.3 was further confirmed by PCR mapping, using as a template purified plasmid and six PCR combinations with nine primers, as follows: (i) rob-1F and rob-1R, rendering a 457-bp product; (ii) rob-1D and rob-1U, rendering a 2,394-bp product; (iii) plSeqF2 and plSeqR2, rendering a 954-bp; (iv) pB1000-F1 and pB1000-R2, rendering a 494-bp product; (v) rob-1F and pB1000-R2, rendering a 1,280-bp product; and (vi) pl-SEQ-rob-1F and rob-1U, rendering a 522-bp product.

To confirm the functionality of the plasmid and the β-lactamase, the pJMA-1 plasmids purified from the eight *H. parasuis* strains shown to carry this plasmid were separately transformed into *H. influenzae* Rd KW20 electrocompetent cells, and transformants were selected on sBHI agar plates containing 25 μg/ml ampicillin. Both *H. influenzae* Rd KW20(pJMA-1) crude DNA and purified plasmids were used for PCR mapping using the six primer pairs described above, rendering in all cases products compatible with pJMA-1. In conclusion, these results identify the presence of a novel plasmid bearing the bla_{ROB-1} gene and conferring resistance to β-lactams in eight *H. parasuis* isolates recovered from the nasal cavities of healthy piglets.

Association of pJMA-1 with *H. parasuis* nasal isolates from healthy pigs at weaning. Whereas pB1000 was described in β-lactam-resistant *H. parasuis* isolates recovered from characteristic lesions of Glässer's disease (11), pJMA-1 was found in β-lactam-resistant *H. parasuis* isolates recovered from healthy animals. To further assess the association of pJMA-1 with *H. parasuis* isolates from the nasal cavities of healthy pigs, we sampled 19 strains isolated from the nasal cavities of animals on farms located in Catalonia and 1 nasal strain from Mallorca, eastern Spain. Interestingly, 10 isolates were bla_{ROB-1} positive (Table 5). A bla_{ROB-1}-specific inverted PCR with primers rob-1D and rob-1U rendered a single amplicon of approximately 2.2 kb in nine isolates and a single amplicon of approximately 4.2 kb in one isolate. The first amplicon was completely sequenced by walking PCR for one representative strain, PM5-4, and its sequences was shown to be identical to that of pJMA-1 (GenBank accession number KP164834). PCR mapping of the remaining eight bla_{ROB-1}-positive strains using the six primer pairs described above rendered PCR products with sizes compatible to those shown by pJMA-1, suggesting that all eight strains bore this plasmid. Differently, walking PCR sequencing of purified plasmid from strain NU5-3 rendered pB1000 (GenBank accession number KP164832). The β-lactam resistance profiles of the 10 bla_{ROB-1}-positive strains were determined (Table 4). Accordingly, we tested the remaining 10 strains belonging to this set of isolates from healthy piglets for β-lactam resistance and found that they presented low antimicrobial MICs, with all isolates being considered susceptible to ampicillin (data not shown). Serovar information for a subset of these nasal strains (Table 5) showed that pJMA-1-bearing strains belonged to serovars 12 and 14 (virulent), 15 (moderately virulent), and 6 and 7 (nonvirulent), which limits the ability to make a clear-cut association between a

TABLE 5 Distribution of the bla_{ROB-1} gene and bla_{ROB-1}-containing plasmid in *H. parasuis* nasal isolates from pigs in Catalonia and Mallorca, Spain, sampled at weaning

H. parasuis nasal isolate	bla_{ROB-1} presence	bla_{ROB-1}-containing plasmid size (kb)	β-Lactamase plasmid	Serovar
CA38-4	−			12
SL8-2	−			ND[a]
F9	+	2.6	pJMA-1	6
MU21-2	+	2.6	pJMA-1	7
ND14-1	−			7
FL1-3	−			10
VS6-2	+	2.6	pJMA-1	15
SC14-1	+	2.6	pJMA-1	15
416N-1	+	2.6	pJMA-1	15
LH9N-4	−			10
MU25-5	+	2.6	pJMA-1	12
NU5-3	+	4.6	pB1000	ND
VB4-1	−			7
ND19-4	−			5
SR2-2	−			12
PO125	−			ND
CD7-3	+	2.6	pJMA-1	14
CD9-1	+	2.6	pJMA-1	15
SN10-1	−			ND
PM5-4	+	2.6	pJMA-1	6

[a] ND, not determined.

particular serovar(s) and the presence of pJMA-1. The pJMA-1 plasmids purified from these nine nasal *H. parasuis* strains were next separately transformed into *H. influenzae* Rd KW20, and transformants were selected on sBHI agar plates with 25 μg/ml ampicillin. *H. influenzae* Rd KW20(pJMA-1) crude DNA and purified plasmids were used for PCR mapping, as described above, rendering in all cases products compatible with pJMA-1. As expected, *H. influenzae* Rd KW20 transformed with pJMA-1 purified from *H. parasuis* PM5-4 (chosen as a representative strain) was shown to be resistant to (to have elevated MICs for) ampicillin and amoxicillin (Table 4).

We next asked if pJMA-1 could also be found in virulent strains isolated from characteristic lesions of Glässer's disease. We sampled 24 strains isolated from polyserositis or pneumonia lesions. All strains were bla_{ROB-1} and bla_{ROB-1} inverted PCR negative (see Table S1 in the supplemental material). In agreement with that finding, when we tested the β-lactam resistance of this set of strains isolated from lesions, we found that all of them were susceptible to ampicillin (data not shown). Finally, we assessed nontypeable *H. influenzae* isolates for the presence of pJMA-1. We screened 20 nasopharyngeal isolates from healthy children, 16 ear isolates from children with otitis media, 6 respiratory isolates from children, 33 respiratory isolates from samples from adult patients suffering from bronchiectasis or chronic obstructive pulmonary disease, and 2 isolates from adult patients with pneumonia. All strains were bla_{ROB-1} and bla_{ROB-1} inverted PCR negative (data not shown).

In sum, our strain survey suggests an association between pJMA-1 and *H. parasuis* strains isolated from the nasal cavities of healthy piglets.

Antibiotic resistance profiles and coexistence with other small plasmids. To get further insight into the antibiotic resistance of the isolates, 1 representative of each of the 14 genetically different *H. parasuis* nasal isolates carrying pJMA-1 or pB1000 was further explored. These isolates displayed a whole range of MIC values for sulfamethoxazole, gentamicin, streptomycin, trimethoprim, chloramphenicol, colistin, kanamycin, and nalidixic acid (see Table S2 in the supplemental material). All isolates were sensitive to ciprofloxacin and florfenicol. Of note, isolates F9, SC14-1, and PM5-4 were resistant (they showed elevated MICs) to tetracycline, displaying tetracycline MICs of 16, 16, and 8 μg/ml, respectively. To assess the genetic determinant(s) linked to tetracycline resistance, we used as a template purified plasmid from strains F9, SC14-1, and PM5-4 for PCR detection of different resistance genes (see the primers listed in Table 2). Strain F9 rendered a positive *tet*(B) amplification, and strain PM5-4 was *tet*(H) positive. Strain SC14-1 was negative for the *tet* genes tested [primer pairs Tet(B)-F and Tet(B)-R, Tet(H)-F and Tet(H)-R, and Tet(M/O/S)-F and Tet(M/O/S)-R]. Moreover, inverse PCR with primers Tet(B)invF and Tet(B)invR, annealing in the *tet*(B) gene, rendered a ca. 4-kb product for strain F9, and inverse PCR with primers Tet(H)invF and Tet(H)invR, annealing in the *tet*(H) gene, rendered a ca. 6-kb product for strain PM5-4.

Stability and fitness cost of pJMA-1. We next assessed the stability of plasmid pJMA-1. To do so, cultures of *H. influenzae* Rd KW20(pJMA-1) were inoculated and propagated every 12 h for eight serial passages in triplicate in medium without antibiotic. One hundred percent of the colonies remained resistant to 25 μg/ml ampicillin after the eight subcultures and maintained pJMA-1, as confirmed by plasmid extraction and PCR mapping as described above.

We also determined the biological cost of pJMA-1. Competition experiments offer discriminative and precise measurements of fitness, since they reflect the competitive disadvantage during all phases of the growth cycle and in several consecutive cycles (56). As in a previous study (29), we performed competition experiments with *H. influenzae* Rd KW20. Briefly, 2 ml of HTM was inoculated with *H. influenzae* Rd KW20 and *H. influenzae* Rd KW20(pJMA-1), and 2×10^6 CFU from this mixture was transferred to 2 ml fresh HTM every 24 h for 5 days. The proportion of bacteria carrying pJMA-1 was measured at the 0-h time point and then every 24 h until day 5. The proportion of pJMA-1-carrying bacteria was fairly well maintained over time. The selection coefficient (*s*) was established to estimate the difference in relative fitness between the strains in competition over the entire experiment. *H. influenzae* Rd KW20 bearing pJMA-1 presented a competitive disadvantage of ca. 1.76% per 10 generations relative to the growth of *H. influenzae* Rd KW20 (Fig. 3).

Altogether, these results indicate that pJMA-1 is stably maintained in the *H. influenzae* Rd KW20 heterologous host and has a weak impact on bacterial fitness.

DISCUSSION

In this study, we explored the genetic diversity of *H. parasuis* strains isolated from the nasal cavities of healthy piglets at weaning to examine two relevant issues associated with colonization and infection by this opportunistic pathogen: (i) the presence of genetic and phenotypic traits related to the virulence potential of colonizing strains and (ii) the presence of mobile genetic elements linked to resistance to antibiotics in colonizing strains. We first focused on *H. parasuis* strains isolated from the nasal cavities of healthy animals from eight farms that had not had a case of Gläss-

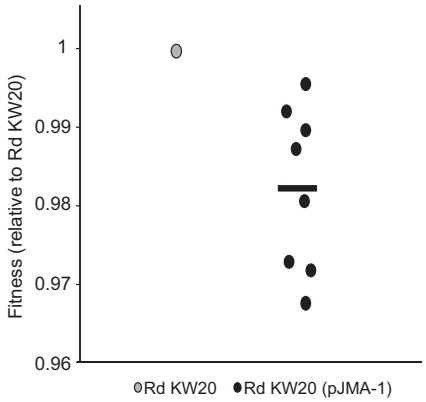


FIG 3 Fitness of Rd KW20(pJMA-1) relative to that of the plasmid-free ancestor strain Rd KW20. Fitness (W) was calculated as $W = 1 + s$, where s is the selection coefficient (see Materials and Methods). Black dots, fitness of Rd KW20(pJMA-1) relative to that of Rd KW20 (gray dot) obtained in eight independent competition assays; black line, average relative fitness of the eight measurements.

er's disease since 2000 and found isolates with the same ERIC-PCR profile recovered from the same animal, from animals sampled on the same farm, and from animals sampled on different farms, suggesting the existence of circulating *H. parasuis* genotypes likely to be commonly found in healthy animals. The group 1 *vtaA* and *lsgB* genes, previously associated with the *H. parasuis* virulence potential (12, 13), were also observed in a subset of the colonizers recovered in this study, in agreement with previous evidence of the presence of strains with virulence potential in the nasal cavities of healthy pigs (57–59). Although most *lsgB*-positive strains were isolated from animals sampled on farm B, to our knowledge, this farm does not present any feature differentiating it from the other seven farms sampled. The absence of disease caused by the potentially virulent strains detected could be explained by the protection acquired by the piglets through lactation, i.e., the protection conferred by maternal immunity. Moreover, the *lsgB* gene was found in four out of five isolates presenting ERIC-PCR profile NA-21 and in two out of three isolates presenting ERIC-PCR profile NA-64. We acknowledge that isolates B4.2 (ERIC-PCR profile NA-21) and F4.3 (ERIC-PCR profile NA-64) could lack the *lsgB* gene or, alternatively, could present a significant number of polymorphisms in this gene, which would make them negative for PCR products after PCR using four different primer pair combinations.

Isolate heterogeneity was found not only at the genomic level but also in terms of the ability of the isolates to reproducibly grow in BHI liquid medium supplemented with β-NAD. This limitation prompted us to select for further analysis a panel of isolates with comparable growth levels. All strains tested triggered the secretion of the proinflammatory cytokine IL-8 by cultured epithelial cells to comparable levels, independently of the presence or absence of the *lsgB* and group 1 *vtaA* genes. Following this notion, *H. parasuis* field strains belonging to serotypes 4 and 5 have been shown to induce the release of IL-8 and IL-6 by porcine brain microvascular endothelial cells to a similar extent (60). A different finding was that *H. parasuis* has previously been shown to stimulate IL-8 and IL-6 release by newborn pig tracheal cells, and serotype 4 field isolates induced higher levels of these mediators than did serotype 5 isolates (44). The previously observed induction of proinflammatory IL-8 by virulent *H. parasuis* strains seems to be linked to the capacity of these strains to evade the immune response and multiply inside the host more than to their intrinsic capacity to induce higher levels of inflammation than nonvirulent strains (61).

In previous studies of biofilm formation by *H. parasuis*, a relationship between colonizers and biofilm growth was suggested (43). In this study, biofilm formation was shown to be heterogeneous; however, the biofilm was sensitive to proteinase K treatment, suggesting the relevance of currently unknown *H. parasuis* proteins in the formation of the biofilm extracellular matrix. An association between biofilm formation and resistance to β-lactam antibiotics has recently been established for *H. parasuis* (28), and a relationship between resistance to β-lactams and the presence of the pB1000 plasmid, bearing bla_{ROB-1}, has been established for *H. parasuis* clinical isolates (11). In this study, as part of the diversity encountered among nasal isolates, we found a novel small plasmid named pJMA-1. pJMA-1 was shown to bear bla_{ROB-1} and to be responsible for resistance to β-lactam antibiotics. Of note, pJMA-1 could be found in strains isolated from the nasal cavities of healthy animals sampled in different geographic regions in Spain but not in *H. parasuis* strains isolated from clinical lesions. pJMA-1 belongs to the ColE1 superfamily and shares a backbone common to other small plasmids described in the *Pasteurellaceae*, such as pB1000 (30), pB1002 (29), and pHS-Tet (52). Comparative analysis of the pJMA-1, pB1000, and pB1002 nucleotide sequences revealed that β-lactamase acquisition by plasmids pB1000/pB1002 and pJMA-1 seems to have occurred via two different transposition and integration events mediated by ISApl*1* within a conserved ColE1 backbone. Thus, the presence of ColE1 plasmids within a population could act both as a reservoir and as a mechanism for the dissemination of antibiotic resistance. Nucleotide analysis of pJMA-1 also showed a putative transfer origin (homologous to that described in pB1000), as well as a region of the *mobC* relaxase, which could mediate this dissemination. Interestingly, previous analysis of the bla_{ROB-1} gene distribution has shown the existence of a 2.6-kb plasmid in *A. pleuropneumoniae* and *P. multocida* swine isolates (62). We currently lack information to evaluate a potential relationship between pJMA-1 and this 2.6-kb plasmid found previously.

Besides its initial identification in *H. parasuis*, pB1000 has also been detected in *P. multocida* and nontypeable *H. influenzae* clinical strains (29, 30). Conversely, we could not detect pJMA-1 in *H. influenzae* strains isolated from healthy carriers or from clinical samples. In *H. influenzae*, high-level resistance to β-lactam antibiotics is mediated by the production of the β-lactamases TEM-1 and ROB-1, with TEM-1 being more prevalent than ROB-1 (63). Given its potentially low frequency, we cannot rule out the possibility that pJMA-1 is present in *H. influenzae*. Wider collections of *H. influenzae* β-lactamase-positive strains should be screened in future work to increase the probability of identifying pJMA-1 in clinical strains of human origin.

Furthermore, while plasmid pB1000 has a high biological cost (30) and has been found only in infected animals, plasmid pJMA-1 seems to have a weak impact on bacterial fitness (we acknowledge that fitness data for both pJMA-1 and pB1000 have been obtained in the heterologous strain *H. influenzae* Rd KW20) and has been found only in commensal populations. This means that in the absence of antibiotic pressure, the dynamics of plasmid loss in the population will be slow, while in the presence of anti-

biotic, bacteria bearing the plasmid will be selected, making the eradication of antibiotic resistance in the animals difficult. In this study, pB1000 was also found in some of the available colonizing strains (2 out of 106) but to a significantly lower extent than pJMA-1 (17 out of 106). pJMA-1 is smaller than pB1000, is stable, can be transferred to other bacterial species, and has a weak impact on bacterial fitness. We speculate that these features could be advantageous for pJMA-1 maintenance in colonizing bacteria as part of the respiratory microbiota of healthy animals and for its potential transfer to other pathogenic bacteria. In addition, we also identified the presence of two plasmids carrying the tetracycline resistance genes *tet*(B) and *tet*(H) in two nasal colonizers, and these will be the subject of future work.

In conclusion, this study emphasizes that the importance of commensal bacteria of the respiratory microbiota as a reservoir of antibiotic resistance in animals is largely unknown and remains to be further explored. These aspects have been tackled for the swine intestinal microbiome and were found to be relevant enough to be considered in agricultural cost-benefit analyses (64, 65). It becomes a critical issue when the genes providing resistance are borne on mobile elements, since commensal bacteria can interact with other bacteria, including zoonotic pathogens, and transfer resistance determinants. Altogether, the cautious and controlled use of antibiotics is needed, as it is the only effective action that can be taken to counteract the acquisition and dissemination of antibiotic resistance and the subsequent inefficacy of antibiotic therapies in animal and human health.

ACKNOWLEDGMENTS

We thank Saioa Burgui (Instituto de Agrobiotecnología) and Nuria Galofré-Milà (CReSA) for their technical help.

J.M. is funded by Ph.D. studentship BES-2013-062644 from the Ministerio Economía y Competitividad-MINECO, Spain. This work has been funded by grants from MINECO (SAF2012-31166) and the Departamento Industria Gobierno Navarra (IlQ14064.R12) to J.G., MINECO (AGL2013-45662) to V.A., and EU projects EvoTAR 282004-FP7 and EFFORT 613754-FP7 to B.G.-Z.

CIBERES is an initiative from ISCIII, Spain.

REFERENCES

1. **Cerda-Cuellar M, Naranjo JF, Verge A, Nofrarias M, Cortey M, Olvera A, Segalés J, Aragón V.** 2010. Sow vaccination modulates the colonization of piglets by *Haemophilus parasuis*. Vet Microbiol **145:**315–320. http://dx.doi.org/10.1016/j.vetmic.2010.04.002.
2. **Costa-Hurtado M, Aragón V.** 2013. Advances in the quest for virulence factors of *Haemophilus parasuis*. Vet J **198:**571–576. http://dx.doi.org/10.1016/j.tvjl.2013.08.027.
3. **Zhang B, Tang C, Liao M, Yue H.** 2014. Update on the pathogenesis of *Haemophilus parasuis* infection and virulence factors. Vet Microbiol **168:**1–7. http://dx.doi.org/10.1016/j.vetmic.2013.07.027.
4. **Oliveira S, Pijoan C.** 2004. *Haemophilus parasuis*: new trends on diagnosis, epidemiology and control. Vet Microbiol **99:**1–12. http://dx.doi.org/10.1016/j.vetmic.2003.12.001.
5. **Olvera A, Cerda-Cuellar M, Nofrarias M, Revilla E, Segalés J, Aragón V.** 2007. Dynamics of *Haemophilus parasuis* genotypes in a farm recovered from an outbreak of Glässer's disease. Vet Microbiol **123:**230–237. http://dx.doi.org/10.1016/j.vetmic.2007.03.004.
6. **Turni C, Blackall PJ.** 2010. Serovar profiling of *Haemophilus parasuis* on Australian farms by sampling live pigs. Aust Vet J **88:**255–259. http://dx.doi.org/10.1111/j.1751-0813.2010.00592.x.
7. **Mullins MA, Register KB, Brunelle BW, Aragón V, Galofre-Mila N, Bayles DO, Jolley KA.** 2013. A curated public database for multilocus sequence typing (MLST) and analysis of *Haemophilus parasuis* based on an optimized typing scheme. Vet Microbiol **162:**899–906. http://dx.doi.org/10.1016/j.vetmic.2012.11.019.
8. **Olvera A, Cerda-Cuellar M, Aragón V.** 2006. Study of the population structure of *Haemophilus parasuis* by multilocus sequence typing. Microbiology **152:**3683–3690. http://dx.doi.org/10.1099/mic.0.29254-0.
9. **Olvera A, Calsamiglia M, Aragón V.** 2006. Genotypic diversity of *Haemophilus parasuis* field strains. Appl Environ Microbiol **72:**3984–3992. http://dx.doi.org/10.1128/AEM.02834-05.
10. **Rafiee M, Bara M, Stephens CP, Blackall PJ.** 2000. Application of ERIC-PCR for the comparison of isolates of *Haemophilus parasuis*. Aust Vet J **78:**846–849. http://dx.doi.org/10.1111/j.1751-0813.2000.tb10507.x.
11. **San Millán A, Escudero JA, Catalán A, Nieto S, Farelo F, Gibert M, Moreno MA, Domínguez L, González-Zorn B.** 2007. β-Lactam resistance in *Haemophilus parasuis* is mediated by plasmid pB1000 bearing bla_{ROB-1}. Antimicrob Agents Chemother **51:**2260–2264. http://dx.doi.org/10.1128/AAC.00242-07.
12. **Olvera A, Pina S, Macedo N, Oliveira S, Aragón V, Bensaid A.** 2012. Identification of potentially virulent strains of *Haemophilus parasuis* using a multiplex PCR for virulence-associated autotransporters (*vtaA*). Vet J **191:**213–218. http://dx.doi.org/10.1016/j.tvjl.2010.12.014.
13. **Martínez-Moliner V, Soler-Llorens P, Moleres J, Garmendia J, Aragón V.** 2012. Distribution of genes involved in sialic acid utilization in strains of *Haemophilus parasuis*. Microbiology **158:**2117–2124. http://dx.doi.org/10.1099/mic.0.056994-0.
14. **Fu S, Yuan F, Zhang M, Tan C, Chen H, Bei W.** 2012. Cloning, expression and characterization of a cell wall surface protein, 6-phosphogluconate dehydrogenase, of *Haemophilus parasuis*. Res Vet Sci **93:**57–62. http://dx.doi.org/10.1016/j.rvsc.2011.07.006.
15. **Xu C, Zhang L, Zhang B, Feng S, Zhou S, Li J, Zou Y, Liao M.** 2013. Involvement of lipooligosaccharide heptose residues of *Haemophilus parasuis* SC096 strain in serum resistance, adhesion and invasion. Vet J **195:**200–204. http://dx.doi.org/10.1016/j.tvjl.2012.06.017.
16. **Zhang B, He Y, Xu C, Xu L, Feng S, Liao M, Ren T.** 2012. Cytolethal distending toxin (CDT) of the *Haemophilus parasuis* SC096 strain contributes to serum resistance and adherence to and invasion of PK-15 and PUVEC cells. Vet Microbiol **157:**237–242. http://dx.doi.org/10.1016/j.vetmic.2011.12.002.
17. **Zhou M, Zhang Q, Zhao J, Jin M.** 2012. *Haemophilus parasuis* encodes two functional cytolethal distending toxins: CdtC contains an atypical cholesterol recognition/interaction region. PLoS One **7:**e32580. http://dx.doi.org/10.1371/journal.pone.0032580.
18. **Mullins MA, Register KB, Bayles DO, Butler JE.** 2011. *Haemophilus parasuis* exhibits IgA protease activity but lacks homologs of the IgA protease genes of *Haemophilus influenzae*. Vet Microbiol **153:**407–412. http://dx.doi.org/10.1016/j.vetmic.2011.06.004.
19. **Wang X, Xu X, Wu Y, Li L, Cao R, Cai X, Chen H.** 2013. Polysaccharide biosynthesis protein CapD is a novel pathogenicity-associated determinant of *Haemophilus parasuis* involved in serum-resistance ability. Vet Microbiol **164:**184–189. http://dx.doi.org/10.1016/j.vetmic.2013.01.037.
20. **Zhang B, Feng S, Xu C, Zhou S, He Y, Zhang L, Zhang J, Guo L, Liao M.** 2012. Serum resistance in *Haemophilus parasuis* SC096 strain requires outer membrane protein P2 expression. FEMS Microbiol Lett **326:**109–115. http://dx.doi.org/10.1111/j.1574-6968.2011.02433.x.
21. **Zhang B, Xu C, Zhang L, Zhou S, Feng S, He Y, Liao M.** 2013. Enhanced adherence to and invasion of PUVEC and PK-15 cells due to the overexpression of RfaD, ThyA and Mip in the Δ*ompP2* mutant of *Haemophilus parasuis* SC096 strain. Vet Microbiol **162:**713–723. http://dx.doi.org/10.1016/j.vetmic.2012.09.021.
22. **Zhou S, He X, Xu C, Zhang B, Feng S, Zou Y, Li J, Liao M.** 2014. The outer membrane protein P2 (OmpP2) of *Haemophilus parasuis* induces proinflammatory cytokine mRNA expression in porcine alveolar macrophages. Vet J **199:**461–464. http://dx.doi.org/10.1016/j.tvjl.2013.12.010.
23. **Miniats OP, Smart NL, Rosendal S.** 1991. Cross protection among *Haemophilus parasuis* strains in immunized gnotobiotic pigs. Can J Vet Res **55:**37–41.
24. **Eaves LE, Blackall PJ, Fegan M.** 1989. Characterisation and antimicrobial sensitivity of haemophili isolated from pigs. Aust Vet J **66:**1–4. http://dx.doi.org/10.1111/j.1751-0813.1989.tb09701.x.
25. **Kofer J, Hinterdorfer F, Awad-Masalmeh M.** 1992. Occurrence and drug resistance of bacteria pathogenic to the lungs from autopsy material of swine. Tierarztl Prax **20:**600–604.
26. **Wissing A, Nicolet J, Boerlin P.** 2001. The current antimicrobial resistance situation in Swiss veterinary medicine. Schweiz Arch Tierheilkd **143:**503–510.
27. **de la Fuente AJ, Tucker AW, Navas J, Blanco M, Morris SJ, Gutiérrez-**

Martín CB. 2007. Antimicrobial susceptibility patterns of *Haemophilus parasuis* from pigs in the United Kingdom and Spain. Vet Microbiol **120:**184–191. http://dx.doi.org/10.1016/j.vetmic.2006.10.014.

28. **Zhang J, Xu C, Shen H, Li J, Guo L, Cao G, Feng S, Liao M.** 2014. Biofilm formation in *Haemophilus parasuis*: relationship with antibiotic resistance, serotype and genetic typing. Res Vet Sci **97:**171–175. http://dx.doi.org/10.1016/j.rvsc.2014.04.014.
29. **San Millán A, Escudero JA, Gutiérrez B, Hidalgo L, García N, Llagostera M, Domínguez L, González-Zorn B.** 2009. Multiresistance in *Pasteurella multocida* is mediated by coexistence of small plasmids. Antimicrob Agents Chemother **53:**3399–3404. http://dx.doi.org/10.1128/AAC.01522-08.
30. **San Millán A, García-Cobos S, Escudero JA, Hidalgo L, Gutiérrez B, Carrilero L, Campos J, González-Zorn B.** 2010. *Haemophilus influenzae* clinical isolates with plasmid pB1000 bearing bla_{ROB-1}: fitness cost and interspecies dissemination. Antimicrob Agents Chemother **54:**1506–1511. http://dx.doi.org/10.1128/AAC.01489-09.
31. **Aragón V, Cerda-Cuellar M, Fraile L, Mombarg M, Nofrarias M, Olvera A, Sibila M, Solanes D, Segalés J.** 2010. Correlation between clinicopathological outcome and typing of *Haemophilus parasuis* field strains. Vet Microbiol **142:**387–393. http://dx.doi.org/10.1016/j.vetmic.2009.10.025.
32. **Puig C, Martí S, Fleites A, Trabazo R, Calatayud L, Linares J, Ardanuy C.** 2014. Oropharyngeal colonization by nontypeable *Haemophilus influenzae* among healthy children attending day care centers. Microb Drug Resist **20:**450–455. http://dx.doi.org/10.1089/mdr.2013.0186.
33. **Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, McKenney K, Sutton G, FitzHugh W, Fields C, Gocyne JD, Scott J, Shirley R, Liu L-I, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghagen NSM, Gnehm CL, McDonald LA, Small KV, Fraser CM, Smith HO, Venter JC.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496–512. http://dx.doi.org/10.1126/science.7542800.
34. **Oliveira S, Galina L, Pijoan C.** 2001. Development of a PCR test to diagnose *Haemophilus parasuis* infections. J Vet Diagn Invest **13:**495–501. http://dx.doi.org/10.1177/104063870101300607.
35. **Del Río ML, Gutiérrez CB, Rodríguez-Ferri EF.** 2003. Value of indirect hemagglutination and coagglutination tests for serotyping *Haemophilus parasuis*. J Clin Microbiol **41:**880–882. http://dx.doi.org/10.1128/JCM.41.2.880-882.2003.
36. **Clinical and Laboratory Standards Institute.** 2013. Performance standards for antimicrobial disk and dilution susceptibility tests for bacteria isolated from animals; second informational supplement. CLSI document VET01-S2. Clinical and Laboratory Standards Institute, Wayne, PA.
37. **Clinical and Laboratory Standards Institute.** 2013. Performance standards for antimicrobial disk and dilution susceptibility tests for bacteria isolated from animals. Approved standard, 4th ed. CLSI document VET01-A4. Clinical and Laboratory Standards Institute, Wayne, PA.
38. **Dayao DA, Kienzle M, Gibson JS, Blackall PJ, Turni C.** 2014. Use of a proposed antimicrobial susceptibility testing method for *Haemophilus parasuis*. Vet Microbiol **172:**586–589. http://dx.doi.org/10.1016/j.vetmic.2014.06.010.
39. **Kaplan JB, Velliyagounder K, Ragunath C, Rohde H, Mack D, Knobloch JK, Ramasubbu N.** 2004. Genes involved in the synthesis and degradation of matrix polysaccharide in *Actinobacillus actinomycetemcomitans* and *Actinobacillus pleuropneumoniae* biofilms. J Bacteriol **186:**8213–8220. http://dx.doi.org/10.1128/JB.186.24.8213-8220.2004.
40. **Mason KM, Munson RS, Jr, Bakaletz LO.** 2003. Nontypeable *Haemophilus influenzae* gene expression induced *in vivo* in a chinchilla model of otitis media. Infect Immun **71:**3454–3462. http://dx.doi.org/10.1128/IAI.71.6.3454-3462.2003.
41. **Gutiérrez B, Escudero JA, San Millán A, Hidalgo L, Carrilero L, Ovejero CM, Santos-López A, Thomas-López D, González-Zorn B.** 2012. Fitness cost and interference of Arm/Rmt aminoglycoside resistance with the RsmF housekeeping methyltransferases. Antimicrob Agents Chemother **56:**2335–2341. http://dx.doi.org/10.1128/AAC.06066-11.
42. **Snyder L, Champness W.** 2007. Plasmids, p 208–209. *In* Molecular genetics of bacteria, 3rd ed. ASM Press, Washington, DC.
43. **Jin H, Zhou R, Kang M, Luo R, Cai X, Chen H.** 2006. Biofilm formation by field isolates and reference strains of *Haemophilus parasuis*. Vet Microbiol **118:**117–123. http://dx.doi.org/10.1016/j.vetmic.2006.07.009.
44. **Bouchet B, Vanier G, Jacques M, Auger E, Gottschalk M.** 2009. Studies on the interactions of *Haemophilus parasuis* with porcine epithelial tracheal cells: limited role of LOS in apoptosis and pro-inflammatory cytokine release. Microb Pathog **46:**108–113. http://dx.doi.org/10.1016/j.micpath.2008.10.008.
45. **Chen Y, Jin H, Chen P, Li Z, Meng X, Liu M, Li S, Shi D, Xiao Y, Wang X, Zhou Z, Bi D, Zhou R.** 2012. *Haemophilus parasuis* infection activates the NF-κB pathway in PK-15 cells through IκB degradation. Vet Microbiol **160:**259–263. http://dx.doi.org/10.1016/j.vetmic.2012.05.021.
46. **Lloyd DH.** 2007. Reservoirs of antimicrobial resistance in pet animals. Clin Infect Dis **45**(Suppl 2):S148–S152. http://dx.doi.org/10.1086/519254.
47. **Kruse H, Sorum H.** 1994. Transfer of multiple drug resistance plasmids between bacteria of diverse origins in natural microenvironments. Appl Environ Microbiol **60:**4015–4021.
48. **San Millán A, Giufre M, Escudero JA, Hidalgo L, Gutiérrez B, Cerquetti M, González-Zorn B.** 2011. Contribution of ROB-1 and PBP3 mutations to the resistance phenotype of a β-lactamase-positive amoxicillin/clavulanic acid-resistant *Haemophilus influenzae* carrying plasmid pB1000 in Italy. J Antimicrob Chemother **66:**96–99. http://dx.doi.org/10.1093/jac/dkq392.
49. **Tristram SG, Littlejohn R, Bradbury RS.** 2010. bla_{ROB-1} presence on pB1000 in *Haemophilus influenzae* is widespread, and variable cefaclor resistance is associated with altered penicillin-binding proteins. Antimicrob Agents Chemother **54:**4945–4947. http://dx.doi.org/10.1128/AAC.00263-10.
50. **Matter D, Rossano A, Sieber S, Perreten V.** 2008. Small multidrug resistance plasmids in *Actinobacillus porcitonsillarum*. Plasmid **59:**144–152. http://dx.doi.org/10.1016/j.plasmid.2007.11.003.
51. **Chang CF, Yeh TM, Chou CC, Chang YF, Chiang TS.** 2002. Antimicrobial susceptibility and plasmid analysis of *Actinobacillus pleuropneumoniae* isolated in Taiwan. Vet Microbiol **84:**169–177. http://dx.doi.org/10.1016/S0378-1135(01)00459-X.
52. **Lancashire JF, Terry TD, Blackall PJ, Jennings MP.** 2005. Plasmid-encoded Tet B tetracycline resistance in *Haemophilus parasuis*. Antimicrob Agents Chemother **49:**1927–1931. http://dx.doi.org/10.1128/AAC.49.5.1927-1931.2005.
53. **Liu J, Tan C, Li J, Chen H, Xu P, He Q, Bei W, Chen H.** 2008. Characterization of ISApl*1*, an insertion element identified from *Actinobacillus pleuropneumoniae* field isolate in China. Vet Microbiol **132:**348–354. http://dx.doi.org/10.1016/j.vetmic.2008.05.031.
54. **Tegetmeyer HE, Jones SC, Langford PR, Baltes N.** 2008. ISApl*1*, a novel insertion element of *Actinobacillus pleuropneumoniae*, prevents ApxIV-based serological detection of serotype 7 strain AP76. Vet Microbiol **128:**342–353. http://dx.doi.org/10.1016/j.vetmic.2007.10.025.
55. **Tomizawa JI, Ohmori H, Bird RE.** 1977. Origin of replication of colicin E1 plasmid DNA. Proc Natl Acad Sci U S A **74:**1865–1869. http://dx.doi.org/10.1073/pnas.74.5.1865.
56. **Foucault ML, Courvalin P, Grillot-Courvalin C.** 2009. Fitness cost of VanA-type vancomycin resistance in methicillin-resistant *Staphylococcus aureus*. Antimicrob Agents Chemother **53:**2354–2359. http://dx.doi.org/10.1128/AAC.01702-08.
57. **Amano H, Shibata M, Kajio N, Morozumi T.** 1994. Pathologic observations of pigs intranasally inoculated with serovar 1, 4 and 5 of *Haemophilus parasuis* using immunoperoxidase method. J Vet Med Sci **56:**639–644. http://dx.doi.org/10.1292/jvms.56.639.
58. **Amano H, Shibata M, Kajio N, Morozumi T.** 1996. Pathogenicity of *Haemophilus parasuis* serovars 4 and 5 in contact-exposed pigs. J Vet Med Sci **58:**559–561. http://dx.doi.org/10.1292/jvms.58.559.
59. **Brockmeier SL, Loving CL, Mullins MA, Register KB, Nicholson TL, Wiseman BS, Baker RB, Kehrli ME, Jr.** 2013. Virulence, transmission, and heterologous protection of four isolates of *Haemophilus parasuis*. Clin Vaccine Immunol **20:**1466–1472. http://dx.doi.org/10.1128/CVI.00168-13.
60. **Bouchet B, Vanier G, Jacques M, Gottschalk M.** 2008. Interactions of *Haemophilus parasuis* and its LOS with porcine brain microvascular endothelial cells. Vet Res **39:**42. http://dx.doi.org/10.1051/vetres:2008019.
61. **Costa-Hurtado M, Olvera A, Martínez-Moliner V, Galofre-Mila N, Martínez P, Domínguez J, Aragón V.** 2013. Changes in macrophage phenotype after infection of pigs with *Haemophilus parasuis* strains with different levels of virulence. Infect Immun **81:**2327–2333. http://dx.doi.org/10.1128/IAI.00056-13.

62. **Juteau JM, Sirois M, Medeiros AA, Levesque RC.** 1991. Molecular distribution of ROB-1 β-lactamase in *Actinobacillus pleuropneumoniae*. Antimicrob Agents Chemother **35:**1397–1402. http://dx.doi.org/10.1128/AAC.35.7.1397.
63. **Farrell DJ, Morrissey I, Bakker S, Buckridge S, Felmingham D.** 2005. Global distribution of TEM-1 and ROB-1 β-lactamases in *Haemophilus influenzae*. J Antimicrob Chemother **56:**773–776. http://dx.doi.org/10.1093/jac/dki281.
64. **Hansen KH, Damborg P, Andreasen M, Nielsen SS, Guardabassi L.** 2013. Carriage and fecal counts of cefotaxime M-producing *Escherichia coli* in pigs: a longitudinal study. Appl Environ Microbiol **79:**794–798. http://dx.doi.org/10.1128/AEM.02399-12.
65. **Looft T, Johnson TA, Allen HK, Bayles DO, Alt DP, Stedtfeld RD, Sul WJ, Stedtfeld TM, Chai B, Cole JR, Hashsham SA, Tiedje JM, Stanton TB.** 2012. In-feed antibiotic effects on the swine intestinal microbiome. Proc Natl Acad Sci U S A **109:**1691–1696. http://dx.doi.org/10.1073/pnas.1120238109.

Supplementary information

Table S1. Distribution of the *blaRob-1* gene and *blaRob-1* encoding plasmid in *H. parasuis* clinical strains isolated from Glässer disease lesions.

***H. parasuis* clinical isolates**	**Lesion**	***rob-1* presence**	***rob-1* containing plasmid size (Kb)**	**Country of isolation**
ER-P6	Pericarditis	-	-	Spain
9904108	Unknown	-	-	Denmark
2725	Arthritis	-	-	Germany
PC4-6P	Peritonitis	-	-	Spain
PV1-12	Pleuritis	-	-	Spain
228/04	Pneumonia	-	-	Spain
373/03A	Pericarditis	-	-	Spain
IT29205	Unknown	-	-	Italy
PO15/96	Pneumonia	-	-	Argentina
264/99	Arthritis	-	-	Spain
Nagasaki	Meningitis	-	-	Japan
61/03	Pneumonia	-	-	Spain
Ba9/09-6	Arthritis	-	-	Spain
112/02	Pericarditis	-	-	Spain
167/03	Pneumonia	-	-	Spain
P555/04	Pericarditis	-	-	Argentina
PO114	Pericarditis	-	-	Poland
2620	Pleuritis	-	-	Germany
CT175-L	Pneumonia	-	-	Spain
PG1-2M	Meningitis	-	-	Portugal
PG1-3M	Meningitis	-	-	Portugal
7204122	Arthritis	-	-	Denmark
7204226	Unknown	-	-	Denmark
9904574	Unknown	-	-	Denmark

Chapter 2. Transformed recombinant enrichment profiling rapidly identifies HMW1 as an intracellular invasion locus in *Haemophilus influenzae*.

Mell JC, Viadas C, **Moleres J**, Sinha S, Fernández-Calvet A, Porsch EA, St Geme JW 3rd, Nislow C, Redfield RJ, Garmendia J. PLoS Pathog. 2016 Apr 28;12(4):e1005576. doi: 10.1371/journal.ppat.1005576. eCollection 2016 Apr.

PLOS | PATHOGENS

RESEARCH ARTICLE

Transformed Recombinant Enrichment Profiling Rapidly Identifies HMW1 as an Intracellular Invasion Locus in *Haemophilus influenzae*

Joshua Chang Mell[1☯]*, Cristina Viadas[2☯], Javier Moleres[2], Sunita Sinha[3], Ariadna Fernández-Calvet[2], Eric A. Porsch[4], Joseph W. St. Geme, III[4], Corey Nislow[3], Rosemary J. Redfield[5], Junkal Garmendia[2,6]*

1 Department of Microbiology and Immunology, Institute for Molecular Medicine and Infectious Diseases, Center for Genomic Sciences, Drexel University College of Medicine, Philadelphia, Pennsylvania, United States of America, **2** Instituto de Agrobiotecnología, CSIC-Universidad Pública Navarra-Gobierno, Navarra, Spain, **3** Department of Pharmaceutical Sciences and the UBC Sequencing Centre, University of British Columbia, Vancouver, British Columbia, Canada, **4** Department of Pediatrics, Children's Hospital of Philadelphia, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **5** Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada, **6** Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Madrid, Spain

☯ These authors contributed equally to this work.
* joshua.mell@drexelmed.edu (JCM); juncal.garmendia@unavarra.es (JG)

Citation: Mell JC, Viadas C, Moleres J, Sinha S, Fernández-Calvet A, Porsch EA, et al. (2016) Transformed Recombinant Enrichment Profiling Rapidly Identifies HMW1 as an Intracellular Invasion Locus in *Haemophilus influenzae*. PLoS Pathog 12 (4): e1005576. doi:10.1371/journal.ppat.1005576

Editor: Nina R. Salama, Fred Hutchinson Cancer Research Center, UNITED STATES

Received: January 21, 2016

Accepted: March 23, 2016

Published: April 28, 2016

Data Availability Statement: All sequence data were deposited at NCBI under BioProject PRJNA308311. BioSample accessions are included in S2 and S3 Tables. Parental strains were submitted to the SRA as BAM files aligned to their own reference sequence. Recombinant pool and clone data were submitted to the SRA as BAM files aligned to the appropriate recipient reference sequence (Hi375 or Rd KW20).

Funding: This work was supported by National Institutes of Health Ruth Kirschstein postdoctoral fellowship F32AI084427 (to JCM); a Canadian

Abstract

Many bacterial species actively take up and recombine homologous DNA into their genomes, called natural competence, a trait that offers a means to identify the genetic basis of naturally occurring phenotypic variation. Here, we describe "transformed recombinant enrichment profiling" (TREP), in which natural transformation is used to generate complex pools of recombinants, phenotypic selection is used to enrich for specific recombinants, and deep sequencing is used to survey for the genetic variation responsible. We applied TREP to investigate the genetic architecture of intracellular invasion by the human pathogen *Haemophilus influenzae*, a trait implicated in persistence during chronic infection. TREP identified the HMW1 adhesin as a crucial factor. Natural transformation of the *hmw1* operon from a clinical isolate (86-028NP) into a laboratory isolate that lacks it (Rd KW20) resulted in ~1,000-fold increased invasion into airway epithelial cells. When a distinct recipient (Hi375, already possessing *hmw1* and its paralog *hmw2*) was transformed by the same donor, allelic replacement of $hmw2A_{Hi375}$ by $hmw1A_{86\text{-}028NP}$ resulted in a ~100-fold increased intracellular invasion rate. The specific role of $hmw1A_{86\text{-}028NP}$ was confirmed by mutant and western blot analyses. Bacterial self-aggregation and adherence to airway cells were also increased in recombinants, suggesting that the high invasiveness induced by $hmw1A_{86\text{-}028NP}$ might be a consequence of these phenotypes. However, immunofluorescence results found that intracellular $hmw1A_{86\text{-}028NP}$ bacteria likely invaded as groups, instead of as individual bacterial cells, indicating an emergent invasion-specific consequence of *hmw1A*-mediated self-aggregation.

Institute of Health Research operating grant (to RJR); Genome British Columbia grant SOF122 (to RJR and JCM); the Faculty of Pharmaceutical Sciences, Canadian Foundation for Innovation (to CN); National Institutes of Heath R01 grant DC002873 (to JWSG); and (to JG) grants from Ministerio Economía y Competitividad-MINECO SAF2012-31166 and SAF2015-66520-R, Dpto. Salud Gobierno Navarra 359/2012 and Ministerio de Educación PRX12/00191. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Author Summary

Many bacteria are naturally competent, actively taking up DNA from their surroundings and incorporating it into their genomes by homologous recombination. This cellular process has had a large impact on the evolution of these species, for example by enabling pathogens to acquire virulence factors and antibiotic resistances from their relatives. But natural competence can also be exploited by researchers to identify the underlying genetic variation responsible for naturally varying phenotypic traits, similar to how eukaryotic geneticists use meiotic recombination during sexual reproduction to create genetically admixed populations. Here we exploited natural competence, phenotypic selection, and deep sequencing to rapidly identify the *hmw1* locus as a major contributor to intracellular invasion of airway epithelial cells by the human pathogen *Haemophilus influenzae*, a trait that likely allows bacterial cells to evade the immune system and therapeutic interventions during chronic infections. Genetic variation in this locus can strongly modulate bacterial intracellular invasion rates, and possession of a certain allele favors adhesion and self-aggregation, which appear to prompt bacteria to invade airway cells as groups, rather than as individuals. Overall, our findings indicate that targeting HMW1 could block the ability of *H. influenzae* to invade airway cells, which would make antibiotic therapy to treat chronic lung infections more effective. Furthermore, our new approach to identifying the genetic basis of natural phenotypic variation is applicable to a wide-range of phenotypically selectable traits within the widely distributed naturally competent bacterial species, including pathogenesis traits in many human pathogens.

Introduction

Genetic mapping in bacteria historically relied on screening mutant libraries for loss-of-function mutations, followed by laborious isolation and identification of the disrupted loci. Recent innovations in mutagenesis approaches like TnSeq can accelerate the process and aid in characterizing gene function (*e.g.* [1,2]), yet such approaches have some limitations. For example: (a) many classes of genetic variation are not evaluated, (b) a suitable loss-of-function screen is typically required, and (c) such techniques ignore naturally occurring within-species phenotypic variation. An alternative is to emulate eukaryotic quantitative genetics approaches, which rely on sexual reproduction to map genetic variation. Rather than isolating the loci responsible for a specific phenotype with disruptive mutations, the QTL (quantitative trait locus) mapping approach identifies the loci and alleles that are directly relevant to phenotypic expression in natural populations.

Bacteria do not reproduce sexually, but genetic transfer mechanisms are widespread, and diverse bacterial species (including many important human pathogens) are naturally competent, able to actively take up and recombine homologous DNA from their surroundings into their chromosomes [3,4]. The value of this genetic transfer mechanism to researchers was seen as early as 1944 by Avery *et al.*, when naturally competent *Streptococcus pneumoniae* were used to show that DNA is the genetic material, or "the transforming principle" [5]. But only recently has exploiting natural competence (and other gene transfer mechanisms) become a practical means to investigate the genetic basis for natural phenotypic variation, as massively parallel sequencing technologies have become cost effective [6–9].

The Gram-negative bacterium *Haemophilus influenzae* has a well-characterized natural competence mechanism (reviewed in [4]) and illustrates how experimental natural transformation can be useful for genetic mapping. Under nutrient limitation, cells actively take up double-

stranded DNA from their environment through their cell envelope, and this DNA can replace homologous segments of the chromosome by recombination. In the laboratory, a competent *H. influenzae* cell will rapidly replace ~0.1–3% of its chromosome with genomic DNA from a divergent *H. influenzae* strain [8,10]. These replacements typically involve multiple independent recombination tracts and contain 100s to 1,000s of donor-specific single-nucleotide polymorphisms (SNPs) that span dozens of genes. Insertions and deletions readily transform as parts of longer recombination tracts, albeit with less efficiency than SNPs; this can add or remove whole genes and operons. Thus, a single transformation experiment can give millions of independently transformed recombinants containing all (or nearly all) of the genetic variation that distinguishes the donor and recipient strains [8]. Such pools can be screened or selected for donor phenotypes, and the donor-specific variation in isolated recombinants can be identified by DNA sequencing. A previous small-scale screen of <100 transformed recombinant clones identified a bacterial QTL with a ~10-fold effect on transformability itself, though this used laborious quantitative assays of individual clones that had already been sequenced [8].

Though typically a commensal of the nasopharynx, *H. influenzae*—especially in nonencapsulated or nontypeable forms (NTHi)—can cause middle ear infections (otitis media), community-acquired pneumonia, exacerbations of chronic obstructive pulmonary disease (COPD), conjunctivitis, and sometimes more severe invasive diseases [11]. Infections often persist and recur despite host production of bactericidal antibodies and the use of antibiotic therapy. Our understanding of the molecular mechanisms involved in the progression and persistence of *H. influenzae* infections remains limited, but identical strains have been repeatedly isolated from the lungs of COPD patients in serial clinic visits, suggesting that *H. influenzae* has traits that promote chronic infection [12,13].

Current evidence indicates that *H. influenzae* is a facultative intracellular pathogen, and host cell invasion may allow bacterial cells to temporarily evade the immune system and therapeutic interventions [14,15]. *H. influenzae* invades a variety of cell types [16–20], and viable NTHi have been found within host cells of adenoid tissues and bronchial biopsies [21,22]. After intracellular invasion, *H. influenzae* cells remains non-proliferative and resides within membrane-bound vacuoles with features of late endosomes [23,24] or freely within the cytoplasm [25], and intracellular bacteria eventually die after persisting for variable lengths of time [26]. While several host factors have been identified as important for intracellular invasion, much less is known about the bacterial factors responsible for this process.

We chose the intracellular invasion phenotype as a model to test the genetic mapping strategy described above, which we have named "transformed recombinant enrichment profiling", or TREP (summarized in Fig 1). We chose this phenotype because entry of *H. influenzae* into airways cells: (a) is easily selectable in lab culture, (b) displays wide phenotypic variation between clinical isolates, and (c) is likely to be an important factor in chronic infections. Application of TREP rapidly identified the *hmw1* adhesin as a factor crucial for intracellular invasion.

Results

Comparison of three diverse *H. influenzae* isolates

Three criteria were used to choose the donor and recipient strains: (a) substantially higher invasiveness of the donor over the recipients, (b) high natural transformability of the recipients, and (c) available genome references with many genetic markers distinguishing the donor from the recipients (S1 Text and S1 Fig). Three strains were assayed: the standard laboratory strain (Rd KW20, hereafter referred to as Rd) and two clinical isolates from pediatric patients with

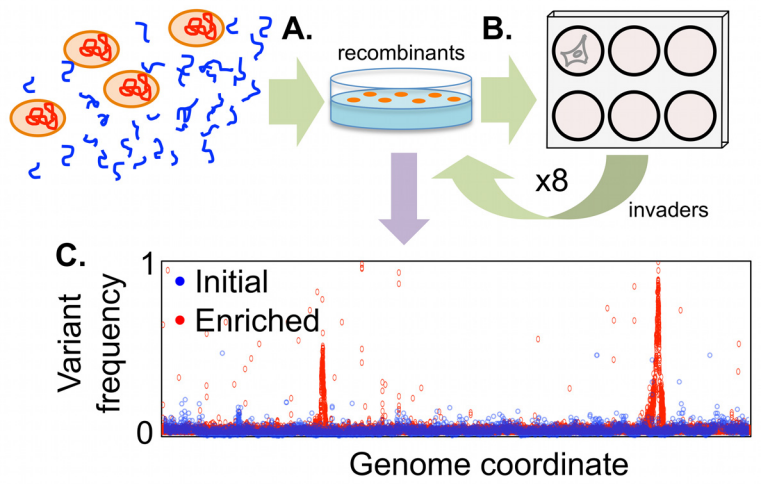


Fig 1. Schematic of transformed recombinant enrichment profiling (TREP) to map *H. influenzae* intracellular invasion genes. (A) Genomic DNA from a strain with high invasiveness is transformed into naturally competent cells of a strain with poor invasiveness. This produces complex pools of recombinants, in which short stretches of donor DNA replace homologous recipient DNA in individual recombinants. **(B)** Rare recombinants with increased invasiveness are enriched by serial passage of recombinant pools through A549 alveolar epithelial cells using gentamicin protection assays. Finally **(C)**, the selected locus/loci are identified from donor allele frequencies measured by deep sequencing.

doi:10.1371/journal.ppat.1005576.g001

otitis media (Hi375 and 86-028NP). All three have complete genome sequences available [27–29]. Rd and Hi375 are highly transformable, while 86-028NP is not [30,31]. Rd is known to be a poor invader of several epithelial cell lines [32]; Hi375 has previously been used in studies of intracellular invasion [15,24]; and 86-028NP is known to be highly virulent in a chinchilla model of otitis media [33–36]. Antibiotic resistant derivatives of all three strains were produced to allow subsequent tracking of transformation events and genetic background, yielding strains Rd Spc^R, Hi375 Str^R, and 86-028NP Nov^R Nal^R, hereafter referred to as RdS, HiT, and NpNN, respectively (S1 Table).

Intracellular invasion frequencies were evaluated by gentamicin protection assays with A549 airway epithelial cells, and quantified as gentamicin-protected bacterial colony forming units (CFU) relative to the original inoculated CFU (hereafter "Invaders/CFU"). Gentamicin protection assays found that NpNN is a highly efficient invader of A549 cells with ~10^{-2} invaders/CFU, whereas both HiT and RdS invade at ~100-fold and ~1,000-fold lower frequencies, respectively (Fig 2A; one-way ANOVA $p<<0.01$, and $p<<0.01$ for all three pairwise comparisons by *post hoc* testing using Tukey's HSD).

Several controls were performed, both for assays of the three parental strains and during the serial passage experiments described below: (a) To ensure that spontaneous gentamicin resistance was not responsible for survival in gentamicin protection assays, equivalent bacterial suspensions were treated in the absence of A549 cells, yielding no viable CFUs (limit of detection $<10^{-8}$), and gentamicin-protected CFUs recovered as colonies on plates remained gentamicin sensitive. (b) To ensure gentamicin treatment was complete, culture supernatants of infected A549 cells treated with gentamicin were plated, rendering no viable CFUs (limit of detection $<10^{-8}$). (c) To ensure that introduction of selectable markers had no effect on intracellular invasion, these strains were compared to their progenitors, finding that they had comparable phenotypes (S2A Fig, one-way ANOVA p-values > 0.1 within each strain background).

Adhesiveness of the three parent strains to A549 cells was assayed similarly to invasiveness, except that the gentamicin treatment was omitted; "Adherents/CFU" was calculated as the

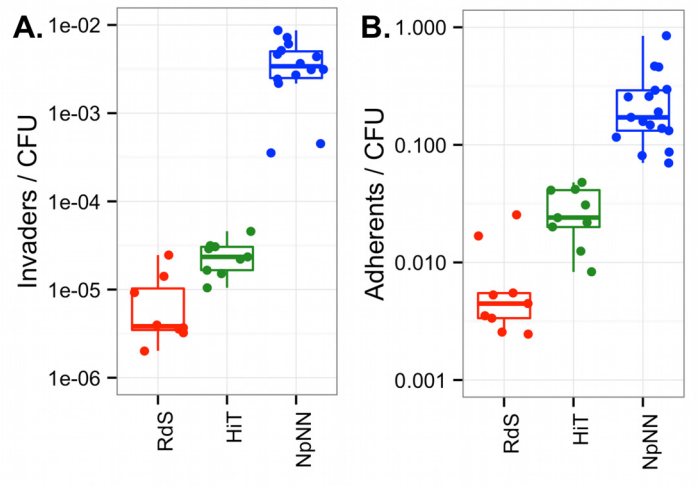


Fig 2. Invasion and adhesion in the parent strains. (A) "Invaders/CFU" was calculated as the total gentamicin-protected CFU / total input CFU. **(B)** "Adherents/CFU" was calculated as above with gentamicin treatment excluded. Triplicate experiments were conducted three times for RdS and HiT recipients and five times for the NpNN donor. Boxplots outline the first and third quartiles, with the thick horizontal line indicating the median, and the whiskers extend 1.5 times the interquartile distance.

doi:10.1371/journal.ppat.1005576.g002

CFU that remained associated with A549 cells after incubation and washing, relative to CFU of the input inoculum. The NpNN strain was the most adherent, with ~10% of the infecting cells remaining associated with A549 cells (Fig 2B). The HiT strain was intermediate (~10-fold lower than NpNN), whereas RdS had ~100-fold lower adherence ($p<<0.01$ by ANOVA and all for three pairwise comparisons). As with invasion, antibiotic resistant derivatives had adhesiveness comparable to that of the progenitors (S2B Fig).

Gentamicin protection strongly selects for intracellular invaders

A calibration experiment was used to test the strength of the experimental selection applied by the gentamicin protection assay. Mixtures of a high invasion strain (86-028NP Nov^R) and a low invasion strain (Rd Str^R) were passaged twice: Bacterial cells were used to infect A549 cells and invaders were recovered after gentamicin treatment. The recovered colonies were then pooled and used for a second infection. This found that the highly invasive strain easily outcompeted the poorly invasive one, even when starting at a 1 to 10,000 disadvantage, dominating the population after the second infection (S3 Fig). The high enrichment was not due to a growth rate advantage of 86-028NP, which has a slower doubling time than both recipients [30]. These data demonstrate that even rare recombinants could be highly enriched from complex pools using serial selection.

Summary of transformed recombinant enrichment profiling (TREP)

The experimental design for isolating bacterial intracellular invasion genes by TREP is depicted in Fig 1, summarized below, and described in detail in subsequent sections. (a) Donor genomic DNA from NpNN was used to transform naturally competent cells of two low invasion strains, RdS or HiT. (b) Pools of ~10^5 recombinant clones were enriched for those that conferred increased invasiveness by serial passages through A549 cells by gentamicin protection. Material from each cycle was stored to use for replicate quantitative assays and DNA extractions. (c) Genomic DNA from pools was sequenced to high coverage, and donor-specific allele

frequencies were calculated at diagnostic SNPs. Sequencing and alignment statistics are summarized in S2, S3 and S4 Tables, and further details are presented in S1 Text and Materials and Methods.

By profiling genome-wide donor allele frequencies from independent experiments, a gene responsible for high invasion by NpNN was rapidly identified in both recipients, RdS or HiT. Individual clones were isolated after the fourth cycle of selection for genome sequencing and phenotypic analysis of recombinants and mutant derivatives. The results reveal a novel role for the adhesin-encoding *hmw1A* gene in intracellular invasion, beyond its previously described role in adhesion [37,38].

Generation of complex recombinant pools of natural transformants

Recombinant pools were made by incubating high molecular weight genomic DNA from NpNN with naturally competent cultures of either RdS or HiT (Fig 1A) [39]. Transformation by the two antibiotic resistance markers (Nov^R and Nal^R) indicated that ~15% of cells in each culture were competent and predicted that a given donor-specific SNP would be found in ~1% of Nov^R or Nal^R colonies, consistent with previously measured values (S5 Table) [8]. Selection for donor-specific antibiotic resistance alleles ensured elimination of untransformed recipient cells, and selection for recipient-specific resistances limited cross-contamination observed in preliminary experiments (Materials and Methods). This procedure generated four separate sets of recombinant clones (for the RdS recipient: Spc^R Nov^R and Spc^R Nal^R; for the HiT recipient: Str^R Nov^R and Str^R Nal^R). For each set, ~10^5 colonies were harvested into pools, thoroughly mixed, and stored prior to infections. We thus predicted that each pool would have millions of cells for which (nearly) any given donor-specific variant would be present. Sequencing of the initial recombinant pools provided several lines of evidence that the TREP approach would be practical (S1 Text): First, a single round of selection for the donor-specific antibiotic resistance alleles was sufficient to map the resistances to single nucleotide resolution (S4A and S4B Fig and S6 Table). Second, recombinant pools contained low frequency donor-specific alleles across the genome, both SNPs and and structural variants (S4C Fig and S7 Table).

Serial selection by gentamicin protection enriches for invasive recombinants

The four initial recombinant pools (Pool 0s) were used in separate infections of A549 cells (Fig 1B). Serial passages of these pools using gentamicin protection resulted in >100-fold increased invasion after only three cycles of enrichment (S5 Fig). Replicate invasion assays using material from the initial enrichments confirmed the dramatic increase in intracellular invasion by pools after serial enrichment (Fig 3A, one-way ANOVA $p<<0.001$, Tukey's HSD for comparisons of Pool 0 to Pools 2–4 were $p<<0.001$, but $p>0.2$ among Pools 2–4). Comparison with donor invasiveness values measured in parallel showed that the HiT recombinants in the two Pool 4s were not significantly less invasive than the donor ($p = 0.082$); the two Pool 4s using the RdS recipient remained marginally less invasive ($p = 0.025$). These data suggest that after the third serial enrichment (Pool 4), donor-specific genetic variants conferring invasiveness were at or near fixation.

Increased invasion was not due to selection for *de novo* mutations that confer increased invasiveness, nor due to changes in bacterial gentamicin sensitivity (controls detailed above). Such events were unlikely given that the number of cell generations across the experiment was <100 in total. Furthermore, a control experiment using untransformed RdS or HiT cultures found no significant increases in invasiveness over 5 serial cycles of selection with either

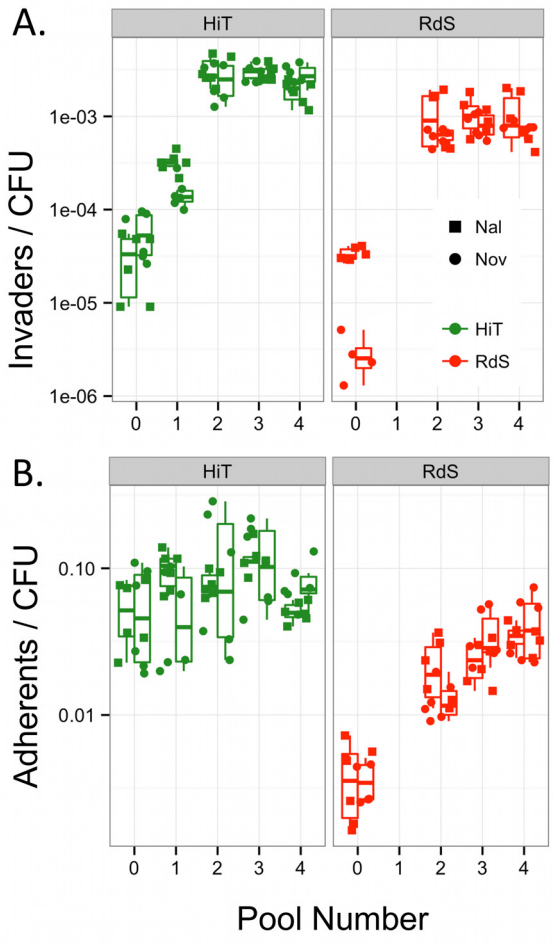


Fig 3. Serial enrichment for invasive recombinants by gentamicin protection. (A) Boxplots of intracellular invasion frequencies from Pools 0–4 (2 experiments in triplicate). **(B)** Boxplots of adhesion frequencies (2 experiments in triplicate). The values show the combined ability of clones in Pool *n* to invade or adhere to airway epithelial cells, while the recovered colonies comprise Pool *n*+1. Nal^R and Nov^R pools are summarized by the left and right boxplots, with individual data points shown as circles and squares, respectively.

doi:10.1371/journal.ppat.1005576.g003

recipient strain (S6 Fig; p>0.1 by one-way ANOVA), strongly suggesting that *de novo* mutations conferring increased invasiveness were not captured in these experiments.

Since higher adhesion might concomitantly increase intracellular invasion, we tested how adhesion was affected by the serial selections (Fig 3B). This found that the RdS pools showed substantial progressive increases in adhesiveness (one-way ANOVA p<<0.001, Tukey's HSD p<<0.001 for comparisons of Pool 0 to Pools 2–4, but p > 0.2 for comparisons among Pools 2–4). The HiT pools trended towards increasing adhesion over serial enrichments, but for this experiment, no significant change was observed (p>0.1). Both pairs of pools still had significantly lower adhesiveness than NpNN cultures run in parallel (p<0.05 for all comparisons against NpNN).

In sum, RdS and HiT recombinants acquired loci or alleles that enhanced intracellular invasion. Adhesion increased, though to a lesser extent. This suggests that, while adhesion might be a prerequisite for invasion, its increase may be insufficient to explain the elevated invasion displayed by invasiveness-enriched recombinants.

Serial enrichment for intracellular invaders selects for overlapping donor segments

Sequencing of genomic DNA across pools and serial passages (Fig 1C) showed that the four recombinant pools became progressively less complex, ultimately resulting in a total of only six recombinant clones dominating the four pools (out of ~$4x10^5$ total). This suggests that the causative alleles transformed competent cells at lower rates than typical SNPs (~1,000-fold; S5 Table). The change in complexity was particularly apparent at the antibiotic-selected sites, where donor allele frequencies shifted from a smooth decline on either side of the resistance alleles to sharply demarcated donor segments (stretches of contiguous donor-specific variation) supporting the presence of only 1 or 2 antibiotic resistance-spanning segments dominating each pool (S7 Fig).

Coincident with decreasing complexity at the antibiotic-selected loci, serial enrichment also increased the frequency of donor segments in other intervals (Figs 4 and 5). At the end of selection, 1–2 recombinant clones dominated each pool, each carrying several donor segments (one segment in each clone spanning the antibiotic resistance allele).

Overlapping donor segment intervals between independent recombinants and pools are the best candidates for carrying invasion loci (*i.e.* the purple circles in Figs 4 and 5). In principle, donor-specific genetic variation found in only some clones (those seen with intermediate frequencies in the pools) could potentially modulate intracellular invasion, but "hitchhiking" segments that are not associated with invasion are expected to occur, since previous work has shown that competent cells typically take up and recombine multiple donor DNA molecules. Similarly, independent recombination tracts carrying the same invasion locus are expected to typically have independent recombination breakpoints [8,10].

All invasion-enriched RdS recombinants had acquired the *hmw1*$_{86\text{-}028NP}$ operon

Sequencing of invader-enriched recombinant pools identified a single donor locus that was enriched to near-fixation in all four TREP experiments: *hmw1*$_{86\text{-}028NP}$ (Fig 6). For the RdS recipient, a narrow interval reached near-fixation for both Nov^R and Nal^R-selected pools (Figs 4 and 6A; Rd coordinates 1,744,519–1,750,336 nt, accompanied by a short nearby interval at 1,760,431–1,760,794 nt with lower levels of enrichment). The Rd gene annotations in this interval have no obvious connection with host cell interactions, but the donor strain carries a large operon here that is absent from Rd: *hmw1ABC*, which is found in ~60% of *H. influenzae* strains [40–43] and located between *yrbI* and *HI1680* (also known as *NTHI1986* in the 86-028NP genome). Depth-of-coverage analysis and reciprocal mapping of sequence reads to the donor genome confirmed the insertion of *hmw1ABC*$_{86\text{-}028NP}$ into the invasion-enriched RdS recombinant pools. Three independent recombination-mediated insertions of *hmw1ABC*$_{86\text{-}028NP}$ dominated the two invasion-enriched pools: one recombinant clone in the Nov^R pool and two in the Nal^R pool, as indicated by the distinct recombination breakpoints flanking the insertion (Fig 4). Additional recombination tracts were detected distant from the putative invasion locus and the antibiotic resistance marker, but these were unique to one of the three Pool 4 clones, as expected for random "hitchhiking" recombination events.

Invader-enriched HiT recombinants had substituted their *hmw2A*$_{Hi375}$ allele with the donor allele *hmw1A*$_{86\text{-}028NP}$

Selecting for invaders from the HiT recombinant pools likewise enriched for donor segments containing *hmw1*$_{86\text{-}028NP}$ adjacent to the *yrbI* gene (Figs 5 and 6B, S9 Table). Only two donor-

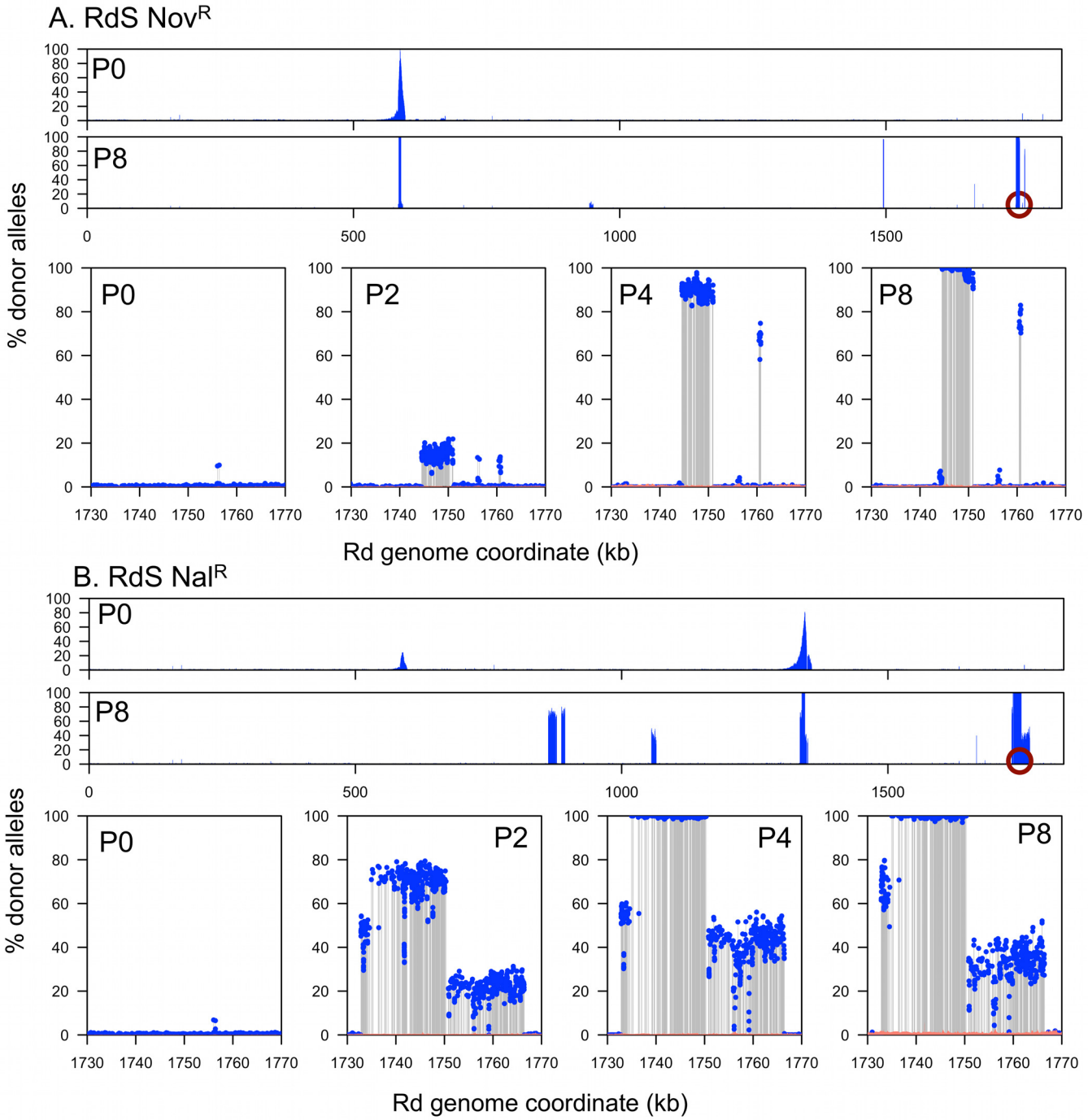


Fig 4. Genomic profile of the RdS recombinant pools. (A) RdS Nov^R. **(B)** RdS Nal^R. The x-axes indicate recipient genome coordinate, and the y-axes indicate the percent donor alleles. Blue dots and grey lines indicate donor-specific SNPs. Salmon color shows the limits of detection, or 1/depth (alignments with mapping quality = 0 were ignored, which excludes multiply mapping reads). Pools are indicated as P*n*. The two wide views at the top of **A** and **B** show the genome-wide donor allele frequencies for Pool 0 and Pool 8, whereas the lower four zoom around the nearly fixed locus selected for by serial gentamicin protection assay (indicated by purple circles) for Pools 0, 2, 4, and 8.

doi:10.1371/journal.ppat.1005576.g004

specific intervals reached fixation in HiT Nov^R: one spanning *gyrB* as expected, and another that completely spanned only the $hmw1A_{86\text{-}028NP}$ gene carrying only short segments of the

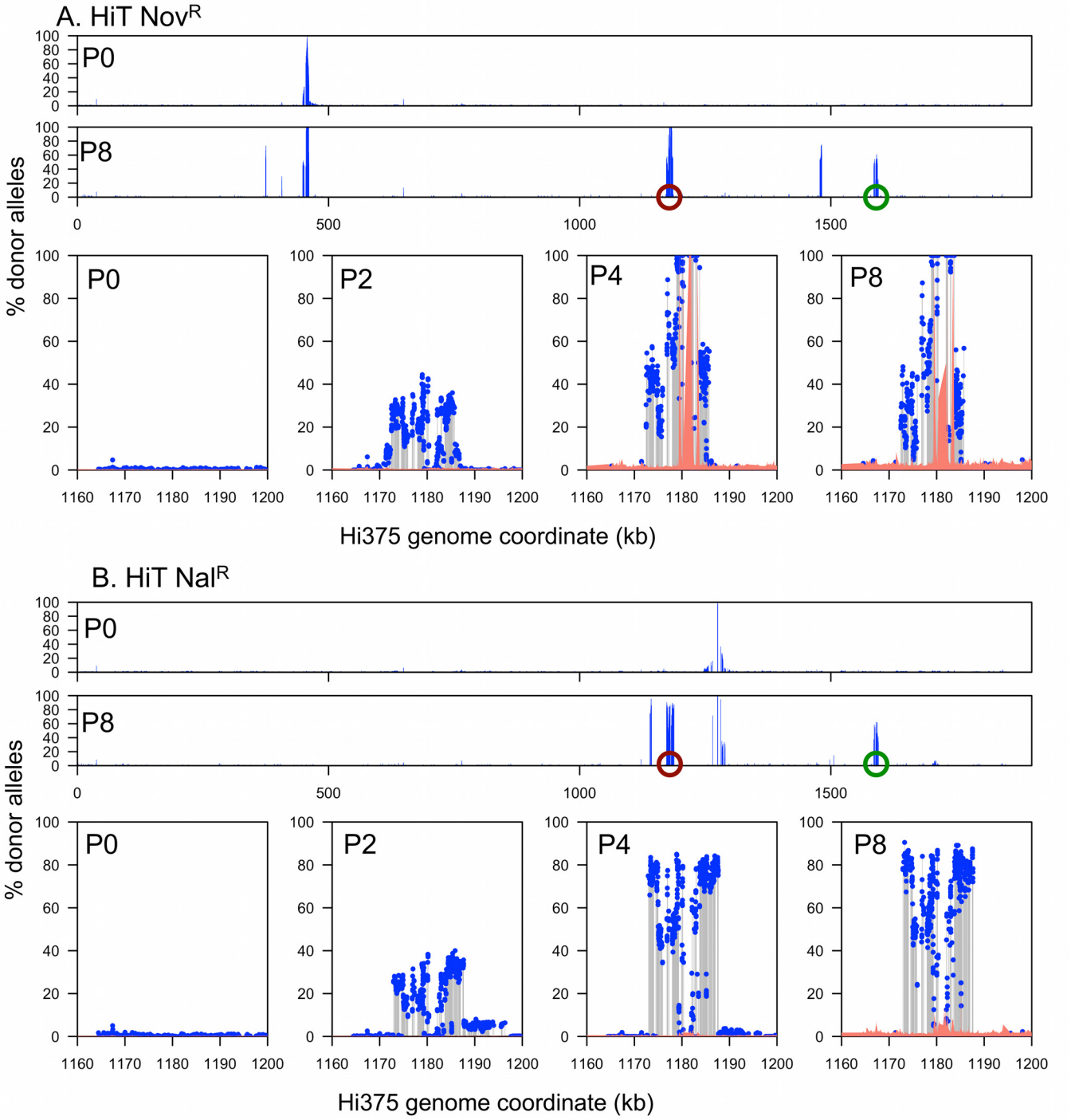


Fig 5. Genomic profile of the HiT recombinant pools. (A) HiT Nov^R. **(B)** HiT Nal^R. Plotting as in Fig 4, with the green circles indicating a multi-mapping artifact. Low limits of detection (1/depth, salmon color) within the invasion locus are due to alignments with mapping quality = 0 (multiply mapping reads) caused by the same sequencing artifact.

doi:10.1371/journal.ppat.1005576.g005

flanking genes (Hi375 genomic coordinates 1,178,771–1,183,728 nt). In HiT Nal^R, a long donor segment spanning this same interval reached ~75%.

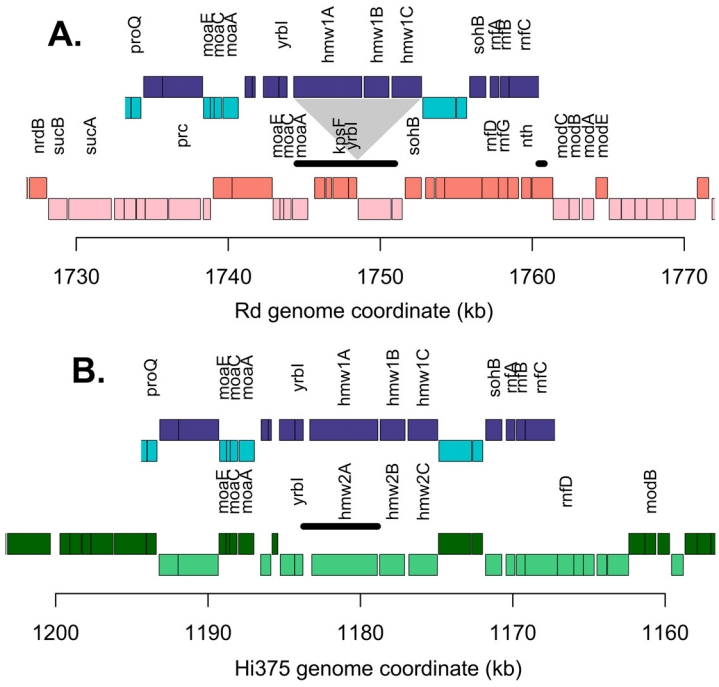


Fig 6. Genomic map at the enriched invasion locus for the three parental strains. For A and B, blue boxes indicate 86-028NP annotations; red is for Rd annotations, and green is for Hi375 annotations. Darker coloring indicates that sense is on the top strand, whereas lighter indicates bottom strand. Genes with names are indicated. The thick black horizontal lines indicate the minimum interval at or near fixation in both the Nov^R and Nal^R recombinant pools. **(A)** RdS recipient. Triangle indicates site of the *hmw1* insertion. **(B)** HiT recipient.

doi:10.1371/journal.ppat.1005576.g006

In contrast to RdS, the HiT recipient already possesses the *hmw1* operon and its paralogous operon *hmw2*, but the locations of the two adhesin genes (*hmw1A* and *hmw2A*) are swapped relative to their location in the donor NpNN. This is made evident through a comparison of the binding domains at the two *hmw* adhesin-encoding loci in HiT, NpNN, and the prototypic

Table 1. HMW adhesin paralog assignment.

Locus	Strain	$HMW1_{strain12}$	$HMW2_{strain12}$	Best Hit
YrbI- adjacent	Strain 12	*	33.8%	HMW1
	86-028NP	39.8%	34.8%	HMW1
	Hi375	35.8%	50.7%	HMW2
RadA- proximal	Strain 12	33.8%	*	HMW2
	86-028NP	32.8%	55.4%	HMW2
	Hi375**	100.0%	33.8%	HMW1

Percent amino acid identity of six HMW binding domains to the prototype HMW1 and HMW2 paralogs in Strain 12, based on ClustalW2 alignments. Shared alignment gaps were excluded from pairwise comparisons. Self-alignments are indicated by *. The binding domains were defined by Strain 12 sequences as amino acids 555–914 in $HMW1A_{strain12}$ and 553–916 in $HMW2A_{strain12}$.

** The NCBI PGAAP annotation for $hmw1A_{Hi375}$ ends with a serine instead of a stop codon and indicates a pseudogene. However, western analysis confirms its expression, and the presence of a stop codon 804 bp downstream accompanied by the expected pair of conserved C-terminal cysteine residues (codons 1624 and 1634) correctly defines the $hmw1A_{Hi375}$ CDS as from coordinates 1,585,978 to 1,590,906 and encoding a 168 kDa protein.

doi:10.1371/journal.ppat.1005576.t001

HMW-positive strain 12 (the strain that HMW adhesins were originally identified in and where they have been most extensively characterized, *aka* R2846) (Table 1). In strain 12 and the donor NpNN, *hmw1* is adjacent to the *yrbI* gene (*NTHI1982*) and *hmw2* is nearby the *radA* gene (*NTHI1453*). However, for Hi375, the *radA*-adjacent *hmw* adhesin has a binding domain with 100% amino acid identity to the *yrbI*-adjacent *hmw1A* gene from strain 12. Thus, whereas the $hmw1_{86\text{-}028NP}$ operon was inserted into invasion-selected RdS recombinants, recombination events that increased HiT invasiveness were the result of an allelic substitution of $hmw2_{Hi375}$ for $hmw1_{86\text{-}028NP}$.

The presence of the paralogous *hmw* locus nearby the *radA* gene caused an alignment artifact from multiply mapping sequence reads (Hi375 coordinates ~1,585–1,595 kb). Highly variable donor allele frequencies were seen across this interval ranging from ~0% to ~50% in Pool 8 for both HiT pools (S8 Fig). Furthermore, no donor-specific variation was detected flanking the *radA*-adjacent $hmw1A_{Hi375}$ locus, in contrast to donor variation flanking the recombinant *yrbI*-adjacent $hmw1A_{86\text{-}028NP}$ locus, which is expected for *recA*-mediated homologous recombination. Conclusive evidence for *radA*-proximal donor variation as a read alignment artifact is provided by allele-specific PCR assays on the isolated clones and mutants, as described below (S9 Fig).

Collectively, these data strongly support a role for $hmw1_{86\text{-}028NP}$ in the increased intracellular invasion seen in enriched recombinants, though they do not strictly rule out a role for flanking donor-specific variation, particularly in the *yrbI* gene, since donor-specific variation in this gene was also near fixation in the invader-enriched recombinant pools. Comparison of sequence variants called between Pool 0s and Pool 8s identified that no novel mutations were fixed during serial selections, consistent with low per base mutation rates and the control experiment with untransformed recipients shown in S4 Fig.

Individual recombinant clones validate and disambiguate the pool sequencing

Four colonies from Pool 4 were isolated from each of the four TREP experiments and further analyzed. Each of the 16 clones was assayed in triplicate for its invasion and adhesion phenotypes (Fig 7). Subsequent genome sequencing (summarized in S3 Table) revealed a total of only six distinct genotypes: three clones from each recipient background (Fig 7A; S9 Table). This is consistent with predictions from the pool data and shows that the isolated clones represented all the high frequency recombinants observed in Pool 4. Intracellular invasion and adhesion were strongly enhanced for all 16 isolated clones compared to recipient controls (Fig 7B and 7C, Tukey's HSD $p<<0.001$ for all six comparisons). Immunoblot analysis confirmed expression of $HMW1A_{86\text{-}028NP}$ protein in two recombinant clones, RdS genotype B (S10A Fig) and HiT genotype E (Fig 8A).

Not all of the recovered genotypes had identical phenotypes. Genotype D colonies (HiT Nal^R from Pool 4) were significantly more invasive and adherent than genotype E and F colonies (one-way ANOVA $p<<0.001$, Tukey's HSD gives $p<0.001$ for comparison of D against E or F, but $p>0.1$ for comparison of E and F). This indicates that donor-specific variation present in genotype D but absent from genotypes E and F slightly enhances adherence and invasion, albeit substantially less so than the *hmw1ABC* locus; the causative variation responsible remains unknown, but includes donor alleles of a QseBC-like two-component system (S9 Table, "D-specific" segments).

The sequencing of Pool 4 clones further supported a read alignment artifact at the *radA*-adjacent $hmw1A_{Hi375}$ locus. Colonies had been collected after re-streaking to ensure they represented single clonal lineages. Despite this, all three HiT clone genotypes (D, E, and F) from

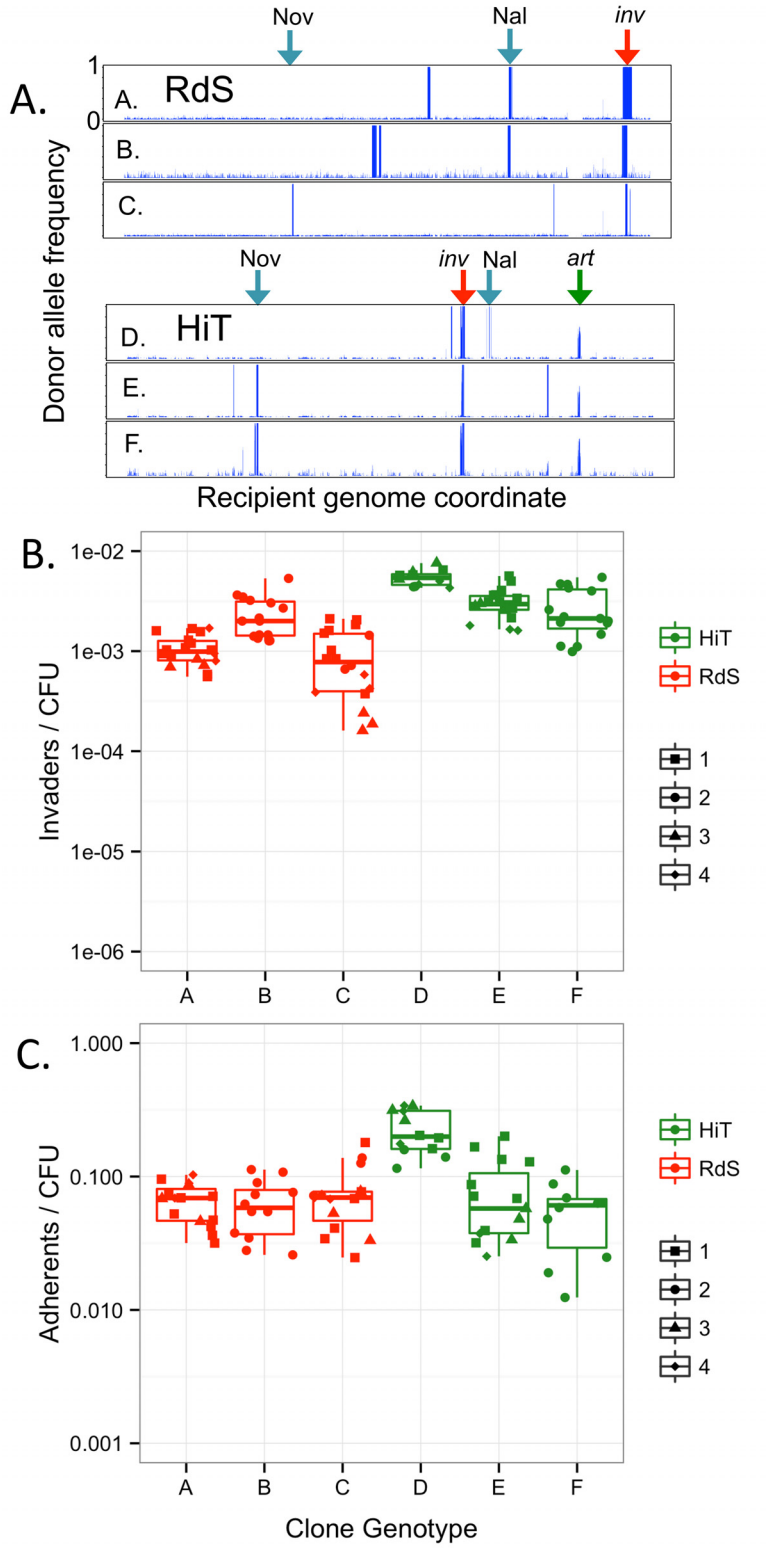


Fig 7. Genotype and phenotype of recombinant clones from Pool 4. Sequencing grouped 16 clones into six genotypes, two RdS Nal^R (A and B), one RdS Nov^R (C), one HiT Nal^R (D) and two HiT Nov^R (E and F). Genome-wide donor allele frequencies of each clone are shown in **(A)**. The x-axis indicates the recipient

genome coordinate and the y-axis shows donor allele frequency from 0 to 1. Blue arrows mark the antibiotic resistance sites; red arrows mark the invasion locus; and the green arrow marks the *hmw2* artifact. Clone assignments are in S8 Table and exact donor segment breakpoints are in S9 Table. All clones were phenotyped in triplicate and were aggregated by genotype for plots depicting their invasion **(B)** and adhesion **(C)** phenotypes. Different symbols within each genotype indicate the specific isolated colony tested (labeled 1–4, as in S8 Table).

doi:10.1371/journal.ppat.1005576.g007

Pool 4 showed a variable mixture of recipient- and donor-specific alleles across the *radA*-proximal *hmw* operon (Fig 6). In contrast, donor-specific allele frequencies observed at the *yrbI*-adjacent $hmw1A_{86\text{-}028NP}$ were near fixation and accompanied by flanking donor-specific variation. As a final confirmation that this mixed signal was not due to merodiploidy or other complex genetic effect, allele-specific PCR assays were conducted that distinguished all four *hmw* adhesin genes (the *yrbI*-proximal $hmw1A_{86\text{-}028NP}$ and $hmw2A_{Hi375}$ genes and the *radA*-proximal $hmw2A_{86\text{-}028NP}$ and $hmw1A_{Hi375}$ genes), and these PCR assays unambiguously showed that HiT recombinants had replaced their *yrbI*-adjacent $hmw2_{Hi375}$ allele with the $hmw1_{86\text{-}028NP}$ allele, whereas the *radA*-adjacent $hmw1_{Hi375}$ alleles were unchanged (S9 Fig).

Donor-specific variation in the *hmw1ABC* left flanking interval does not contribute to enhanced invasiveness

Despite the unambiguous acquisition of $hmw1_{86\text{-}028NP}$ in all recombinant clones, other donor-specific variation in the recombinant clones could conceivably be responsible. Since recombination tracts that carried $hmw1ABC_{86\text{-}028NP}$ also carried flanking donor-specific SNPs, we directly tested for a contribution by variation in the flanking interval. We cloned the 86-028NP alleles of the two genes upstream of *hmw1ABC*: *kpsF* and *yrbI*, encoding arabinose-5-phosphate isomerase and Kdo 8-phosphatase, respectively. Donor-specific variation in *yrbI* in particular might contribute to intracellular invasion, since the minimum interval that overlapped between all four experiments contained donor-specific variation in this gene (Fig 6). The resulting HA-tagged pSU20 plasmid (pSU20-*kpsF-yrbI*-HA) was then electroporated into Rd. Confirming expression from the plasmid, a ~19.3-kDa full-length $YrbI_{86\text{-}028NP}$-HA protein was detected in whole cell extracts by immunoblot with an anti-HA antibody (S10B Fig). Strains Rd, Rd pSU20, and Rd pSU20-*kpsF-yrbI*-HA were tested for intracellular invasion into A549 cells. No significant difference was observed among strains (S10C Fig, one-way ANOVA p-value = 0.29), thereby excluding a significant role for these flanking loci in intracellular invasion.

Mutation of $hmw1A_{86\text{-}028NP}$ confirms its role in intracellular invasion

Genetic confirmation of a role for $hmw1A_{86\text{-}028NP}$ in intracellular invasion (rather than other donor-specific variation acquired by recombinants) was performed using the HiT recipient and one of the invasive HiT recombinant clones (strain P551 with genotype E, hereafter called rHiT). We generated a panel of knockouts of the genes encoding HMW adhesins in the HiT and rHiT strains, either the locus adjacent to *yrbI* ($hmw2A_{Hi375}$ or $hmw1A_{86\text{-}028NP}$) or the locus nearby *radA* ($hmw1A_{Hi375}$). In the case of rHiT, the double mutant was also produced (both $hmw1A_{86\text{-}028NP}$ and $hmw1A_{Hi375}$ deleted). Western blot analysis confirmed expression of the expected HMW adhesins in each strain (Fig 8A).

All mutant strains were assayed for invasion in parallel with NpNN, HiT, and rHiT controls (Fig 8B, one-way ANOVA $p<<0.001$). We hypothesized that knocking out $hmw1A_{86\text{-}028NP}$ in the recombinant rHiT would show a strong defect in intracellular invasion, but that knocking out either *hmw* gene in the HiT recipient would have little or no effect. Indeed, deletion of

Strain background	NP NN	HiT	HiT	HiT	rHiT	rHiT	rHiT	rHiT
yrbI-adjacent	1A	2A	2A	Δ	1A	Δ	1A	Δ
radA-proximal	2A	1A	Δ	1A	1A	1A	Δ	Δ

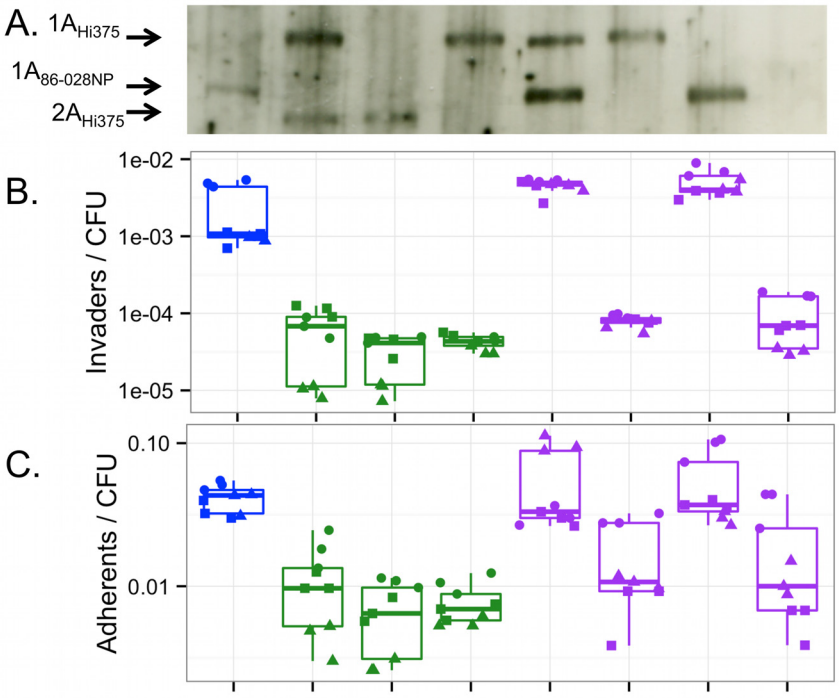


Fig 8. The *hmw1A*$_{86\text{-}028NP}$ gene confers increased adhesion and intracellular invasion. (A) Western blot detection of HMW adhesins with a guinea pig anti-HMW1A gp85 antibody. The 154 kDa protein corresponds to HMW1A$_{86\text{-}028NP}$; the 168 kDa protein corresponds to HMW1A$_{Hi375}$, and the 150 kDa protein corresponds to HMW2A$_{Hi375}$. Expression of HMW2A$_{86\text{-}028NP}$ was not detectable with the antibody used. **(B)** Intracellular invasion and **(C)** adhesion by NpNN and HiT derivatives including mutant, recombinant and recombinant mutant strains. Blue indicates NpNN, green indicates the HiT recipient and mutant derivatives, and purple indicates the rHiT recombinant and mutant derivatives. Experiments were run in triplicate on three separate days, as indicated by the three distinct symbols.

doi:10.1371/journal.ppat.1005576.g008

hmw1A$_{86\text{-}028NP}$ from rHiT reduced invasion frequencies 56-fold (Tukey's HSD $p<<0.01$) down to HiT recipient levels. By contrast, deletion of the *radA*-proximal *hmw1A*$_{Hi375}$ had no significant effect in either strain background, nor did deletion of either locus in the HiT recipient (Tukey's HSD $p>0.5$ for all comparisons).

These results confirmed the role of *hmw1A*$_{86\text{-}028NP}$ in the increased intracellular invasion of the HiT recombinant, and suggested that the HiT alleles of the HMW adhesins do not appreciably contribute to the HiT strain's ability to invade A549 cells. Parallel adhesion assays of the knockout panel found qualitatively similar results (Fig 8C, one-way ANOVA $p<<0.001$). The recombinant rHiT was significantly more adherent than the HiT recipient (4.7-fold increase, Tukey's HSD $p<0.001$)—comparable to adhesion by the donor NpNN ($p = 0.99$)—whereas deletion of *hmw1A*$_{86\text{-}028NP}$ from rHiT brought adhesion down 3.3-fold to HiT recipient levels (Tukey's HSD $p = 0.0014$ versus rHiT, $p = 0.99$ versus HiT).

The effect of *hmw1A*$_{86\text{-}028NP}$ on adhesion is >10-fold lower than its effect on intracellular invasion, such that significant increases in adhesion had not been detected in the original pool experiments and were only marginally significant in adhesion assays with the isolated recombinant clones (see above). Nevertheless, these results indicate that *hmw1A*$_{86\text{-}028NP}$ might contribute to intracellular invasion in part through an indirect effect of increasing adherence. While this is one contributing factor, the immunofluorescence data reported below indicate an unexpected intracellular invasion phenotype that cannot readily be explained by increased adherence alone.

Possession of *hmw1ABC*$_{86\text{-}028NP}$ confers a self-aggregation phenotype to recombinants

In the course of working with the parental strains, we noted that when cultures were left standing on the bench, NpNN (and 86-028NP) settled more quickly than RdS (and Rd), denoting a clumping or self-aggregation phenotype. Because self-aggregation could modulate bacterial-host cell interplay, we quantitatively tested this phenotype for a panel of strains: the three parents, an invasive recombinant RdS clone (P540, genotype B, hereafter rRdS), the rHiT recombinant, and the rHiTΔ*hmw1A*$_{86\text{-}028NP}$ mutant. This clearly demonstrated that *hmw1A*$_{86\text{-}028NP}$ plays a major role in the high self-aggregation seen in the recombinants (Fig 9; one-way ANOVA $p<<0.001$ at the t = 140 min time point and higher). Both recombinants settled quicker than the recipient

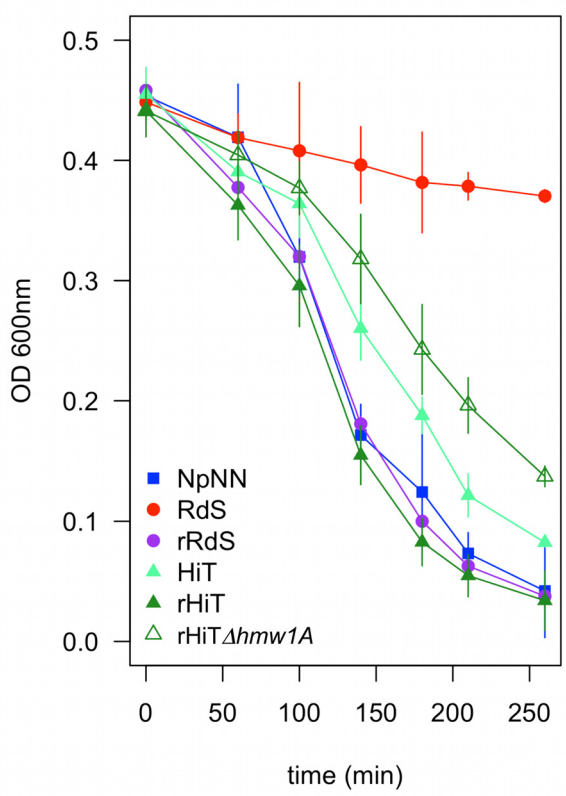


Fig 9. Self-aggregation is increased by *hmw1A*$_{86\text{-}028NP}$. Bacteria scraped from chocolate agar plates were suspended into 35 ml of sBHI normalized to OD_{600} = 0.5 in a 50 ml conical tube and allowed to sit on the lab bench. The OD_{600} at the top of the cultures was followed over time as a proxy for clumping/self-aggregation. Error bars indicate the standard deviation from four replicate assays run on different days.

doi:10.1371/journal.ppat.1005576.g009

strains (Tukey's HSD $p<0.001$) but were indistinguishable from NpNN ($p>0.6$). While the HiT recipient settled substantially faster than RdS, it and the rHiTΔ*hmw1A*$_{86\text{-}028NP}$ mutant settled slower than NpNN ($p<0.01$). These observations raised the question of how this clumping or self-aggregation phenotype modulates adhesion and intracellular invasion.

Immunofluorescence microscopy reveals that *hmw1ABC*$_{86\text{-}028NP}$ confers a novel aggregated intracellular bacterial invasion phenotype

To directly assess intracellular location of bacterial cells, we used immunofluorescence microscopy (Fig 10). As expected, the donor (NpNN) and two recombinant (rRdS and rHiT) strains infected A549 cells at high rates, whereas the recipients (RdS and HiT) and the mutant recombinant rHiTΔ*hmw1A*$_{86\text{-}028NP}$ infected A549 cells at substantially lower rates (Table 2). Co-localization with the endosomal marker Lamp-1 was substantial for all strains, indicating *bona fide* intracellular invasion, rather than gentamicin resistance or occlusion by A549 cells. These results confirm: (a) that both recipients (including RdS) successfully enter A549 cells, albeit at low rates; (b) that donor and recombinants have substantially higher invasion rates than recipients; and (c) that *hmw1A*$_{86\text{-}028NP}$ is responsible for elevated intracellular invasion rates in the rHiT recombinant.

The number of bacteria per infected cell was also distinct between strains. Cells infected by either recipient or the rHiTΔ*hmw1A*$_{86\text{-}028NP}$ mutant were infected with <10 bacteria/cell, whereas cells infected by the donor strain had >10 bacteria/cell. Although visually similar to the NpNN donor, scoring indicated that the rRdS and rHiT recombinants had an intermediate phenotype (Fig 10, Table 1). Whereas gentamicin protection assays with the recombinants gave invasion rates that approached donor levels, these data suggest that additional unidentified donor-specific factors besides *hmw1A*$_{86\text{-}028NP}$ may augment the ability of bacteria to invade.

Unexpectedly, bacterial invaders had a distinct pattern of co-localization with Lamp-1-positive endosomes for the donor and two recombinants. When infected by bacteria of either recipient strain or the rHiTΔ*hmw1A*$_{86\text{-}028NP}$ mutant, Lamp-1-positive subcellular compartments typically enclosed single bacterial cells, as previously observed [24]. In contrast, cells of the donor or either recombinant had enlarged Lamp-1-positive subcellular compartments that surrounded groups of bacteria. The size of these groups varied, with ~5–50 bacteria per compartment (Fig 10).

These results indicate that increased intracellular invasion by the donor and recombinants relates—at least in part—to internalized bacteria localizing as groups in the same subcellular compartment, rather than as individual cells. Previous work showed that *H. influenzae* invaders do not replicate [24] and, in our assays, the time between infection and observation was sufficiently brief that the groups of intracellular bacteria seen are unlikely to reflect intracellular replication. Hi375 has previously been seen to clump at the surface of A549 cells prior to invasion, but with only single bacteria undergoing internalization [24]. Thus, our self-aggregation and immunofluorescence results strongly suggests that groups of *hmw1ABC*$_{86\text{-}028NP}$ bacteria remain clumped during epithelial cell entry, thereby increasing overall invasion rates above and beyond the indirect effect of elevated adhesion, and also explaining the observed Lamp-1 reorganization around groups of bacteria.

Addition of *hmw1ABC*$_{strain12}$ to Rd increases intracellular invasion

Previous studies of *hmw1* focused on its role in adhesion, including detailed characterization of an Rd derivative carrying the *hmw1* operon from strain 12 with only minimal flanking variation from strain 12 [44]. We used the Rd *hmw1*$_{strain12}$ strain to independently test for the role

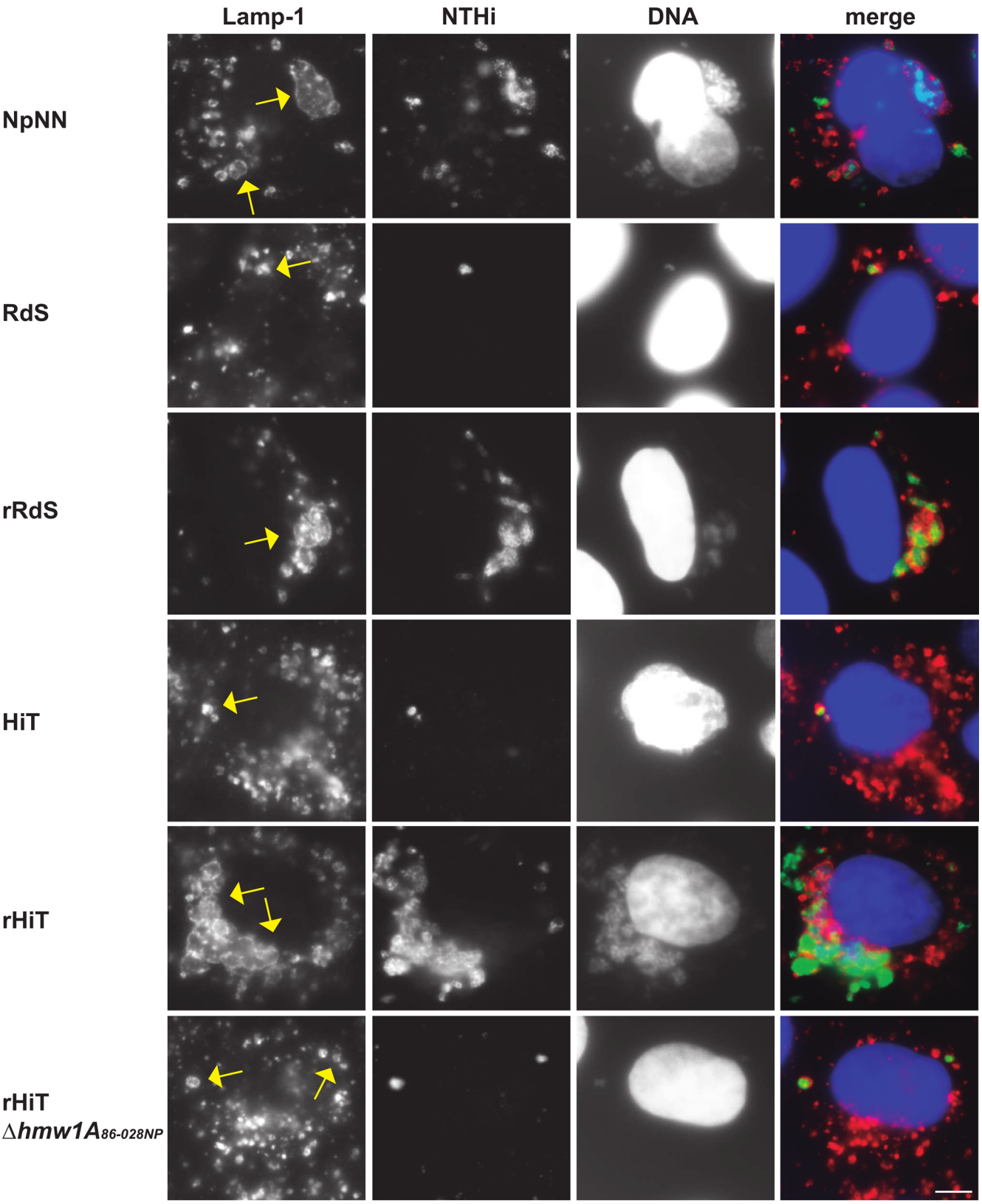


Fig 10. Co-localization of intracellular bacteria and the Lamp-1 endosomal marker. A549 cells were infected by NpNN, RdS, rRdS, HiT, rHiT, or rHiTΔ*hmw1A*$_{86\text{-}028NP}$. In the merged images, reactivity with an anti-NTHi antibody is shown in green, Lamp-1 stain is shown in red, and DNA stained with

Hoechst 33342 is shown in blue. Images were taken at 1 h post-gentamicin treatment. The scale bar in the lower right indicates 5 microns; individual bacterial cells are ~1 micron in diameter.

doi:10.1371/journal.ppat.1005576.g010

of *hmw1* in intracellular invasion. We evaluated invasiveness and adhesiveness of Rd, Rd $hmw1_{strain12}$, and strain 12 using A549 cells, as described above. Rd $hmw1_{strain12}$ had intermediate levels of both invasion and adhesion between both parents (Fig 11, p-values<0.01 for one-way ANOVA and all three comparisons by Tukey's HSD).

To test how these results depended on the specific protocol or cell type used, we evaluated invasion and adhesion following an alternative protocol in both A549 and Chang cells [18]. Rd $hmw1_{strain12}$ had significantly higher adhesion and invasion than Rd for both cell types, though the effect was much stronger with Chang cells (S11 Fig, p-value < 0.01 for all three comparisons, except $p = 0.054$ for adherence to Chang cells by Rd $hmw1_{strain12}$ and Strain 12). Rd $hmw1_{strain12}$ invaded Chang cells nearly as well as strain 12. In contrast, when infecting A549 cells, Rd $hmw1_{strain12}$ had an intermediate phenotype. These results indicate that *hmw1*'s contribution to invasion depends on the host cell type and confirm a more general role for *hmw1* in intracellular invasion beyond that seen for the 86-028NP allele.

Discussion

Is *hmw1* directly involved in *H. influenzae* intracellular invasion?

In contrast to bacterial pathogens with well-characterized intracellular life styles like *Salmonella*, *Listeria*, *Legionella* or *Brucella* [45–47], the mechanism and functional role of intracellular invasion in *H. influenzae* has been less well understood. It has been suggested that intracellular invasion of airway epithelial cells by non-typeable *H. influenzae* (NTHi) allows the bacterium to evade the immune system (antibodies, surfactant, antimicrobial peptides, galectins, professional phagocytes) and therapeutic interventions (antibiotics, anti-inflammatory agents), and to facilitate access to essential nutrients [14,48–50]. Thus, entry into airways cells may better equip bacterial cells for survival during long-term infections, particularly in the context of chronic infections that are often treated with intense antibiotic regimes.

Potential factors that contribute to invasion include those known to facilitate *H. influenzae*'s interactions with host cell surfaces. Among these are bacterial surface proteins that participate

Table 2. Intracellular bacteria immunofluorescence microscopy.

Strain	% infected (n = 250 cells)[a]	% bacteria:Lamp-1 co-localization (n = 150 infected cells)	% cells with # of infecting bacteria (n = 250 infected cells)	
			<10	>10
RdS (P532)	11.6	74.1	100	0
HiT (P531)	33.6	42.0	98.8	1.2
NpNN (P351)	100	88.0	0	100
rRdS (P540)	90.0	87.6	61.3	38.7
rHiT (P551)	80.8	93.9	76.2	23.8
rHiTΔ$hmw1A_{86\text{-}028NP}$ (P834)	44.7	65.3	92.1	7.9

[a]Frequencies are higher than from the plating assays above due to bacterial lysis during the saponin treatment that lyses A549 cells, as previously described [24].

doi:10.1371/journal.ppat.1005576.t002

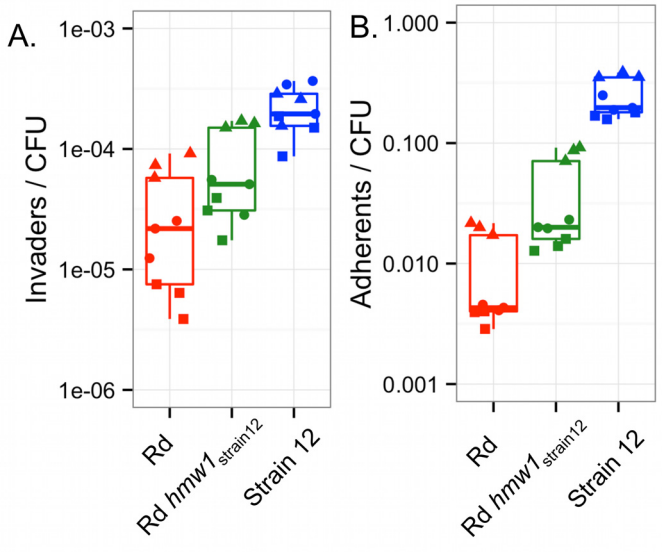


Fig 11. The *hmw1*$_{strain12}$ allele confers increased intracellular invasion to Rd KW20. (**A**) Invasion and (**B**) adherence phenotypes of Rd, Rd *hmw1*$_{strain12}$, and strain 12 were evaluated in A549 cells in triplicate on three separate days (shown by distinct symbols).

doi:10.1371/journal.ppat.1005576.g011

in *H. influenzae* binding to extracellular matrix proteins, mucin, or epithelial cells (including P5, OapA, PE, Hap, Hia, the HMW1 and HMW2 adhesins, IgA1 proteases A and B, and type IV pili [26,33,38,51–54]). Some factors, such as IgaA1 protease, appear to be more directly involved in *H. influenzae* entry into epithelial cells [26]. The PE and Hap adhesins have also been implicated in *H. influenzae* entry into epithelial cells [55,56]. While adherence to host cells may be a prerequisite for invasion, we lacked information on the specific involvement of adhesins or other factors that modulate intracellular invasion by *H. influenzae*. TREP was designed to identify invasion-promoting genes in an unbiased manner, but we were nonetheless surprised to isolate the well-characterized adhesin-encoding *hmw1* operon.

Gain of *hmw1*$_{86\text{-}028NP}$ by a poorly invading strain naturally lacking both *hmw1* and *hmw2* (RdS) dramatically enhanced both adhesion and invasion, showing that *hmw1*$_{86\text{-}028NP}$ is sufficient to confer high adhesion and invasion levels. When a recipient strain already possessing both *hmw1* and its paralog *hmw2* (HiT) was transformed, allelic substitution of the *hmw2A*$_{Hi375}$ allele with *hmw1A*$_{86\text{-}028NP}$ also strongly enhanced invasion. Characterization of recombinant clones and mutants lacking functional *hmw1*$_{86\text{-}028NP}$ confirmed these results. Important roles for other donor-specific variation carried by the transformed recombinants can be excluded, except that one of the HiT recombinants (genotype D) showed significantly higher adhesion and invasion than the others, suggesting at least one additional invasion-promoting factor, albeit a smaller contributor; further analysis will be of interest in future studies.

The striking intracellular phenotype we observed by immunofluorescence for *hmw1*$_{86\text{-}028NP}$ strains—in which groups of bacteria occupy engorged intracellular vesicles—suggests that increased adhesion alone is insufficient to fully explain the role of *hmw1* in intracellular invasion. We instead suggest that elevated invasion is an emergent property of HMW1-mediated self-aggregation; whereas adhesion is increased as an indirect result, we speculate that invasion by bacterial groups directly enhances overall invasion rates. Alternatively, possession of *hmw1*$_{86\text{-}028NP}$ may increase independent entry by bacteria into cells, followed by subsequent aggregation of bacterium-containing vesicles. Ruling out self-aggregation *per se*, recent results

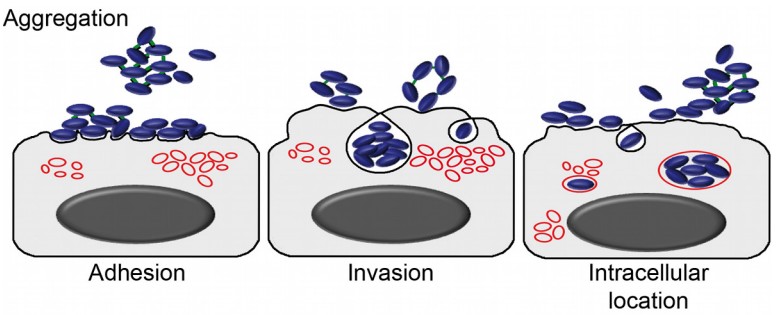


Fig 12. Model of enhanced intracellular invasion by HMW1$_{86\text{-}028NP}$. Bacteria self-aggregate and adhere to the epithelial cell surface in clumps, and HMW1-mediated attachments are maintained during uptake by cells into subcellular compartments with endosomal features.

doi:10.1371/journal.ppat.1005576.g012

show that deletion of the Hap autotransporter from Hi375 (naturally present in both 86-028NP and Hi375 but absent from Rd) eliminates self-aggregation, but epithelial cell adhesion and invasion were not significantly affected [57]. Altogether, we propose that, whether or not it can truly be called an "invasin", allelic variation in HMW1A affects invasion by increasing adhesion and also directly through a novel mechanism that allows for entry by groups of aggregated bacterial cells (model summary in Fig 12).

The HMW1 and HMW2 adhesins

The *hmw1* and *hmw2* operons are found in ~60% of *H. influenzae* isolates, and they always co-occur, despite being at different chromosomal loci (one adjacent to *HI1679* in Rd and the other to *HI1598*) [40,58,59]. Co-occurrence of the *hmw1* and *hmw2* loci in all tested clinical isolates suggests that the laboratory-created *hmw1*-only strains studied here must be at some unknown fitness disadvantage in nature.

The *hmw1* and *hmw2* operons are phase-variable, and expression is inversely correlated with the number of 7-bp tandem repeats found within their promoter regions [60–63]. Though subtle expression variation was not ruled out, western blot analysis indicated that HMW1 adhesin levels were mostly unchanged in recombinants (Fig 8A; 16 repeats upstream of $hmw1_{86\text{-}028NP}$ but 17 upstream of $hmw2_{Hi375}$). This indicates that allelic variation in the *hmw* coding sequences is likely responsible for differences in adhesion and invasion.

HMW1A and HMW2A display wide amino acid diversity both within and between isolates, with the region of lowest sequence identity in the host cell binding domain, which has been predicted to affect tissue tropism and immune evasion [42,59,63–67]. Phylogenetic analyses of the HMW adhesin binding domain has revealed four distinct sequence clusters, and the majority of sequences belonging to one of two dominant sequence clusters [41]. Of note, 86-028NP and strain 12 *hmw1A* binding domains belong to clusters 4 and 2, respectively, which might contribute to their strong and intermediate phenotypes; future studies using mosaic proteins with binding domains from distinct clusters and using multiple human cell lines could identify any clade-specific functions for HMW proteins.

The two *hmw* operons encode high molecular weight non-pilus adhesins (HMW1/HMW2) [37,64], along with two co-factor proteins encoded by the downstream genes. The co-factors are required for proper surface localization of HMW adhesins, and the paralogs of the co-factors are functionally interchangeable [37,58]. The *hmw1B/hmw2B* genes encode outer membrane pore-forming translocators that export HMW1 and HMW2 to the cell surface [68]. The *hmw1C/hmw2C* genes encode glycosyltransferases responsible for adding mono-hexose or di-

hexose residues at asparagines in conserved NX(S/T) motifs of HMW1 and HMW2 [69], likely involved in stabilizing the adhesins during or after their synthesis [70]. Due to the high diversity in HMW adhesin sequences, differential glycosylation patterns might in part be responsible for distinct activities of different alleles.

Another key distinction between the HMW1 and HMW2 adhesins is that the former recognizes sialylated glycoprotein receptors on cultured human epithelial cells [71]. HMW1 confers high adherence to Chang, Hep-2, HaCaT and NCI-H292 cells mediated by interactions with α-2,3 N-linked sialic acids. By contrast, HMW2 confers adherence to HaCaT and NCI-H292 cells via a sialic acid-independent mechanism [59,67,71]. This suggests that the role of $hmw1A_{86\text{-}028NP}$ in intracellular invasion may involve specific interactions between *H. influenzae* cells and sialylated glycoprotein receptors, both on the bacterial cell surface to mediate self-aggregation and possibly also specific host sialylated glycoproteins on epithelial cells. Glycoproteins play important roles in many cellular activities, and new methods for investigating their expression and sialylation states are being developed and applied to multiple cell types including A549 [72], opening new avenues to identify host glycoproteins hijacked by bacterial proteins such as HMW adhesins during the infection process.

To our knowledge, this is the first report of an involvement for *hmw1A* in *H. influenzae* intracellular invasion and, more strikingly, we further found that intracellular invasion is modulated by allelic diversity at *hmw1A*. Finding that $hmw1_{86\text{-}028NP}$ results in clumps of intracellular bacteria was unexpected and indicates that increased self-aggregation and adhesion *per se* are not sufficient to explain its effects, offering new avenues of investigation. Bacterial factors contributing to adhesion are already potential targets for antimicrobial therapies, and the additional role of HMW1 in intracellular invasion further increases its attractiveness as a target. Understanding the relationship between *hmw1* allelic variation and within-host adaptive evolution poses interesting challenges for future studies.

Transformed recombinant enrichment profiling

To better understand intracellular invasion by *H. influenzae*, we have successfully employed a gain-of-function genetic mapping strategy, TREP, which takes advantage of within-species phenotypic variation, natural competence, and deep sequencing. In total, our experiment isolated six highly invasive recombinants from a total of ~400,000 independent recombinants. Thus, while transformation rates of SNPs was much higher (*e.g.* ~0.4% for the antibiotic resistance alleles), this approach was able to isolate even very rare recombinants, in the case of the Rd strain requiring the insertion of a particularly long operon (>9kb). TREP proved to be a rapid method to map genes. Once the donor and recipient were assayed and the effectiveness of the selections was determined, the total hands-on time was only about six weeks, from generating the recombinant pools, performing the serial selections, extracting DNA, making libraries and sequencing, with serial selections comprising the most time-consuming step. Owing to the strong selection used, we found that strains that have slight advantages in invasion were able to overtake the pools after serial selection. Thus, it was crucial to add selectable markers to our recipient background and to ensure that the serial selections were performed without the donor strain or any other strains assayed in parallel.

The TREP method holds great promise for studying a wide range of traits that show natural phenotypic variation in other naturally competent species, which includes many virulence traits and pathogens important to human health. In contrast to screening/selecting clones transformed by plasmids [73], TREP does not depend on dominance, a suitable vector, nor is it restricted to monogenetic traits. The approach should be readily applicable to any selectable trait in any bacterial species for which a natural competence protocol has been developed, and

the number of such species continues to grow. Similar approaches have recently been reported in other organisms, for example to identify conjugation genes in *Mycobacterium* and causative alleles responsible for antibiotic resistance in *Streptococcus* [7,74]. Importantly, TREP is a general genetic mapping strategy agnostic to the type of variation (*i.e.* SNPs or whole loci can be identified), and we expand the utility of the transformation-based genetic mapping to include quantitative differences that go beyond absolute phenotypic differences (*i.e.* resistance *versus* sensitivity) by incorporating serial selection.

Traditionally, microbial experimental evolution studies rely on "hard" selective sweeps, in which newly arising beneficial mutations fix in a laboratory population [75]; more recently this has also included experimental evolution of pathogenic traits [76,77]. But "soft" sweeps, in which pre-existing genomic variation recombines within/into a population [78], may also play an important role in the adaptation of naturally competent species to new environments [79]. Allowing for introgression of natural variation has been used in experimental evolution in sexual eukaryotes (*e.g.* [80,81]), but its role in bacterial adaptation has been explored only recently and only in the context of standard experimental evolution studies that start with clonal populations [82,83]. Here, we found that rare recombinants generated in a single round of natural transformation could reach fixation after a small number of serial selections, illustrating the powerful contribution of natural competence to adaptation.

Finally, applying TREP to understand bacterial pathogenesis could use large "zoos" of donor genomic DNAs, rather than single donor-recipient combinations. This would better mimic the situation in chronic infections, where diverse polyclonal infections are common, and it would more fully sample the genomic diversity of these organisms in single experiments. However, caution must be exercised with such an approach: depending on the organism and trait under study, this could inadvertently generate novel hyper-virulent strains by combining multiple pathogenicity factors from different genetic backgrounds; a similar ethical concern has already been raised for studying pathogens using gain-of-function mutagenesis [84,85].

Materials and Methods

Bacterial strains

Bacterial strains and plasmids used are listed in S1 Table, and all PCR primers used are in S10 Table. General culturing and manipulation of *Haemophilus influenzae* followed standard methods [39]. Strains were grown at 37°C with 5% CO_2, on chocolate agar or brain heart infusion (BHI) supplemented with 10 μg/ml hemin and 10 μg/ml β-nicotinamide (sBHI). Antibiotics were added as required: novobiocin (Nov) at 2.5 μg/ml, nalidixic acid (Nal) at 3 μg/ml, spectinomycin (Spc) at 25 μg/ml, streptomycin (Str) at 100 μg/ml, chloramphenicol (Cm) at 2 μg/ml, and erythromycin (Erm) at 11 μg/ml. *Escherichia coli* strains were grown at 37°C on Luria Bertani (LB), and Cm at 30 μg/ml was added as required.

Donor and recipient strains. Donor genomic DNA was obtained from P351, a derivative of the non-typeable otitis media isolate 86-028NP that also carries Nov^R and Nal^R alleles of *gyrB* and *gyrA* from the multi-antibiotic resistant strain MAP7 [10,28,39] (hereafter NpNN). Two recipient strains were used for natural transformation by this donor DNA: P532, a Spc^R derivative of the laboratory strain Rd KW20 [27] (hereafter RdS), and P531, a Str^R derivative of the non-typeable otitis media isolate Hi375 [29,86] (hereafter HiT). RdS was produced by transformation of Rd with MAP7 genomic DNA, selection for Spc^R, followed by screening against other MAP7 resistance alleles. HiT was produced by transformation of Hi375 with a PCR amplicon spanning the Str^R allele of the *rpsL* gene from an Rd KW20 Str^R strain (P193) produced with "RdS" primers 1497+1798, and selection of transformants on sBHI-agar containing Str 100 μg/ml.

Cloning of the HMW1 flanking interval. To test whether the region flanking *hmw1*$_{86\text{-}028NP}$ played a role in invasion, an interval encompassing *NTHI1981* (*kpsF*) and *NTHI1982* (*yrbI*), and 675 bp upstream, was PCR amplified using 86-028NP genomic DNA as template and "cloning" primers 1219 and 1218. Primer 1218 included an HA tag to add at the 3′end of the *yrbI* gene. This 2,241bp blunt PCR product was phosphorylated with T4 kinase, and cloned into pSU20 [87] pre-digested with *Hinc*II and dephosphorylated with Antarctic phosphatase, generating pSU20-*kpsF-yrbI*-HA. pSU20 and pSU20-*Pr*::*kpsF-yrbI*-HA were transformed into electrocompetent Rd. Tranformants were selected by plating to sBHI-agar containing Cm.

HMW adhesin knockouts. Deletion mutations of *hmw* adhesins were generated using one of two approaches. In the first recombineering approach, the first 1kb of the *hmw1A*$_{86\text{-}028NP}$ gene (7,000bp) was replaced with a Spc^R cassette. This was used to generate an *hmw1A*$_{86\text{-}028NP}$ knock-out in the rHiT recombinant clone (P551). Briefly, a ~3kb interval was amplified from NpNN using "interval" primers 13CAN+1273, which spanned the 1kb region targeted for deletion ±1kb, and the purified amplicon was cloned into the pGEMT-easy vector. Separately, a Spc^R selectable marker was amplified from pRSM2832 [88] using primers carrying 50 bp overhangs flanking the deletion target ("Deletion" primers 1199 and 1274). The plasmid and amplicon were co-electroporated into DY380/SW102, which carries a heat-shock inducible λ Red recombinase. After recombinase induction, recovery, and selection for Amp^R Spc^R resistant *E. coli* colonies [88], disruption cassettes were confirmed by PCR. Finally, the "interval" primers (13CAN and 1273) were used to amplify the complete disruption cassette from the plasmid prior to natural transformation into rHiT and selection on Spc^R. Mutants were selected on sBHI-agar containing Spc, and correct targeting was confirmed by PCR and western blot. This procedure generated strain P834.

In the second approach, PCR was used to amplify two ~1kb intervals flanking the region targeted for deletion (Flanks A and B), and an erythromycin resistance cassette (Erm^R) was added between them. This was used to generate knockouts in HiT (P531), rHiT (P551), and rHiT Δ*hmw1A*$_{86\text{-}028NP}$ (P834) in two genes: *hmw1A*$_{Hi375}$ (Flank A primers 1463 and 1464; Flank B primers 1465 and 1466) and *hmw2A*$_{Hi375}$ (Flank A primers 1467 and 1468; Flank B primers 1469 and 1470). Flanks A and B were amplified with *Sma*I sites included in the R primer of Flank A and the F primer of Flank B, digested with *Sma*I, and blunt-end cloned into pJET1.2 by tri-molecular ligation. The Erm^R cassette was excised from pBSLerm [89] using *Sma*I (1,188 bp fragment) and added to the pJET1.2 clone by blunt-ended ligation into the *Sma*I site joining Flanks A and B. Finally, the whole disruption cassette was amplified and transformed into MIV-competent cells of the appropriate strain using the standard protocol [39]. Mutants were selected on sBHI-agar containing Erm, and correct targeting was confirmed by PCR and western blot. This procedure generated strains P836-P839.

TREP design

The spectrum and distribution of recombinants in transformed pools that carry invasion alleles/loci depends upon the number of loci involved, their genetic interactions, the rates of recombination at those loci, and the experimental environment. To maximize the chance of enriching invasive recombinants from transformed pools: (a) We selected for a donor-specific marker (either Nov^R or Nal^R) after transformations to ensure that recombinant clones were not derived from non-competent cells in the original culture [8,10]. (b) We selected for a recipient-specific marker (Spc^R or Str^R) to limit cross-contamination. (c) Colonies were pooled, so that each independent recombinant in the pools was represented by many ($>10^6$) cells. (d) We maximized the complexity of the recombinant pools emerging from the first round of selection. (e) We progressively increased the frequency of invasive recombinants by serial selection,

using a pool of the total CFU output from the gentamicin protection assay as the infecting material for the next cycle of selection.

Natural transformations

Recipient strains were made naturally competent using the standard protocol [39], except scaled up to 10 ml (~10^{10} CFU). Briefly, exponentially dividing cells growing in rich medium (sBHI) were transferred to starvation medium (MIV) for 100 min. Purified genomic DNA from the donor was incubated with naturally competent cultures at a concentration of ~1 genome per cell, or ~2 μg / 10^9 CFU / ml, for 30 min on a roller drum at 37°C, followed by a 1:5 dilution into sBHI and further incubation for 80 min to allow for expression of resistance alleles. Cultures were diluted and plated on sBHI-agar ± antibiotics to measure transformation and co-transformation frequencies (as Nov^R or Nal^R resistant colonies / CFU). Percent competence was calculated as (Nov^R Nal^R / CFU) / (Nov^R/CFU * Nal^R/CFU), as previously described [8,10,90]. To generate high complexity pools of recombinants, we plated 0.75 ml of a 10^{-2} dilution to 20 large petri dishes (20 cm diameter): 10 containing Nov and 10 with Nal (plus Spc or Str, depending on the recipient). This yielded ~10^4 resistant colonies per plate. Colonies from each set of 10 plates were scraped into a single 10 ml sBHI pool, titrated by dilution and plating to sBHI+antibiotics, and immediately stored as 1.25 ml aliquots in 15% glycerol at -80°C. This generated a total of four pools with an initial complexity of ~10^5 independent recombinants each, two for each recipient (Rd KW20 Spc^R and Hi375 Str^R), selected for either the Nov^R or the Nal^R donor allele, as well as a second antibiotic to select for the appropriate recipient background.

Infection of cultured epithelial cells and measurements of adhesion and intracellular invasion frequencies

The carcinomic human alveolar basal epithelial cell line A549 (ATCC CCL-185) was maintained in RPMI 1640 medium supplemented with 10 mM Hepes, 10% heat-inactivated fetal calf serum (FCS) and antibiotics (penicillin 100 units/ml and Str 0.1 mg/ml) in 25 cm^2 tissue culture flasks at 37°C in a humidified 5% CO_2 atmosphere. Chang cells (ATCC CCL-13) were cultured under the same atmospheric conditions in Minimum Essential Medium Eagle supplemented with 10% FCS and 1x MEM non-essential amino acid mixture (Sigma). Cells were seeded to 6×10^4 or to 1.2×10^5 cells / well in 24- or in 6-well tissue culture plates, respectively, for 32 h, and then serum starved for 16 h before infection. A ~90% confluence was reached by the time of infection. Adhesion and intracellular invasion assays in 24-well plates were conducted as previously described [15,24], starting with *H. influenzae* cells scraped from chocolate-agar plates (freshly grown for 16 h at 37°C with 5% CO_2) into PBS and adjusted to OD_{600} = 1. A small aliquot of this adjusted suspension was diluted and plated on sBHI-agar to titrate the input CFU.

For invasion assays, A549 cells were incubated with 0.2 ml of each adjusted bacterial suspension for 2 h, washed 3 times with PBS and incubated for 1 h with RPMI 1640 containing 10% FCS, Hepes 10 mM and gentamicin 200 μg/ml to kill extracellular bacteria (the bacterial isolates used all had minimum inhibitory concentrations of < 5 μg/mL), washed 3 times with PBS, and human cells were lysed with 300 μl of PBS-saponin 0.025% for 10 min at room temperature. To quantify intracellular invasion frequencies, lysates were serial diluted and plated onto sBHI-agar with appropriate antibiotics. Recovered CFU was divided by the input CFU to calculate "Invaders/CFU". Unless otherwise indicated, all infections were carried out in triplicate on three separate occasions. Adhesion assays were carried out similarly, excluding gentamicin treatment to calculate "Adherents/CFU". For adhesion assays, cells were incubated with 0.1 ml of each adjusted bacterial suspension for 30 min. Wells were then washed 5 times with

PBS and lysed as above. An alternative method was also used for both invasion and adhesion for comparisons of Rd, Strain 12, and Rd/HMW1$_{Strain12}$ following a previously published protocol [18]. The primary differences were the lack of a serum starvation step, a low-speed centrifugation step to quickly bring bacteria into contact with the monolayer, and a lower MOI.

Serial enrichment for invasive recombinants

To enrich for recombinants carrying donor-specific invasion alleles, we performed eight serial selections for invasive clones for each recombinant pool. To maximize the complexity of the initial recombinant pools (Pool 0), one frozen aliquot per pool (~10^{10} CFU of ~10^5 independent recombinants per aliquot) was used to infect three wells of A549 cells seeded onto 6-well plates. Pool 0 aliquots were first recovered by thawing, pelleting, resuspending in 5 ml sBHI, and incubating on a roller drum for 60 min at 37°C under 5% CO_2. Cultures were pelleted prior to proceeding with the invasion protocol, performed as described above, scaled up to cells seeded on 6-well plates (in 4 ml EBSS, with 0.8 ml of bacterial adjusted suspension / well), starting with resuspension of pellets in PBS, and ending with the total lysate plated on sBHI-agar (-+appropriate antibiotics) at varying dilutions. This allowed measurement of intracellular invasion frequency and provided material for the next cycle.

For each subsequent serial invasion cycle, all CFU were scraped off plates into PBS and thoroughly mixed before normalizing to OD_{600} = 1 and proceeding with the infection. For all cycles, unused material was stored as (i) 15% glycerol stocks at -80°C for repeat assays and isolation of individual clones, and (ii) as a pellet at -20°C for DNA extractions (except for the RdS Pool 1 material, for which none was left over). In practical terms, this serial infection procedure was repeated for four enrichment cycles, at which point recovered pools were frozen as 15% glycerol stocks to allow for a new set of confluent A549 cells to grow up; frozen stocks were then restarted to carry out four additional cycles of selection. Untransformed recipient controls were run in parallel to exclude potential issues related to cell seeding. At Pool 4, several single gentamicin-protected clones per enrichment were isolated on sBHI-agar plates and stored at -80°C in 15% glycerol for adhesion and invasion assays, as well as clone sequencing.

Western blot

To monitor YrbI-HA expression, whole cell extracts from strain Rd alone, Rd carrying pSU20, and Rd carrying pSU20-*Pr*::*kpsF-yrbi*-HA were prepared from bacterial cultures grown to OD_{600} = 0.9 in sBHI containing Cm, when required. YrbI-HA expression was analyzed by western blot with a primary rabbit anti-HA antibody (Sigma) diluted 1:4000, and a secondary goat anti-rabbit IgG (whole molecule, Sigma) antibody conjugated to horseradish peroxidase, diluted 1:1000.

To investigate HMW adhesin protein expression in strains NpNN, RdS, rRdS, HiT, HiTΔ*hmw1A*$_{Hi375}$, HiTΔ*hmw2A*$_{Hi375}$, rHiT, rHiTΔ*hmw1A*$_{86\text{-}028NP}$, rHiTΔ*hmw1A*$_{Hi375}$, rHiTΔ*hmw1A*$_{86\text{-}028NP}$Δ*hmw1A*$_{Hi375}$, whole cell extracts were prepared from bacterial suspensions recovered from overnight grown chocolate-agar plates and adjusted to OD_{600} = 1 in PBS. HMW1A expression was analyzed by western blot with a primary guinea pig anti-HMW1A (gp85) antibody diluted 1:2000 [91], and a secondary goat anti-guinea pig IgG (Santa Cruz) antibody conjugated to horseradish peroxidase, diluted 1:5000.

Bacterial self-aggregation

H. influenzae cells were scraped from chocolate-agar plates freshly grown for 16 h at 37°C with 5% CO_2 into PBS solution, and adjusted to OD_{600} = 0.45 in a 35 ml volume, and left standing at room temperature for at least 260 min. OD_{600} readings were performed at regular time

intervals on 500 μl aliquots gently collected from the top of each bacterial suspension. Four independent experiments were performed for each strain.

Immunofluorescence microscopy

A549 cells were seeded on 13 mm circular coverslips in 24-well tissue culture plates. Cells were infected at an MOI ~1:8 (5 μl) of each adjusted bacterial suspension for 2 h, and infected cells were incubated in RPMI 1640 containing 10% FCS, Hepes 10mM and gentamicin 200 μg/ml for 1 h. Cells were washed three times with PBS and fixed with 3.7% paraformaldehyde (PFA) in PBS pH 7.4 for 15 min at room temperature. Immunofluorescence staining was carried out as previously described [24]. *H. influenzae* cells were stained with a rabbit anti-NTHi serum (raised against a pool of strains Hi375, 2019, and 398 [24]) diluted 1:600. Late endosomes were stained with mouse monoclonal anti-human Lamp-1 H4A3 antibody (Developmental Studies Hybridoma Bank) diluted 1:100. DNA was stained with Hoechst 33342 (Invitrogen) diluted 1:2500. Donkey anti-rabbit conjugated to Cy2 and donkey anti-goat or donkey anti-mouse conjugated to Rhodamine secondary antibodies (Jackson) were diluted 1:100.

Samples were analyzed with a Carl Zeiss Axioskop 2 plus fluorescence microscope and a Carl Zeiss Axio Cam MRm monochrome camera. We quantified: (a) the percentage of infected cells, counting at least 250 cells per sample; (b) the number of bacteria per infected cell in at least 250 cells per sample type, scoring <10 bacteria/cell or >10 bacteria/cell; (c) co-localization of bacteria and Lamp-1—an NTHi-containing vacuole (NTHi-CV) was considered positive for Lamp-1 when the marker was detected throughout the area occupied by the bacterium, or around/enclosing the bacterium. To determine the percentage of bacteria that co-localized with Lamp-1, all bacteria located inside a minimum of 150 infected cells were scored in each experiment. Results were calculated from two independent experiments.

DNA sequencing

Genomic DNA was extracted from the donor and recipients, stored pools, and isolated clones by phenol/chloroform extraction as described [8]. Purity and quality were evaluated by Nanodrop spectrophotometry (Thermo Scientific) and agarose gel electrophoresis, and quantification was performed with Qbit fluorometry prior to sequencing library construction. Multiplexed sequencing libraries were produced using the Nextera XT kit following manufacturer recommendations (Illumina). Paired-end sequencing (2x101nt) was conducted on an Illumina HiSeq in RapidRun mode over several independent runs/lanes. Raw base call data (bcl) was converted into FastQ format (Illumina version 1.8) using the bcl2fastq conversion software from Illumina (version 1.8.3, setting—no-eamss). For recombinant clones, paired-end sequencing (2x151nt) was conducted on an Illumina MiSeq, which automates demultiplexing to provide raw FastQ files. Properties of the genomic DNA samples and sequencing statistics (including donor and recipient controls) are in S2 and S3 Tables.

Read alignments and variant calling

The genome sequence references for the donor and recipients were: 86-028NP (NC_007146.2) [28], Rd_KW20 (NC_000907.1) [27], and Hi375 (CP009610.1) [29]. For the Rd genome, all non-ACGT bases were first converted to Ns (some non-N ambiguous IUPAC nucleotide characters lead to errors running samtools mpileup). Reads from control strains were used to identify variation between the derivative strains' genomes and their deposited parental reference sequences (as described below). For all raw Illumina sequence processing, paired-end reads were trimmed of adapter sequences with Trimmomatic (v0.32) [92] and overlapping pairs were merged with COPE (v1.1.3; simple-connect mode) [93]. Next, reads were mapped using

bwa mem (v0.7.8) with default settings [94], duplicates were marked with SamBlaster (v0.1.14) [95], and aligned reads were sorted and compressed using SamBamba (v0.4.6) [96]. Subsequent steps filtered out reads with a mapping quality = 0, which excludes multiply mapping reads that align equally well to different reference genome coordinates.

For donor and recipient controls, as well as recombinant clones, single-nucleotide polymorphism (SNP) and small indel variant calling used samtools mpileup and bcftools view (v0.1.19) [97]. Variant frequency calling from recombinant pools used a python script (available at https://github.com/photonchang/allelecount/) to count reads supporting each of the 4 bases at each reference position directly from samtools mpileup output (for base calls with quality score >10), and subsequent parsing used linux commands (mostly awk). BedTools (v2.19.1) [98] was used for subsetting (using the intersect tool) with the variants detected between donor and recipient genomes. Variant tables were first corrected for "self" variants identified between reads and their own reference (with the exception of resistance-associated markers). This allowed calculation of recipient-specific, donor-specific, and erroneous base frequencies (*i.e.* bases with neither donor nor recipient identity). Manual validation of recombination breakpoints and clone assignments used the Integrative Genomics Viewer (v2.3.1) [99]. Identification of novel alleles that had approached fixation compared the variants called from Pool 8 reads to those from Pool 0 reads (using samtools mpileup and bcftools view). Due to systematic alignment artifacts that arise when mapping donor reads to recipient genomes in regions of high divergence, putative novel variation that was also identified only in reciprocal alignments of control reads ("unreliable" SNP positions) was excluded, leaving no observed fixed new mutations.

PCR validation. To distinguish between the four possible *hmw* genotypes, allele- and locus-specific PCR was used, with allele-specific primer pairs listed under "Allele ID" in S10 Table. Each pair is specific for one of the four possibilities and generates a distinct PCR product size as determined by standard agarose gel electrophoresis. Primers 1456+1458 were used for $hmw1A_{Hi375}$ (product size 1,364 bp), primers 1456+1457 for $hmw2A_{Hi375}$ (product size 1,076 bp); primers 1459+1460 for $hmw1A_{86\text{-}028NP}$ (product size 744 bp); and primers 1461+1462 for $hmw2A_{86\text{-}028NP}$ (product size 582 bp).

Statistics and plotting

Significant differences in invasion, adhesion, and self-aggregation phenotypes among strains and pools were evaluated using one-way ANOVA with *post hoc* hypothesis testing using Tukey's HSD ("honest significant differences"). Invasion and adhesion frequencies were first log-transformed prior to testing to account for the highly unequal variances observed between strains/pools that were quantified at distinct plating dilutions. Pairwise student's t-tests with untransformed data and Bonferroni correction gave qualitatively similar results. Plotting used the R statistical programming language including add-on packages seqinr, genoplotr, ggplot2, and Rcolorbrewer.

Data deposition

All sequence data were deposited at NCBI under BioProject PRJNA308311. BioSample accessions are included in S2 and S3 Tables. Parental strains were submitted to the SRA as BAM files aligned to their own reference sequence. Recombinant pool and clone data were submitted to the SRA as BAM files aligned to the appropriate recipient reference sequence (Hi375 or Rd KW20).

Supporting Information

S1 Text. Supplementary Results and Supplementary References.
(DOCX)

S1 Fig. Comparison of the donor to the two recipients at different scales. Turquoise lines above the x-axis indicate the position of SNPs distinguishing donor from recipient, while grey lines below the x-axis indicate positions in the recipient genome missing from the donor genome (at indels). **(A)** Hi375 recipient. **(B)** Rd KW20 recipient. Note that SNPs between Hi375 and 86-028NP are punctate, with stretches of very low SNP density punctuated by stretches of high SNP density. Genomic positions exclusive to the recipient strains are shown in grey; these coincide with areas that appear as regions of low SNP density, but these artifacts are insufficient to explain the pattern seen in Hi375. Conversely Rd-specific positions do largely explain low SNP density regions in Rd KW20.
(TIF)

S2 Fig. Invasion and adhesion phenotypes of parental and related strains. **(A)** Invasion of and **(B)** adhesion to A549 cells is shown for *H*. *influenzae* strains Rd KW20, Hi375, 86-028NP, and antibiotic resistant derivatives, including the parental strains.
(TIF)

S3 Fig. Competition for invasion between Rd and 86-028NP strain backgrounds. Two serial cycles of selection for intracellular invaders were conducted using three mixtures of 86-028NP Nov^R and Rd Str^R cells, at 1:100, 1:1,000, or 1:10,000 ratios. Prior to the first infection (input.A), the bacterial cell suspension was titrated for the total Nov^R and Str^R CFU used per well, and this closely matched the expected frequencies. After the first round of selection (output.A), dramatically fewer CFU were recovered, but Nov^R were proportionally much more abundant. Total unselected CFUs were pooled and titrated (input.B), showing that the proportion of Nov^R remained relatively the same in between cycles of selection for invasion. Finally, the second cycle of selection resulted in a higher yield with an even higher proportion of Nov^R colonies, representing a strong enrichment of 86-028NP over Rd, even when at a low relative abundance in the starting mixture.
(TIF)

S4 Fig. Donor allele frequencies in the transformed input pools. **(A)** and **(B)** NpNN-specific SNP frequencies as a function of chromosome coordinate for the RdS and HiT recipients, respectively, at Pool 0, prior to enrichment for invasive recombinants. Left panels: Nov^R-selected pools. Right panels: Nal^R-selected pools. Top panels: chromosome-wide view. Bottom panels: zoom on 60 kb windows around the antibiotic resistance markers. The peak SNP is the one conferring antibiotic resistance. **(C)** "Bean plots" summarizing 16 histograms of non-recipient allele frequencies for untransformed controls and the initial transformed recombinant pools. The left side (salmon-colored) of each bean shows a histogram for allele frequencies with donor allele identities, whereas the right side (light blue) shows a histogram for "novel" alleles (neither recipient nor donor). The latter are sequencing errors, while the former are sequencing errors for the control strains and a combination of sequencing errors and transformants for the transformed pools.
(TIF)

S5 Fig. Serial selection of invasive recombinants by gentamicin protection. Invaders/CFU for pools during the initial eight serial selections for invasive recombinants. Recovered CFU that survived gentamicin treatment (Pool 1) served as input for the next cycle (which generated Pool 2). The values show the combined ability of clones in Pool *n* to invade airway epithelial cells, while the recovered colonies comprise Pool *n*+1. This procedure was carried out eight times. The apparent decline in invasiveness seen at Pool 4 appears to be an artifact, since no such decline was seen in the replicate assays (Fig 3A). Instead, this drop likely reflects that Pool 4 bacteria had been frozen and re-inoculated prior to the next cycle, combined with batch-to-batch variation of the confluent A549 cells used.
(TIF)

S6 Fig. No improvement by selection on untransformed recipients. Control experiment using untransformed recipients cultures in triplicate found no increase in invasiveness over 5 serial selections. This experiment was conducted independently for each of the recipients and separately from the experimental enrichments to minimize enrichment of cross-contaminants.
(TIF)

S7 Fig. Complexity of recombination tracts decreases at the antibiotic resistance markers over serial passages. Genomic profiling at antibiotic-selected sites for both the **(A)** RdS and **(B)** HiT recipients at Nov^R (top) and Nal^R (bottom) sites (*gyrB* and *gyrA* respectively, see S6 Table) for Pools 0, 2, 4, and 8. Axes are as in other figures with x-axes indicating recipient genome coordinate (in kb) and the y-axis indicating donor allele frequency. RdS Nov^R contains a single clone at ~95% by Pool 8, while RdS Nal^R contains two dominant clones, one at ~70% and the other ~30%. HiT Nov^R contains two dominant clones (at ~30% and 70%), whereas HiT Nal^R appears to contain two clones at ~80% and ~20%. For this pool, only a single genotype (the one at ~80%) was recovered in the four individual clones collected from Pool 4. No other donor segments appeared at ~20%, so this is likely due to incomplete fixation of the invasive genotype after several rounds of selection.
(TIF)

S8 Fig. Read alignment artifact at the *radA*-proximal *hmw1* locus in the HiT pools. (A) Pools 0, 2, 4, and 8 for HiT Nov^R and HiT Nal^R as in other figures (x-axis is HiT recipient coordinate in kb, and y-axis is donor allele frequency). **(B)** Genomic map around the same interval. The thick black horizontal line shows the entire range of positions containing donor frequencies >5%. The affected interval spans only the hmw1 locus; no flanking variation was detected, unlike the situation at the *yrbI*-adjacent $hmw2_{Hi375}$, which was replaced by the $hmw1_{86\text{-}028NP}$ allele. Donor allele frequencies are highly variable in this region. They are also highly consistent between the two pools, which was unexpected, as all other overlapping donor segments detected had distinct recombination breakpoints. Allele-specific PCR assays confirm this as read alignment artifact and confirm that the *radA*-proximal adhesin remain $hmw1_{Hi375}$ across strains (S9 Fig).
(TIF)

S9 Fig. Agarose gel showing allele/locus-specific PCR products amplified for the four possible *hmw* alleles. Strains are listed as it follows: (1) NpNN, (2) RdS, (3) rRdS, (4) HiT, (5) HiTΔ$hmw1A_{Hi375}$, (6) HiTΔ$hmw2A_{Hi375}$, (7) rHiT, (8) rHiTΔ$hmw1A_{86\text{-}028NP}$, (9) rHiTΔ$hmw1A_{Hi375}$, (10) rHiTΔ$hmw1A_{86\text{-}028NP}$Δ$hmw1A_{Hi375}$, and primers are in S10 Table. **(A)** Primers 1456+1458 identify $hmw1A_{Hi375}$ (1,364 bp product); **(B)** primers 1456+1457 identify $hmw2A_{Hi375}$ (1,076 bp product); **(C)** primers 1459+1460 identify $hmw1A_{86\text{-}028NP}$ (744 bp product); and **(D)** primers 1461+1462 for $hmw2A_{86\text{-}028NP}$ (582 bp product). Lanes 8 and 10 rendered a correct size band upon PCR with primers 1459+1460 because mutant strains lacking $hmw1A_{86\text{-}028NP}$ were generated by partial deletion that maintains the annealing sites for the primers and product size.
(TIF)

S10 Fig. Confirmation of HMW adhesion expression and testing for a role by *kpsF* and *yrbI*. (A) Western blot showing expression of $HMW1A_{86\text{-}028NP}$ (154 KDa) adhesin. Whole cell extracts of NpNN, RdS and rRdS (P540, genotype B) were prepared and used to detect HMW by immunoblot with the guinea pig anti-HMW1A gp85 antibody. **(B and C)** Addition of the *kpsF* and *yrbI* alleles from 86-028NP on a plasmid does not increase intracellular invasion frequencies. **(B)** Western blot showing expression from plasmid carrying an interval carrying *kpsF-yrbI* from 86-028NP. Whole cell extracts of cultures (Rd, Rd pSU20, and Rd pSU20-*kpsF-yrbI*-HA) were prepared and used to detect Hap-HA by immunoblot with a rabbit anti-HA antibody, finding expression of the expected ~19.3-kDa protein in the expected strain. **(C)** The same strains were

used to infect A549 cells and measure bacterial intracellular invasion. Experiments were performed three times in triplicate (different symbols denote independent experiments).
(TIF)

S11 Fig. A role for HMW1 is seen for a distinct strain, for another epithelial cell type, and with an alternative protocol. Invasion **(A)** and adhesion **(B)** by Rd, Rd $hmw1_{strain12}$, and Strain 12 bacterial strains into Chang and A549 epithelial cell lines. An alternative protocol that includes centrifugation to quickly bring bacteria into contact with the cell monolayer was used for these experiments, showing that both cell type and details of the infection procedure give qualitatively similar results.
(TIF)

S1 Table. Strains and Plasmids Used
(DOCX)

S2 Table. Pool and control sequencing statistics
(DOCX)

S3 Table. Recombinant clone sequencing statistics
(DOCX)

S4 Table. Alignment statistics for the untransformed controls
(DOCX)

S5 Table. Transformation frequencies and estimated competence
(DOCX)

S6 Table. Allele frequencies around antibiotic resistances in Pool 0
(DOCX)

S7 Table. Non-reference alleles at reliable SNP positions in transformed pools and untransformed controls
(DOCX)

S8 Table. Clone genotype assignments
(DOCX)

S9 Table. Donor segments detected in each isolated genotype
(DOCX)

S10 Table. Primers used in this study
(DOCX)

Acknowledgments

We thank Nathaniel Lin for work on pileup parsing script and Begoña Euba for help with DNA preparations and cloning.

Author Contributions

Conceived and designed the experiments: JCM JG RJR. Performed the experiments: JCM CV JG SS JM AFC EAP. Analyzed the data: JCM CV JG JM AFC EAP SS RJR JWSG CN. Contributed reagents/materials/analysis tools: JG JCM CN RJR JWSG. Wrote the paper: JCM JG RJR CN.

References

1. Wong SM, Bernui M, Shen H, Akerley BJ (2013) Genome-wide fitness profiling reveals adaptations required by *Haemophilus* in coinfection with influenza A virus in the murine lung. Proc Natl Acad Sci U S A 110: 15413–15418. doi: 10.1073/pnas.1311217110 PMID: 24003154
2. Akerley BJ, Rubin EJ, Novick VL, Amaya K, Judson N, et al. (2002) A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. Proc Natl Acad Sci U S A 99: 966–971. PMID: 11805338
3. Johnston C, Martin B, Fichant G, Polard P, Claverys JP (2014) Bacterial transformation: distribution, shared mechanisms and divergent control. Nat Rev Microbiol 12: 181–196. doi: 10.1038/nrmicro3199 PMID: 24509783
4. Mell JC, Redfield RJ (2014) Natural competence and the evolution of DNA uptake specificity. J Bacteriol 196: 1471–1483. doi: 10.1128/JB.01293-13 PMID: 24488316
5. Avery OT, Macleod CM, McCarty M (1944) Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. J Exp Med 79: 137–158. PMID: 19871359
6. Dalia AB, McDonough E, Camilli A (2014) Multiplex genome editing by natural transformation. Proc Natl Acad Sci U S A 111: 8937–8942. doi: 10.1073/pnas.1406478111 PMID: 24889608
7. Gray TA, Krywy JA, Harold J, Palumbo MJ, Derbyshire KM (2013) Distributive conjugal transfer in mycobacteria generates progeny with meiotic-like genome-wide mosaicism, allowing mapping of a mating identity locus. PLoS Biol 11: e1001602. doi: 10.1371/journal.pbio.1001602 PMID: 23874149
8. Mell JC, Lee JY, Firme M, Sinha S, Redfield RJ (2014) Extensive cotransformation of natural variation into chromosomes of naturally competent *Haemophilus influenzae*. G3 (Bethesda) 4: 717–731.
9. Freddolino PL, Goodarzi H, Tavazoie S (2014) Revealing the genetic basis of natural bacterial phenotypic divergence. J Bacteriol 196: 825–839. doi: 10.1128/JB.01039-13 PMID: 24317396
10. Mell JC, Shumilina S, Hall IM, Redfield RJ (2011) Transformation of natural genetic variation into *Haemophilus influenzae* genomes. PLoS Pathog 7: e1002151. doi: 10.1371/journal.ppat.1002151 PMID: 21829353
11. Agrawal A, Murphy TF (2011) *Haemophilus influenzae* infections in the *H*. *influenzae* type b conjugate vaccine era. J Clin Microbiol 49: 3728–3732. doi: 10.1128/JCM.05476-11 PMID: 21900515
12. Garmendia J, Viadas C, Calatayud L, Mell JC, Marti-Lliteras P, et al. (2014) Characterization of nontypable *Haemophilus influenzae* isolates recovered from adult patients with underlying chronic lung disease reveals genotypic and phenotypic traits associated with persistent infection. PLoS One 9: e97020. doi: 10.1371/journal.pone.0097020 PMID: 24824990
13. Murphy TF, Brauer AL, Schiffmacher AT, Sethi S (2004) Persistent colonization by *Haemophilus influenzae* in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 170: 266–272. PMID: 15117742
14. Clementi CF, Murphy TF (2011) Non-typeable *Haemophilus influenzae* invasion and persistence in the human respiratory tract. Front Cell Infect Microbiol 1: 1. doi: 10.3389/fcimb.2011.00001 PMID: 22919570
15. Lopez-Gomez A, Cano V, Moranta D, Morey P, Garcia del Portillo F, et al. (2012) Host cell kinases, alpha5 and beta1 integrins, and Rac1 signalling on the microtubule cytoskeleton are important for nontypable *Haemophilus influenzae* invasion of respiratory epithelial cells. Microbiology 158: 2384–2398. doi: 10.1099/mic.0.059972-0 PMID: 22723286
16. Ahren IL, Williams DL, Rice PJ, Forsgren A, Riesbeck K (2001) The importance of a beta-glucan receptor in the nonopsonic entry of nontypeable *Haemophilus influenzae* into human monocytic and epithelial cells. J Infect Dis 184: 150–158. PMID: 11424011
17. Ketterer MR, Shao JQ, Hornick DB, Buscher B, Bandi VK, et al. (1999) Infection of primary human bronchial epithelial cells by *Haemophilus influenzae*: macropinocytosis as a mechanism of airway epithelial cell entry. Infect Immun 67: 4161–4170. PMID: 10417188
18. St Geme JW 3rd, Falkow S (1990) *Haemophilus influenzae* adheres to and enters cultured human epithelial cells. Infect Immun 58: 4036–4044. PMID: 2254028
19. Swords WE, Buscher BA, Ver Steeg Ii K, Preston A, Nichols WA, et al. (2000) Non-typeable *Haemophilus influenzae* adhere to and invade human bronchial epithelial cells via an interaction of lipooligosaccharide with the PAF receptor. Mol Microbiol 37: 13–27. PMID: 10931302
20. Virji M, Kayhty H, Ferguson DJ, Alexandrescu C, Moxon ER (1991) Interactions of *Haemophilus influenzae* with cultured human endothelial cells. Microb Pathog 10: 231–245. PMID: 1895925

21. Bandi V, Apicella MA, Mason E, Murphy TF, Siddiqi A, et al. (2001) Nontypeable *Haemophilus influenzae* in the lower respiratory tract of patients with chronic bronchitis. Am J Respir Crit Care Med 164: 2114–2119. PMID: 11739144

22. Forsgren J, Samuelson A, Ahlin A, Jonasson J, Rynnel-Dagoo B, et al. (1994) *Haemophilus influenzae* resides and multiplies intracellularly in human adenoid tissue as demonstrated by in situ hybridization and bacterial viability assay. Infect Immun 62: 673–679. PMID: 7507900

23. St Geme JW 3rd (2002) Molecular and cellular determinants of non-typeable *Haemophilus influenzae* adherence and invasion. Cell Microbiol 4: 191–200. PMID: 11952636

24. Morey P, Cano V, Marti-Lliteras P, Lopez-Gomez A, Regueiro V, et al. (2011) Evidence for a non-replicative intracellular stage of nontypable *Haemophilus influenzae* in epithelial cells. Microbiology 157: 234–250. doi: 10.1099/mic.0.040451-0 PMID: 20929955

25. Woo JI, Oh S, Webster P, Lee YJ, Lim DJ, et al. (2014) NOD2/RICK-dependent beta-defensin 2 regulation is protective for nontypeable *Haemophilus influenzae*-induced middle ear infection. PLoS One 9: e90933. doi: 10.1371/journal.pone.0090933 PMID: 24625812

26. Clementi CF, Hakansson AP, Murphy TF (2014) Internalization and trafficking of nontypeable *Haemophilus influenzae* in human respiratory epithelial cells and roles of IgA1 proteases for optimal invasion and persistence. Infect Immun 82: 433–444. doi: 10.1128/IAI.00864-13 PMID: 24218477

27. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269: 496–512. PMID: 7542800

28. Harrison A, Dyer DW, Gillaspy A, Ray WC, Mungur R, et al. (2005) Genomic sequence of an otitis media isolate of nontypeable *Haemophilus influenzae*: comparative study with *H. influenzae* serotype d, strain KW20. J Bacteriol 187: 4627–4636. PMID: 15968074

29. Mell JC, Sinha S, Balashov S, Viadas C, Grassa CJ, et al. (2014) Complete Genome Sequence of *Haemophilus influenzae* Strain 375 from the Middle Ear of a Pediatric Patient with Otitis Media. Genome Announc 2.

30. Maughan H, Redfield RJ (2009) Tracing the evolution of competence in *Haemophilus influenzae*. PLoS One 4: e5854. doi: 10.1371/journal.pone.0005854 PMID: 19516897

31. Maughan H, Redfield RJ (2009) Extensive variation in natural competence in *Haemophilus influenzae*. Evolution 63: 1852–1866. doi: 10.1111/j.1558-5646.2009.00658.x PMID: 19239488

32. Daines DA, Cohn LA, Coleman HN, Kim KS, Smith AL (2003) *Haemophilus influenzae* Rd KW20 has virulence properties. J Med Microbiol 52: 277–282. PMID: 12676864

33. Hong W, Mason K, Jurcisek J, Novotny L, Bakaletz LO, et al. (2007) Phosphorylcholine decreases early inflammation and promotes the establishment of stable biofilm communities of nontypeable *Haemophilus influenzae* strain 86-028NP in a chinchilla model of otitis media. Infect Immun 75: 958–965. PMID: 17130253

34. Mason KM, Munson RS Jr., Bakaletz LO (2003) Nontypeable *Haemophilus influenzae* gene expression induced in vivo in a chinchilla model of otitis media. Infect Immun 71: 3454–3462. PMID: 12761130

35. Mason KM, Munson RS Jr., Bakaletz LO (2005) A mutation in the sap operon attenuates survival of nontypeable *Haemophilus influenzae* in a chinchilla model of otitis media. Infect Immun 73: 599–608. PMID: 15618200

36. Novotny LA, Mason KM, Bakaletz LO (2005) Development of a chinchilla model to allow direct, continuous, biophotonic imaging of bioluminescent nontypeable *Haemophilus influenzae* during experimental otitis media. Infect Immun 73: 609–611. PMID: 15618201

37. St Geme JW 3rd, Falkow S, Barenkamp SJ (1993) High-molecular-weight proteins of nontypable *Haemophilus influenzae* mediate attachment to human epithelial cells. Proc Natl Acad Sci U S A 90: 2875–2879. PMID: 8464902

38. St Geme JW 3rd, Yeo HJ (2009) A prototype two-partner secretion pathway: the *Haemophilus influenzae* HMW1 and HMW2 adhesin systems. Trends Microbiol 17: 355–360. doi: 10.1016/j.tim.2009.06.002 PMID: 19660953

39. Poje G, Redfield RJ (2003) Transformation of *Haemophilus influenzae*. Methods Mol Med 71: 57–70. PMID: 12374031

40. De Chiara M, Hood D, Muzzi A, Pickard DJ, Perkins T, et al. (2014) Genome sequencing of disease and carriage isolates of nontypeable *Haemophilus influenzae* identifies discrete population structure. Proc Natl Acad Sci U S A 111: 5439–5444. doi: 10.1073/pnas.1403353111 PMID: 24706866

41. Davis GS, Patel M, Hammond J, Zhang L, Dawid S, et al. (2014) Prevalence, distribution, and sequence diversity of hmwA among commensal and otitis media non-typeable *Haemophilus influenzae*. Infect Genet Evol 28: 223–232. doi: 10.1016/j.meegid.2014.09.035 PMID: 25290952

42. Ecevit IZ, McCrea KW, Pettigrew MM, Sen A, Marrs CF, et al. (2004) Prevalence of the *hifBC*, *hmw1A*, *hmw2A*, *hmwC*, and *hia* Genes in *Haemophilus influenzae* Isolates. J Clin Microbiol 42: 3065–3072. PMID: 15243061

43. Vuong J, Wang X, Theodore JM, Whitmon J, Gomez de Leon P, et al. (2013) Absence of high molecular weight proteins 1 and/or 2 is associated with decreased adherence among non-typeable *Haemophilus influenzae* clinical isolates. J Med Microbiol 62: 1649–1656. doi: 10.1099/jmm.0.058222-0 PMID: 23988628

44. Grass S, Buscher AZ, Swords WE, Apicella MA, Barenkamp SJ, et al. (2003) The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. Mol Microbiol 48: 737–751. PMID: 12694618

45. Fonseca MV, Swanson MS (2014) Nutrient salvaging and metabolism by the intracellular pathogen *Legionella pneumophila*. Front Cell Infect Microbiol 4: 12. doi: 10.3389/fcimb.2014.00012 PMID: 24575391

46. Liss V, Hensel M (2015) Take the tube: remodelling of the endosomal system by intracellular *Salmonella enterica*. Cell Microbiol 17: 639–647. doi: 10.1111/cmi.12441 PMID: 25802001

47. Winchell CG, Steele S, Kawula T, Voth DE (2015) Dining in: intracellular bacterial pathogen interplay with autophagy. Curr Opin Microbiol 29: 9–14. doi: 10.1016/j.mib.2015.09.004 PMID: 26462048

48. Raffel FK, Szelestey BR, Beatty WL, Mason KM (2013) The *Haemophilus influenzae* Sap transporter mediates bacterium-epithelial cell homeostasis. Infect Immun 81: 43–54. doi: 10.1128/IAI.00942-12 PMID: 23071138

49. Garmendia J, Marti-Lliteras P, Moleres J, Puig C, Bengoechea JA (2012) Genotypic and phenotypic diversity of the noncapsulated *Haemophilus influenzae*: adaptation and pathogenesis in the human airways. Int Microbiol 15: 159–172. PMID: 23844475

50. Euba B, Moleres J, Segura V, Viadas C, Morey P, et al. (2015) Genome Expression Profiling-Based Identification and Administration Efficacy of Host-Directed Antimicrobial Drugs against Respiratory Infection by Nontypeable *Haemophilus influenzae*. Antimicrob Agents Chemother 59: 7581–7592. doi: 10.1128/AAC.01278-15 PMID: 26416856

51. Prasadarao NV, Lysenko E, Wass CA, Kim KS, Weiser JN (1999) Opacity-associated protein A contributes to the binding of Haemophilus influenzae to chang epithelial cells. Infect Immun 67: 4153–4160. PMID: 10417187

52. Rosadini CV, Ram S, Akerley BJ (2014) Outer membrane protein P5 is required for resistance of nontypeable *Haemophilus influenzae* to both the classical and alternative complement pathways. Infect Immun 82: 640–649. doi: 10.1128/IAI.01224-13 PMID: 24478079

53. Singh B, Brant M, Kilian M, Hallstrom B, Riesbeck K (2010) Protein E of *Haemophilus influenzae* is a ubiquitous highly conserved adhesin. J Infect Dis 201: 414–419. doi: 10.1086/649782 PMID: 20028233

54. Spahich NA, St Geme JW 3rd (2011) Structure and function of the *Haemophilus influenzae* autotransporters. Front Cell Infect Microbiol 1: 5. doi: 10.3389/fcimb.2011.00005 PMID: 22919571

55. Ronander E, Brant M, Eriksson E, Morgelin M, Hallgren O, et al. (2009) Nontypeable *Haemophilus influenzae* adhesin protein E: characterization and biological activity. J Infect Dis 199: 522–531. doi: 10.1086/596211 PMID: 19125675

56. St Geme JW 3rd, de la Morena ML, Falkow S (1994) A *Haemophilus influenzae* IgA protease-like protein promotes intimate interaction with human epithelial cells. Mol Microbiol 14: 217–233. PMID: 7830568

57. Euba B, Moleres J, Viadas C, Ruiz de los Mozos I, Valle J, et al. (2015) Relative Contribution of P5 and Hap Surface Proteins to Nontypable *Haemophilus influenzae* Interplay with the Host Upper and Lower Airways. PLoS One 10: e0123154. doi: 10.1371/journal.pone.0123154 PMID: 25894755

58. St Geme JW 3rd, Grass S (1998) Secretion of the *Haemophilus influenzae* HMW1 and HMW2 adhesins involves a periplasmic intermediate and requires the HMWB and HMWC proteins. Mol Microbiol 27: 617–630. PMID: 9489673

59. Buscher AZ, Burmeister K, Barenkamp SJ, St Geme JW 3rd (2004) Evolutionary and functional relationships among the nontypeable *Haemophilus influenzae* HMW family of adhesins. J Bacteriol 186: 4209–4217. PMID: 15205423

60. Dawid S, Barenkamp SJ, St Geme JW 3rd (1999) Variation in expression of the *Haemophilus influenzae* HMW adhesins: a prokaryotic system reminiscent of eukaryotes. Proc Natl Acad Sci U S A 96: 1077–1082. PMID: 9927696

61. Cholon DM, Cutter D, Richardson SK, Sethi S, Murphy TF, et al. (2008) Serial isolates of persistent *Haemophilus influenzae* in patients with chronic obstructive pulmonary disease express diminishing

quantities of the HMW1 and HMW2 adhesins. Infect Immun 76: 4463–4468. doi: 10.1128/IAI.00499-08 PMID: 18678658

62. Davis GS, Marino S, Marrs CF, Gilsdorf JR, Dawid S, et al. (2014) Phase variation and host immunity against high molecular weight (HMW) adhesins shape population dynamics of nontypeable *Haemophilus influenzae* within human hosts. J Theor Biol 355: 208–218. doi: 10.1016/j.jtbi.2014.04.010 PMID: 24747580

63. Giufre M, Carattoli A, Cardines R, Mastrantonio P, Cerquetti M (2008) Variation in expression of HMW1 and HMW2 adhesins in invasive nontypeable *Haemophilus influenzae* isolates. BMC Microbiol 8: 83. doi: 10.1186/1471-2180-8-83 PMID: 18510729

64. Barenkamp SJ, Leininger E (1992) Cloning, expression, and DNA sequence analysis of genes encoding nontypeable *Haemophilus influenzae* high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of *Bordetella pertussis*. Infect Immun 60: 1302–1313. PMID: 1548058

65. Dawid S, Grass S, St Geme JW 3rd (2001) Mapping of binding domains of nontypeable *Haemophilus influenzae* HMW1 and HMW2 adhesins. Infect Immun 69: 307–314. PMID: 11119519

66. Giufre M, Muscillo M, Spigaglia P, Cardines R, Mastrantonio P, et al. (2006) Conservation and diversity of HMW1 and HMW2 adhesin binding domains among invasive nontypeable *Haemophilus influenzae* isolates. Infect Immun 74: 1161–1170. PMID: 16428765

67. van Schilfgaarde M, van Ulsen P, Eijk P, Brand M, Stam M, et al. (2000) Characterization of adherence of nontypeable *Haemophilus influenzae* to human epithelial cells. Infect Immun 68: 4658–4665. PMID: 10899870

68. Li H, Grass S, Wang T, Liu T, St Geme JW 3rd (2007) Structure of the *Haemophilus influenzae* HMW1B translocator protein: evidence for a twin pore. J Bacteriol 189: 7497–7502. PMID: 17693509

69. Grass S, Lichti CF, Townsend RR, Gross J, St Geme JW 3rd (2010) The *Haemophilus influenzae* HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin. PLoS Pathog 6: e1000919. doi: 10.1371/journal.ppat.1000919 PMID: 20523900

70. McCann JR, St Geme JW 3rd (2014) The HMW1C-like glycosyltransferases—an enzyme family with a sweet tooth for simple sugars. PLoS Pathog 10: e1003977. doi: 10.1371/journal.ppat.1003977 PMID: 24722584

71. St Geme JW 3rd (1994) The HMW1 adhesin of nontypeable *Haemophilus influenzae* recognizes sialylated glycoprotein receptors on cultured human epithelial cells. Infect Immun 62: 3881–3889. PMID: 8063405

72. Liang Y, Hua Q, Pan P, Yang J, Zhang Q (2015) Development of a novel method to evaluate sialylation of glycoproteins and analysis of gp96 sialylation in Hela, SW1990 and A549 cell lines. Biol Res 48: 52. doi: 10.1186/s40659-015-0041-8 PMID: 26363641

73. Sanders JD, Cope LD, Hansen EJ (1994) Identification of a locus involved in the utilization of iron by *Haemophilus influenzae*. Infect Immun 62: 4515–4525. PMID: 7927717

74. Todorova K, Maurer P, Rieger M, Becker T, Bui NK, et al. (2015) Transfer of penicillin resistance from *Streptococcus oralis* to *Streptococcus pneumoniae* identifies *murE* as resistance determinant. Mol Microbiol 97: 866–880. doi: 10.1111/mmi.13070 PMID: 26010014

75. Barrick JE, Lenski RE (2013) Genome dynamics during experimental evolution. Nat Rev Genet 14: 827–839. doi: 10.1038/nrg3564 PMID: 24166031

76. Traverse CC, Mayo-Smith LM, Poltak SR, Cooper VS (2013) Tangled bank of experimentally evolved *Burkholderia* biofilms reflects selection during chronic infections. Proc Natl Acad Sci U S A 110: E250–259. doi: 10.1073/pnas.1207025110 PMID: 23271804

77. van Ditmarsch D, Boyle KE, Sakhtah H, Oyler JE, Nadell CD, et al. (2013) Convergent evolution of hyperswarming leads to impaired biofilm formation in pathogenic bacteria. Cell Rep 4: 697–708. doi: 10.1016/j.celrep.2013.07.026 PMID: 23954787

78. Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol Evol 28: 659–669. doi: 10.1016/j.tree.2013.08.003 PMID: 24075201

79. Moradigaravand D, Engelstadter J (2013) The evolution of natural competence: disentangling costs and benefits of sex in bacteria. Am Nat 182: E112–126. doi: 10.1086/671909 PMID: 24021408

80. Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, et al. (2010) Genome-wide analysis of a long-term evolution experiment with *Drosophila*. Nature 467: 587–590. doi: 10.1038/nature09352 PMID: 20844486

81. Zhou D, Udpa N, Gersten M, Visk DW, Bashir A, et al. (2011) Experimental selection of hypoxia-tolerant *Drosophila melanogaster*. Proc Natl Acad Sci U S A 108: 2349–2354. doi: 10.1073/pnas.1010643108 PMID: 21262834

82. Engelmoer DJ, Donaldson I, Rozen DE (2013) Conservative sex and the benefits of transformation in *Streptococcus pneumoniae*. PLoS Pathog 9: e1003758. doi: 10.1371/journal.ppat.1003758 PMID: 24244172

83. Cooper TF (2007) Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. PLoS Biol 5: e225. PMID: 17713986

84. Duprex WP, Fouchier RA, Imperiale MJ, Lipsitch M, Relman DA (2015) Gain-of-function experiments: time for a real debate. Nat Rev Microbiol 13: 58–64. doi: 10.1038/nrmicro3405 PMID: 25482289

85. Casadevall A, Howard D, Imperiale MJ (2014) An epistemological perspective on the value of gain-of-function experiments involving pathogens with pandemic potential. MBio 5: e01875–01814. doi: 10.1128/mBio.01875-14 PMID: 25227471

86. Hood DW, Makepeace K, Deadman ME, Rest RF, Thibault P, et al. (1999) Sialic acid in the lipopolysaccharide of *Haemophilus influenzae*: strain distribution, influence on serum resistance and structural characterization. Mol Microbiol 33: 679–692. PMID: 10447878

87. Sanchez R (1998) A medium-copy-number plasmid for insertional mutagenesis of *Streptococcus mutans*. Plasmid 40: 247–251. PMID: 9806863

88. Tracy E, Ye F, Baker BD, Munson RS Jr. (2008) Construction of non-polar mutants in *Haemophilus influenzae* using FLP recombinase technology. BMC Mol Biol 9: 101. doi: 10.1186/1471-2199-9-101 PMID: 19014437

89. Allen S, Zaleski A, Johnston JW, Gibson BW, Apicella MA (2005) Novel sialic acid transporter of *Haemophilus influenzae*. Infect Immun 73: 5291–5300. PMID: 16113244

90. Goodgal SH, Herriott RM (1961) Studies on transformations of *Hemophilus influenzae*. I. Competence. J Gen Physiol 44: 1201–1227. PMID: 13707010

91. Buscher AZ, Grass S, Heuser J, Roth R, St Geme JW 3rd (2006) Surface anchoring of a bacterial adhesin secreted by the two-partner secretion pathway. Mol Microbiol 61: 470–483. PMID: 16771846

92. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120. doi: 10.1093/bioinformatics/btu170 PMID: 24695404

93. Liu B, Yuan J, Yiu SM, Li Z, Xie Y, et al. (2012) COPE: an accurate k-mer-based pair-end reads connection tool to facilitate genome assembly. Bioinformatics 28: 2870–2874. doi: 10.1093/bioinformatics/bts563 PMID: 23044551

94. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997v2.

95. Faust GG, Hall IM (2014) SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics 30: 2503–2505. doi: 10.1093/bioinformatics/btu314 PMID: 24812344

96. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P (2015) Sambamba: fast processing of NGS alignment formats. Bioinformatics.

97. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987–2993. doi: 10.1093/bioinformatics/btr509 PMID: 21903627

98. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842. doi: 10.1093/bioinformatics/btq033 PMID: 20110278

99. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, et al. (2011) Integrative genomics viewer. Nat Biotechnol 29: 24–26. doi: 10.1038/nbt.1754 PMID: 21221095

SUPPLEMENTARY RESULTS AND REFERENCES

"Transformed recombinant enrichment profiling rapidly identifies HMW1 as an intracellular invasion locus in *Haemophilus influenzae*" by Mell *et al.*

SUPPLEMENTARY RESULTS

Genomic comparisons of donor to recipients. The three parental strains can be distinguished by tens of thousands of single-nucleotide polymorphisms distinguishable by genomic DNA sequencing, as well as by hundreds of "accessory" or "distributed" loci present in only some strains. Re-sequencing the parental strains was used as a control to identify unambiguous and reliable genetic markers, as well as to identify genomic positions with systematically low coverage or ambiguous base calls. Sequence reads collected from the recipients were aligned to their respective references, also to the donor, and *vice versa*.

To evaluate the quality of the DNA sequencing and read alignment and to correct the references for intra-strain variation, we examined variant frequencies at each position after mapping reads from the parents to their own genomic reference (S4 Table). Variants could be any of the 3 non-reference bases aligned to each reference genome position. Nearly all positions had 10 or more reads aligned to them (>99.9% of positions for all five controls), with mean depths per control averaging 372 reads per position. For all controls, >50% of positions showed no variation whatsoever from the reference base, with a mean frequency for non-reference variants (VarFreq) of ~10^{-3} per position or less (median VarFeq=0).

A small number of positions showed evidence for high frequency strand-specific sequencing errors (as determined by Fisher's exact test with a p-value cut-off of 0.05), and these had a disproportionately high VarFreq (S4 Table). When only positions with no evidence for strand bias were considered, <2.5% of had a VarFreq > 0.01. Together, these data indicate that sequencing and alignment artifacts are rare, with most positions having no evidence of non-reference bases, even at very high sequencing depths. These data also suggest that, as long as

donor allele frequencies exceed a few percent, sequencing-based measurements are unaffected by errors at nearly every genomic position.

Control sequence alignments also identified positions that differed between the reference sequences and the parental strain derivatives used for the experiments (last row, S4 Table). Variants in HiT are due to the addition of the Str^R marker, which also carried several Rd-specific SNPs, and similarly variants in NPNN are due to addition of the Nov^R and Nal^R markers, which added adjacent Rd-specific SNPs [2,3]. Differences between the Rd-derivatives and their reference are largely due to errors in the Rd KW20 genome sequence produced in 1995, as previously described [2], but also include MAP7-specific genetic variation, including seven antibiotic resistance markers (as previously described in [3]), in the case of the RdS strain, only the Spc^R marker. Identifying these variants between the strains and their references allowed for correction of reference base at these positions in the pool and clone-based genotyping used for TREP mapping.

To identify unambiguous and reliable genetic markers, reciprocal alignments between donor and recipient reads to donor and recipient genomes was used to identify mutually consistent variant calls. This set of reliable SNP positions required that alignment of donor and recipient reads to their own reference supported the reference base and a variant base when aligned to their reciprocal reference. This procedure also eliminated putative SNP positions that, when aligned to either reference, generated mixed (or “heterozygous”) genotype calls, including high-frequency strand-biased positions. Whole-genome alignnments generated by Mauve provided a coordinate liftover table for cross-referencing positions between donor and recipient genomes. Comparisons of 86-028NP and Rd have previously been made [1,2]; here we found 36,869 reliable SNPs. Comparison of 86-028NP and Hi375 found this pair of strains to be more similar, reliably distinguished by 19,520 SNPs. Notably, the genetic variation between 86-028NP and Hi375 was punctuated by segments of much lower sequence identity (S2 Fig.). This mosaic divergence seemed unusual, suggesting that these strains may have once had a history of co-occurrence and genetic exchange, despite the

isolation of the strains from different parts of the world (86-028NP in USA and Hi375 in Finland).

Isolation of the antibiotic resistance markers to nucleotide resolution. The frequency of donor-specific SNPs along the genome in the transformed recombinant pools (Pool 0) identified the known variants responsible for antibiotic resistance after a single round of strong selection (S4 Fig. parts A and B). All four pools started with nearly 100% donor-specific Nov^R or Nal^R alleles. Three recombinant pools were as expected (nearly 100% Nov^R or Nal^R donor-specific alleles), but a manipulation mistake during pooling from independent plates led to the RdS Nal^R pool being ~20% Nov^R (S4 Fig. part A, right panels); this did not interfere with mapping invasion loci/alleles. "Novel" alleles (base calls with neither donor nor recipient identity) occurred at a frequency consistent with overall error rates (S4 Table and S5 Table, and see below).

Because the donor and recipient genomes are divergent, clone sequencing could have identified recombinant segments spanning the selected site but not have isolated the exact resistance alleles [3]. Here, the high complexity of the recombinant pools after selection (~10^5 colonies) allowed for identification of the exact causative nucleotide, since donor alleles immediately adjacent to the resistance alleles had lower frequencies (S6 Table). The smooth decline in donor SNP frequencies on either side of the antibiotic-selected sites further underscores the pools' high complexities (S4 Fig. parts A and B), since less complex pools have a more saw-toothed or step-wise pattern [3]. Consistent with previous observations using a small set of sequenced recombinant clones [3], the decay in donor allele frequencies was asymmetric and non-uniform, suggesting variation in transformation frequencies for positions flanking the selected sites, at least in part due to the presence of nearby structural variation.

These results validate TREP, showing that causative single nucleotide differences can be identified directly from selected pools. In this situation, the transformation frequency of the causative alleles was high, the selection was very strong, and the alleles had very large effect sizes. For other phenotypes, like intracellular invasion,

serial selections would allow for weaker selections and concomitant enrichment of alleles with weaker effects, though possibly at the expense of reduced complexity in enriched pools, especially if transformation by the relevant alleles is infrequent.

Genome-wide transformation in the recombinant pools. Because TREP relies on the enrichment of donor-specific genetic variation through serial selection, these experiments do not need to be specifically designed to measure the low transformation frequencies (<1%) in the initial recombinant pools prior to selection, *per* se, which would require considerably higher sequencing depth, along with other experimental design and bioinformatics analysis considerations not needed for TREP, mostly having to do with transformation frequencies being similar to average Illumina sequencing error rates. Nevertheless, we were able to check the initial recombinant pools (Pool 0s) for evidence that transformation had occurred genome-wide, albeit with a relatively high limit-of-detection and in the context of sequencing errors. Several lines of evidence show that the recombinant pools carried donor-specific genetic variation of all types across the genome prior to serial selections for intracellular invasion and not only near the antibiotic-selected sites: (a) Compared to untransformed controls, non-reference variants with a donor base identity in the transformed pools had non-zero frequencies at substantially more positions (~3-fold or more, S6 Table). (b) "Donor" variant frequencies in the recombinant pools were significantly higher than in untransformed controls, where "donor" variants are just 1 of 3 possible sequencing/alignment errors (S4 Fig. part C, $p<<0.01$ for one-way ANOVA). (c) In contrast, untransformed controls had putative "donor" variant frequencies similar to or lower than "novel" variant frequencies (the frequency of the two bases that had neither donor nor recipient identity), as would be expected for sequencing/alignment artifacts (S4 Fig. part C; Tukey's HSD $p>0.1$ for comparisons of "donor" frequencies to error rate in controls, but $p<<0.001$ for transformed pools, compared to "novel" allele frequencies). Finally, (d) donor-specific structural variants (including donor-specific genes) were also incorporated into recombinant chromosomes in pools, indicated by sequence reads failing to

align to the recipient reference but aligning to the donor reference >5-fold better than reads from untransformed controls (S7 Table).

Suggestive of potential “hot” and “cold” regions for transformation, donor-specific SNP frequencies in the recombinant pools—but not the control strains—had a markedly bimodal distribution, with the low frequency mode being comparable to the rate of erroneous “donor” variants in untransformed controls (S4 Fig. part C). Unfortunately, these experiments do not accurately capture global transformation frequencies, due to insufficient coverage (limits of detection >0.1%). However, for purposes of TREP, the data demonstrate that the starting recombinant pools are highly complex and likely carry much of the genetic variation distinguishing the donor from the recipient. As long as phenotypic selection increases the frequency of recombinant (or mutant) clones above a few percent, sequencing artifacts will not confound identifying the donor-specific (or novel) genetic variation in selected clones (outside of the <1% of genomic positions with low coverage or ambiguous base calls).

SUPPLEMENTARY REFERENCES

1. Harrison A, Dyer DW, Gillaspy A, Ray WC, Mungur R, et al. (2005) Genomic sequence of an otitis media isolate of nontypeable *Haemophilus influenzae*: comparative study with *H. influenzae* serotype d, strain KW20. J Bacteriol 187: 4627-4636.
2. Mell JC, Shumilina S, Hall IM, Redfield RJ (2011) Transformation of natural genetic variation into *Haemophilus influenzae* genomes. PLoS Pathog 7: e1002151.
3. Mell JC, Lee JY, Firme M, Sinha S, Redfield RJ (2014) Extensive cotransformation of natural variation into chromosomes of naturally competent *Haemophilus influenzae*. G3 (Bethesda) 4: 717-731.
4. Wilcox KW, Smith HO (1975) Isolation and characterization of mutants of *Haemophilus influenzae* deficient in an adenosine 5'-triphosphate-dependent deoxyribonuclease activity. J Bacteriol 122: 443-453.
5. Catlin BW, Bendler JW, 3rd, Goodgal SH (1972) The type b capsulation locus of *Haemophilus influenzae*: map location and size. J Gen Microbiol 70: 411-422.
6. Hood DW, Makepeace K, Deadman ME, Rest RF, Thibault P, et al. (1999) Sialic acid in the lipopolysaccharide of *Haemophilus influenzae*: strain distribution, influence on serum resistance and structural characterization. Mol Microbiol 33: 679-692.
7. Barenkamp SJ, Leininger E (1992) Cloning, expression, and DNA sequence analysis of genes encoding nontypeable *Haemophilus influenzae* high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of *Bordetella pertussis*. Infect Immun 60: 1302-1313.
8. Grass S, Buscher AZ, Swords WE, Apicella MA, Barenkamp SJ, et al. (2003) The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. Mol Microbiol 48: 737-751.
9. Tracy E, Ye F, Baker BD, Munson RS, Jr. (2008) Construction of non-polar mutants in *Haemophilus influenzae* using FLP recombinase technology. BMC Mol Biol 9: 101.
10. Sanchez R (1998) A medium-copy-number plasmid for insertional mutagenesis of *Streptococcus mutans*. Plasmid 40: 247-251.

Supplementary Figures

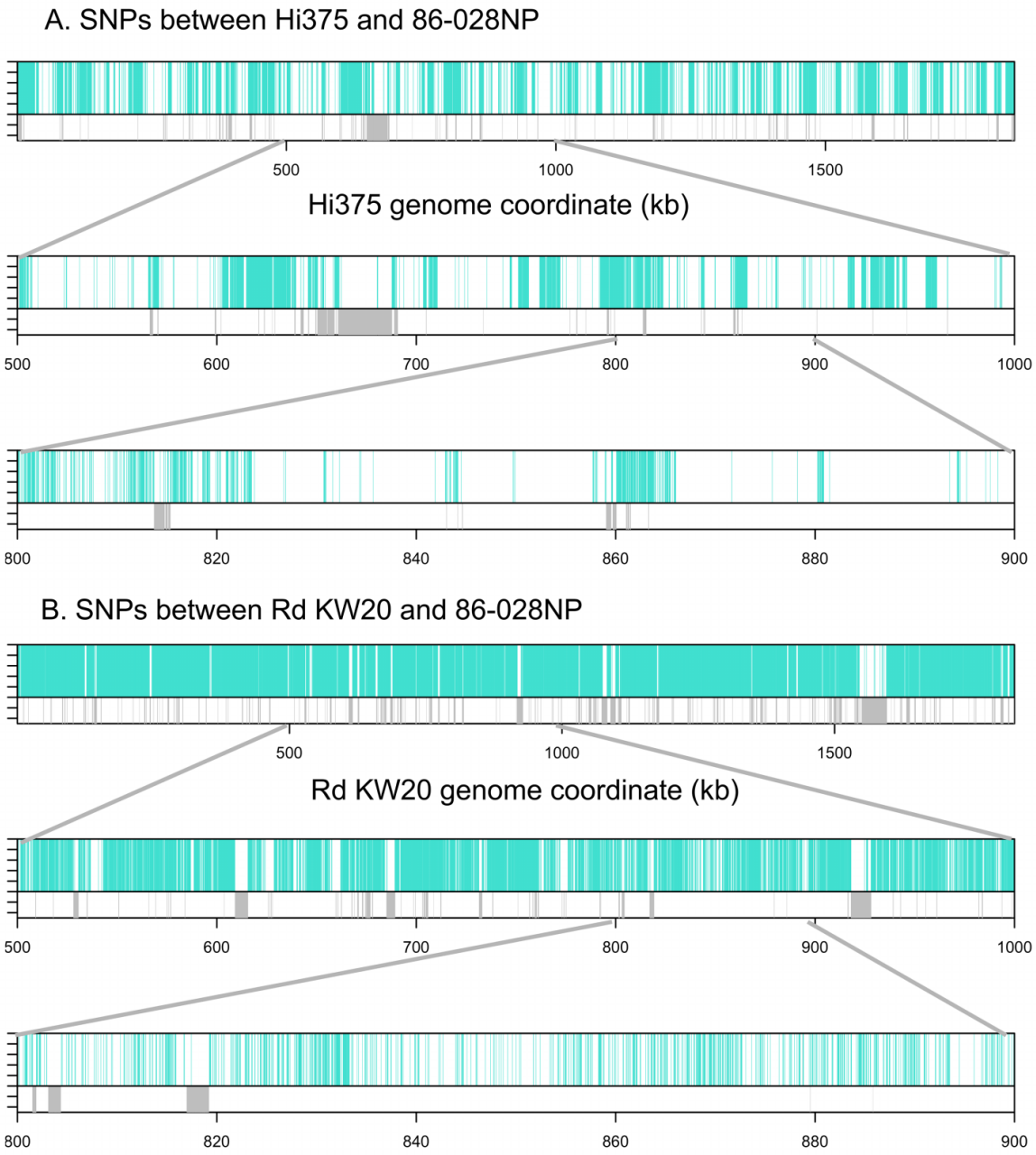


Figure S2.1. Comparison of the donor to the two recipients at different scales. Turquoise lines above the x-axis indicate the position of SNPs distinguishing donor from recipient, while grey lines below the x-axis indicate positions in the recipient genome missing from the donor genome (at indels). (**A**) Hi375 recipient. (**B**) Rd KW20 recipient. Note that SNPs between Hi375 and 86-028NP are punctate, with stretches of very low SNP density punctuated by stretches of high SNP density. Genomic positions exclusive to the recipient strains are shown in grey; these coincide with areas that appear as regions of low SNP density, but these artifacts are insufficient to explain the pattern seen in Hi375. Conversely Rd-specific positions do largely explain low SNP density regions in Rd KW20.

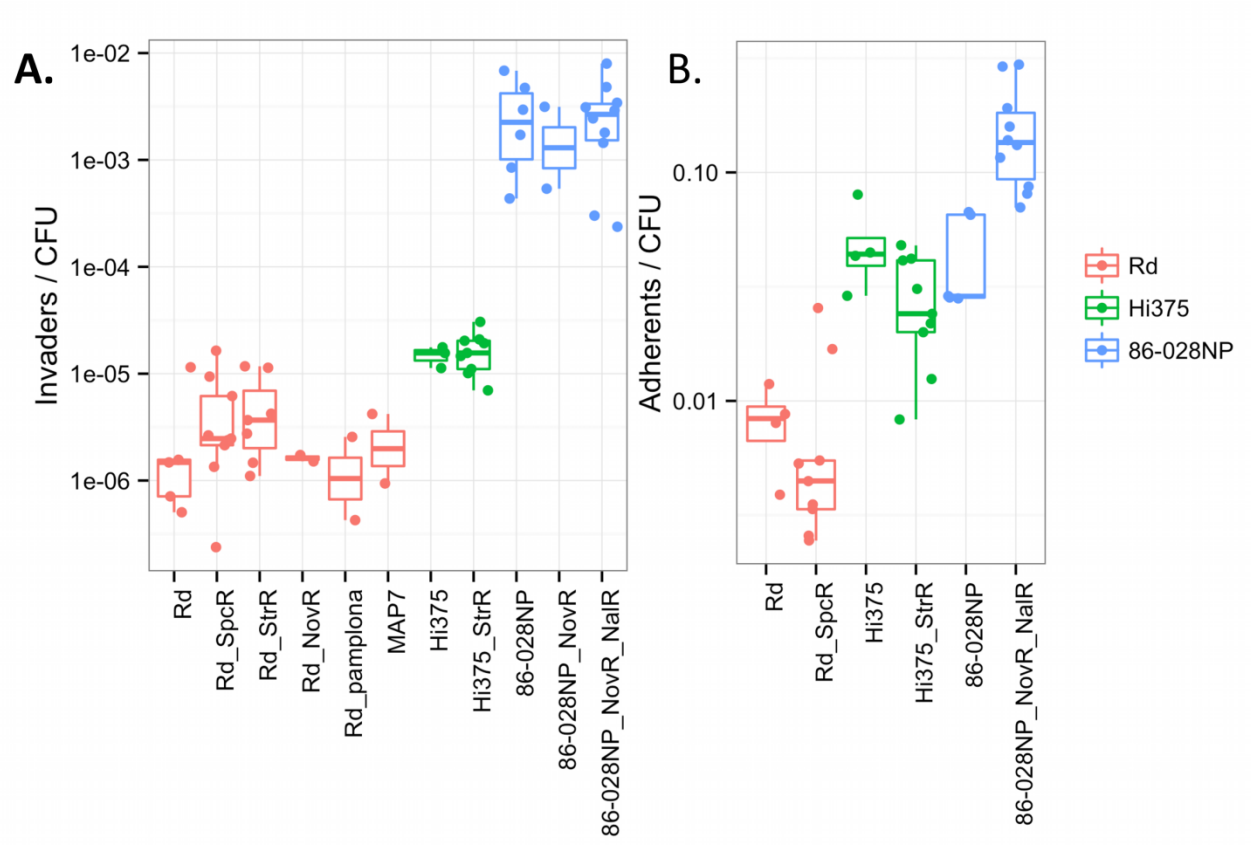


Figure S2.2. Invasion and adhesion phenotypes of parental and related strains. (**A**) Invasion of and (**B**) adhesion to A549 cells is shown for H. influenzae strains Rd KW20, Hi375, 86-028NP, and antibiotic resistant derivatives, including the parental strains.

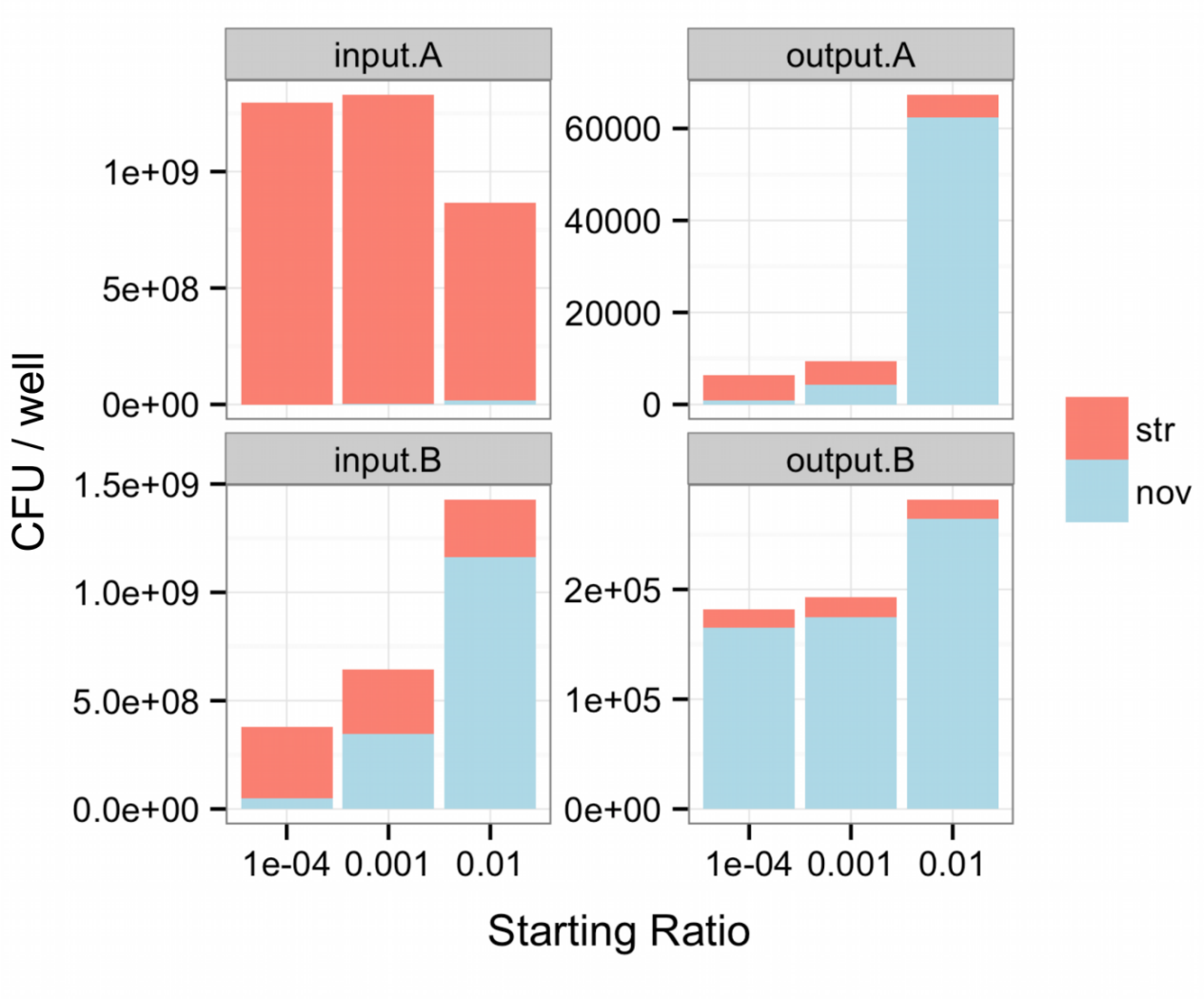


Figure S2.3. Competition for invasion between Rd and 86-028NP strain backgrounds. Two serial cycles of selection for intracellular invaders were conducted using three mixtures of 86-028NP Nov^R and Rd Str^R cells, at 1:100, 1:1,000, or 1:10,000 ratios. Prior to the first infection (input.A), the bacterial cell suspension was titrated for the total Nov^R and Str^R CFU used per well, and this closely matched the expected frequencies. After the first round of selection (output.A), dramati- cally fewer CFU were recovered, but Nov^R were proportionally much more abundant. Total unse- lected CFUs were pooled and titrated (input.B), showing that the proportion of Nov^R remained relatively the same in between cycles of selection for invasion. Finally, the second cycle of selection resulted in a higher yield with an even higher proportion of Nov^R colonies, representing a strong enrichment of 86-028NP over Rd, even when at a low relative abundance in the starting mixture.

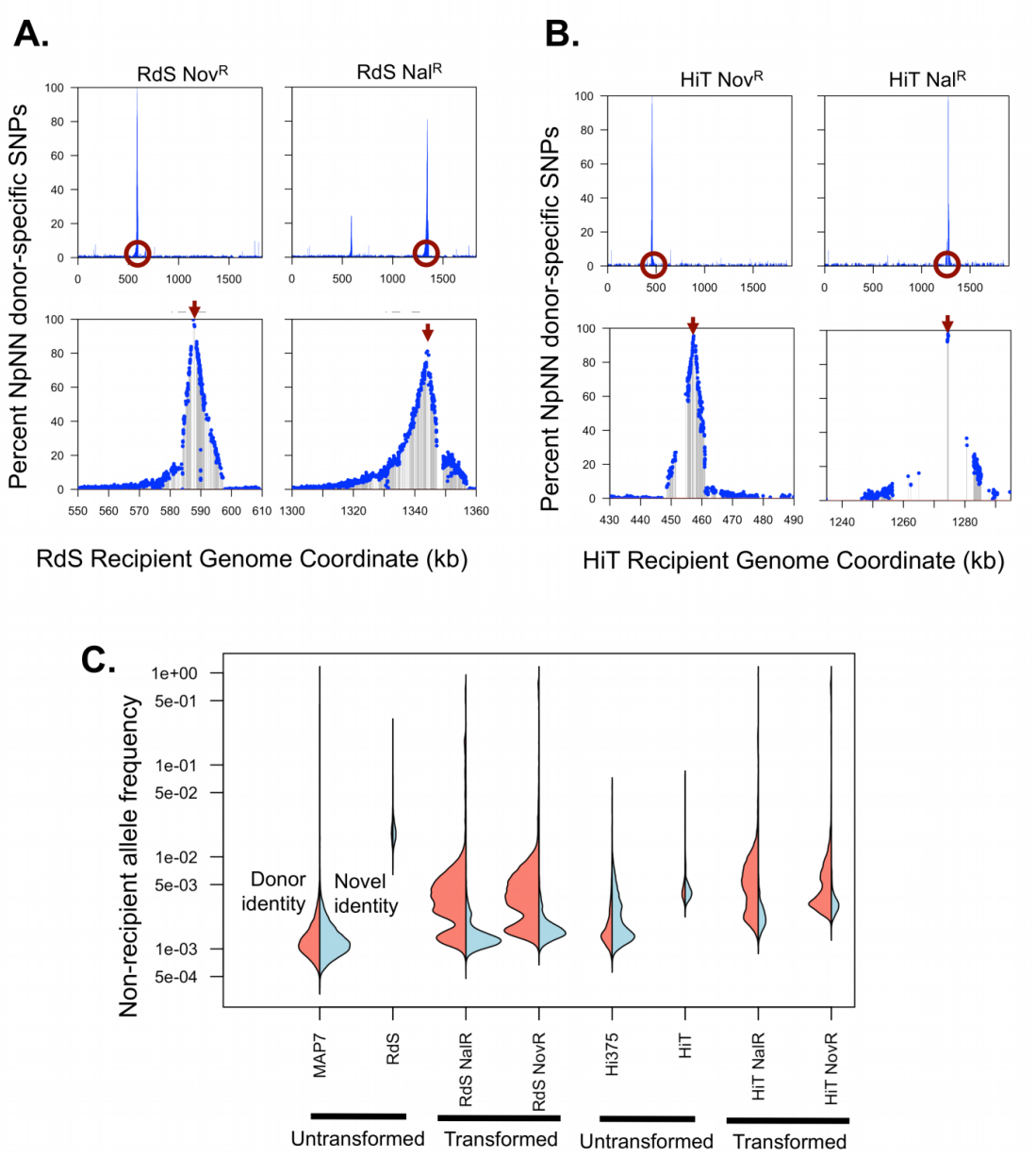


Figure S2.4. Donor allele frequencies in the transformed input pools. (**A**) and (**B**) NpNN-specific SNP frequencies as a function of chromosome coordinate for the RdS and HiT recipients, respec- tively, at Pool 0, prior to enrichment for invasive recombinants. Left panels: Nov^R-selected pools. Right panels: Nal^R-selected pools. Top panels: chromosome-wide view. Bottom panels: zoom on 60 kb windows around the antibiotic resistance markers. The peak SNP is the one conferring antibiotic resistance. (**C**) "Bean plots" summarizing 16 histograms of non-recipient allele fre- quencies for untransformed controls and the initial transformed recombinant pools. The left side (salmon-colored) of each bean shows a histogram for allele frequencies with donor allele identi- ties, whereas the right side (light blue) shows a histogram for "novel" alleles (neither recipient nor donor). The latter are sequencing errors, while the former are sequencing errors for the con- trol strains and a combination of sequencing errors and transformants for the transformed pools.

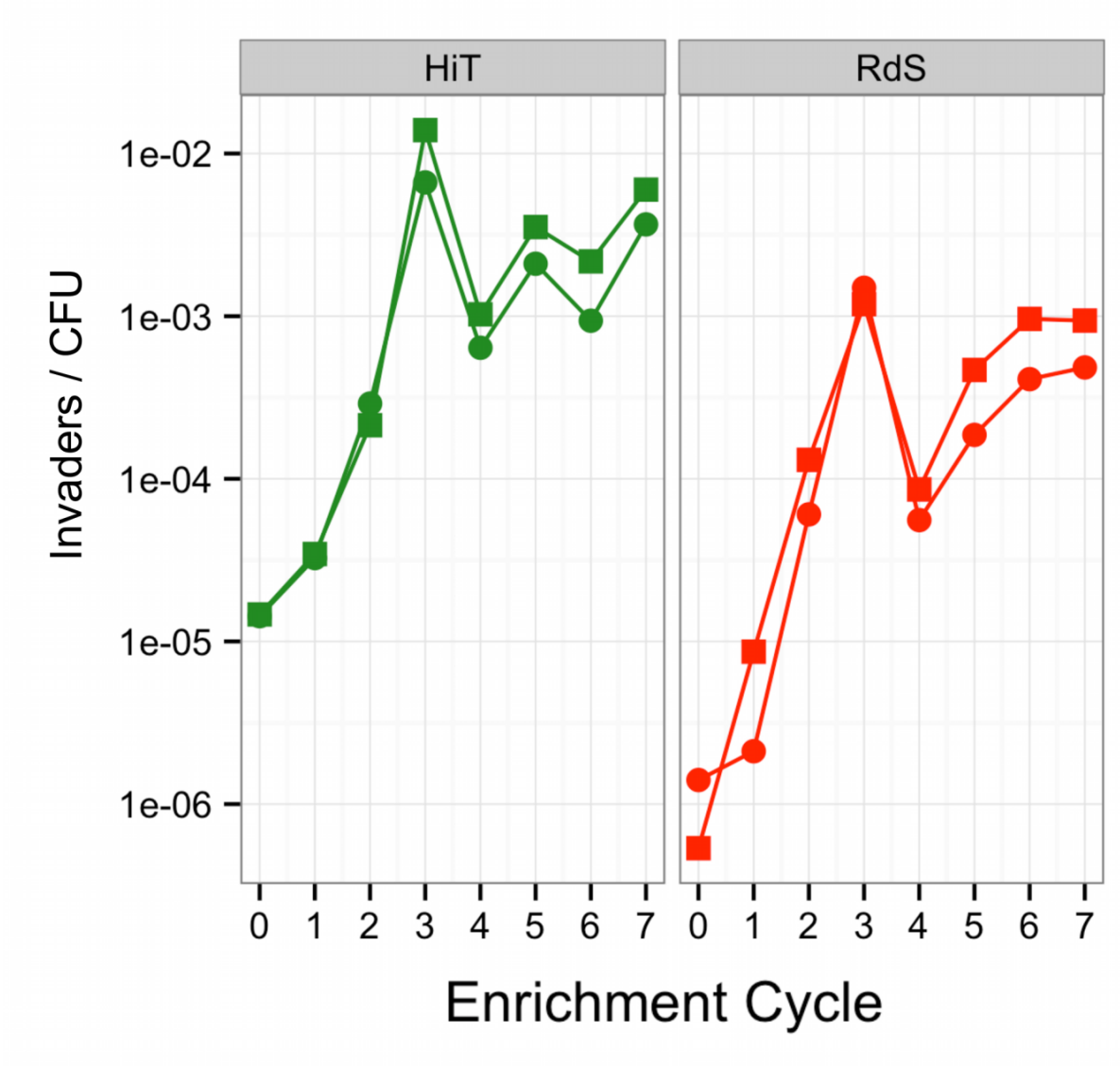


Figure S2.5. Invaders/CFU for pools during the initial eight serial selections for invasive recombinants. Recovered CFU that survived gentamicin treatment (Pool 1) served as input for the next cycle (which generated Pool 2). The values show the combined ability of clones in Pool n to invade airway epithelial cells, while the recovered colonies comprise Pool n+1. This procedure was carried out eight times. The apparent decline in invasiveness seen at Pool 4 appears to be an artifact, since no such decline was seen in the replicate assays (**Figure 2.3A**). Instead, this drop likely reflects that Pool 4 bacteria had been frozen and re-inoculated prior to the next cycle, combined with batch-to- batch variation of the confluent A549 cells used.

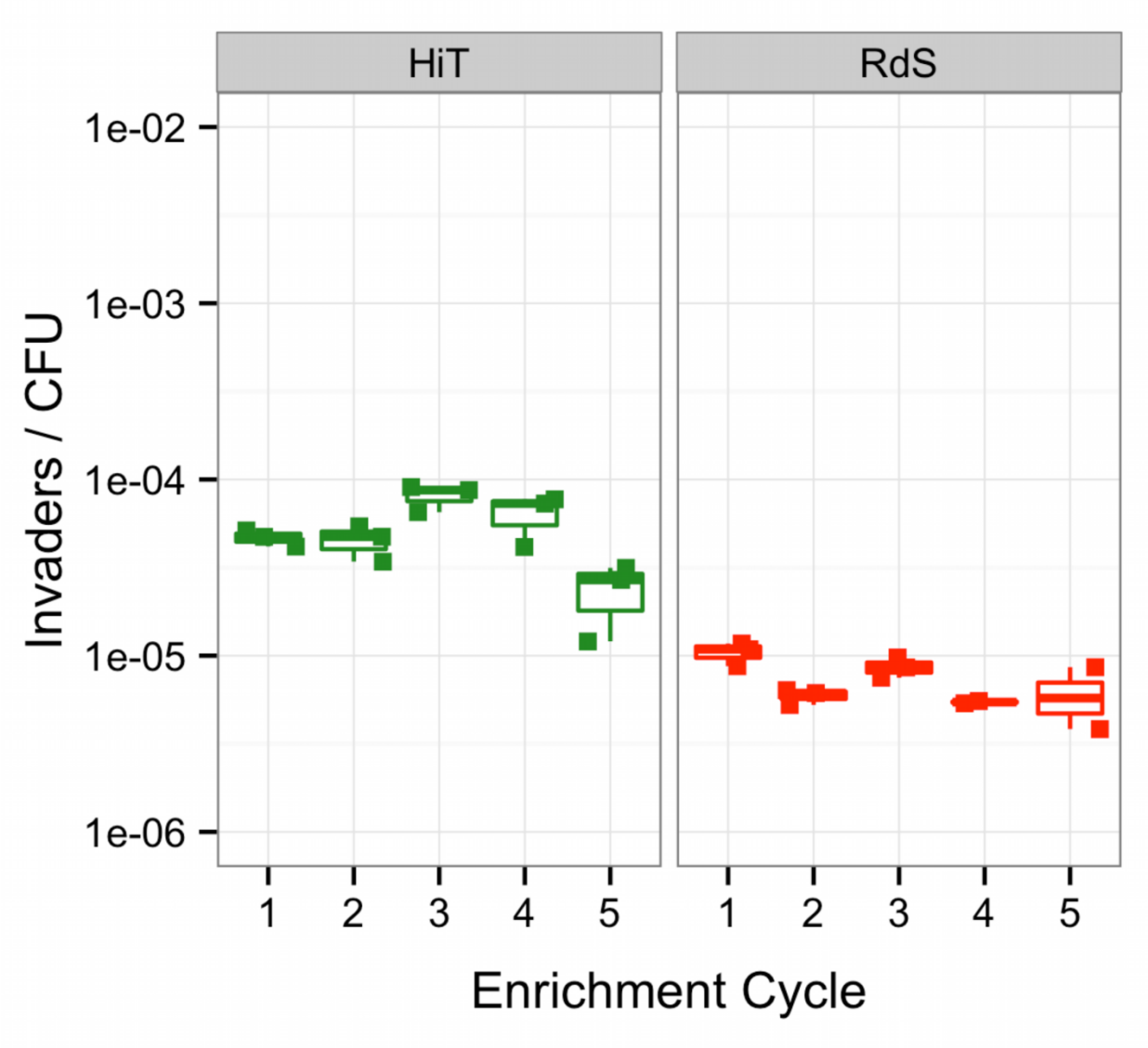


Figure S2.6. No improvement by selection on untransformed recipients. Control experiment using untransformed recipients cultures in triplicate found no increase in invasiveness over 5 serial selections. This experiment was conducted independently for each of the recipients and separately from the experimental enrichments to minimize enrichment of cross-contaminants.

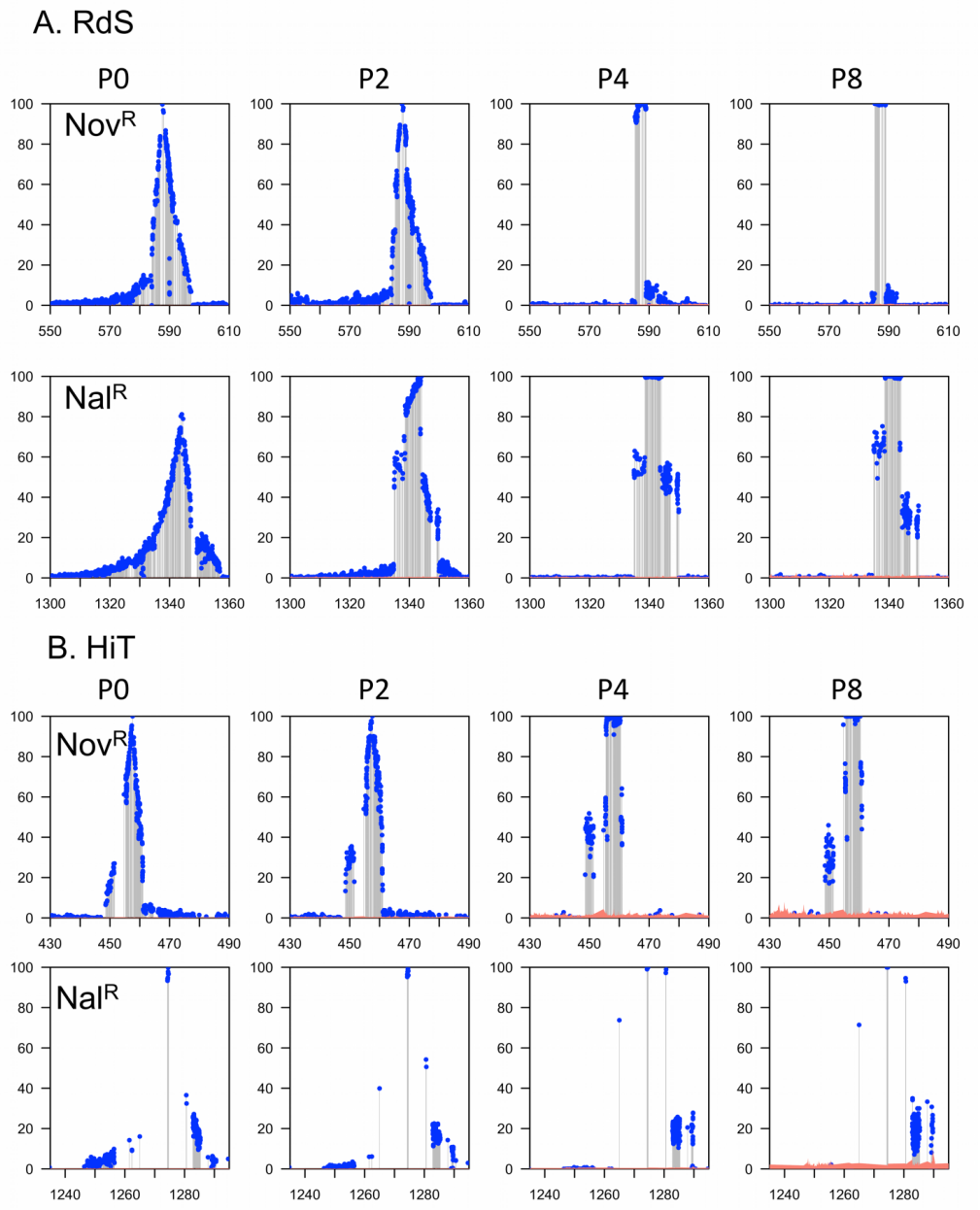


Figure S2.7. Complexity of recombination tracts decreases at the antibiotic resistance markers over serial passages. Genomic profiling at antibiotic-selected sites for both the (**A**) RdS and (**B**) HiT recipients at Nov^R (top) and Nal^R (bottom) sites (gyrB and gyrA respectively, see **Table S2.6**) for Pools 0, 2, 4, and 8. Axes are as in other figures with x-axes indicating recipient genome coordinate (in kb) and the y-axis indicating donor allele frequency. RdS Nov^R contains a single clone at ~95% by Pool 8, while RdS Nal^R contains two dominant clones, one at ~70% and the other ~30%. HiT Nov^R contains two dominant clones (at ~30% and 70%), whereas HiT Nal^R appears to contain two clones at ~80% and ~20%. For this pool, only a single geno- type (the one at ~80%) was recovered in the four individual clones collected from Pool 4. No other donor segments appeared at ~20%, so this is likely due to incomplete fixation of the inva- sive genotype after several rounds of selection.

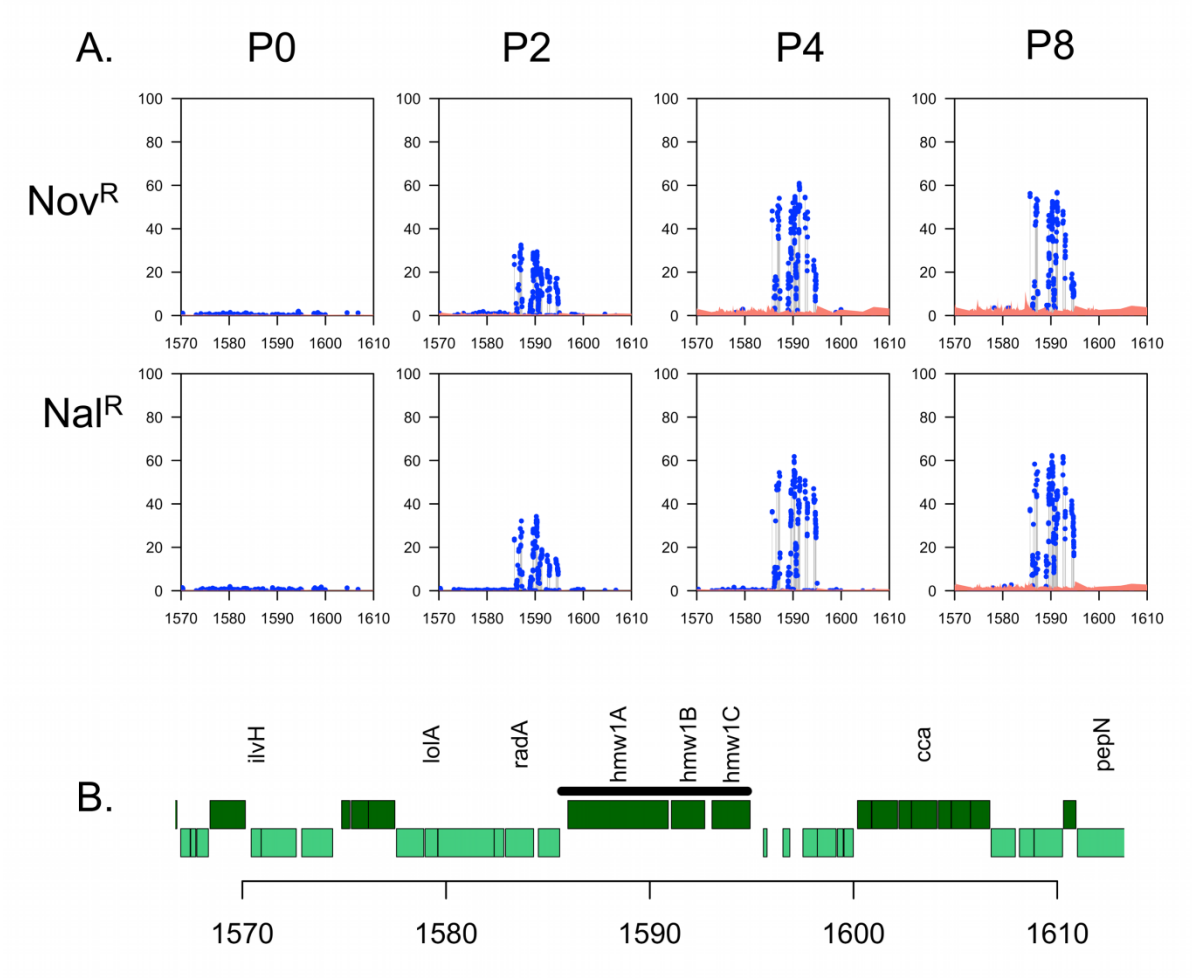


Figure S2.8. Read alignment artifact at the radA-proximal hmw1 locus in the HiT pools. (**A**) Pools 0, 2, 4, and 8 for HiT NovR and HiT NalR as in other figures (x-axis is HiT recipient coordi- nate in kb, and y-axis is donor allele frequency). (**B**) Genomic map around the same interval. The thick black horizontal line shows the entire range of positions containing donor frequencies >5%. The affected interval spans only the hmw1 locus; no flanking variation was detected, unlike the situation at the yrbI-adjacent $hmw2_{Hi375}$, which was replaced by the $hmw1_{86\text{-}028NP}$ allele. Donor allele frequencies are highly variable in this region. They are also highly consistent between the two pools, which was unexpected, as all other overlapping donor segments detected had distinct recombination breakpoints. Allele-specific PCR assays confirm this as read alignment artifact and confirm that the radA-proximal adhesin remain $hmw1_{Hi375}$ across strains (**Figure S2.9**).

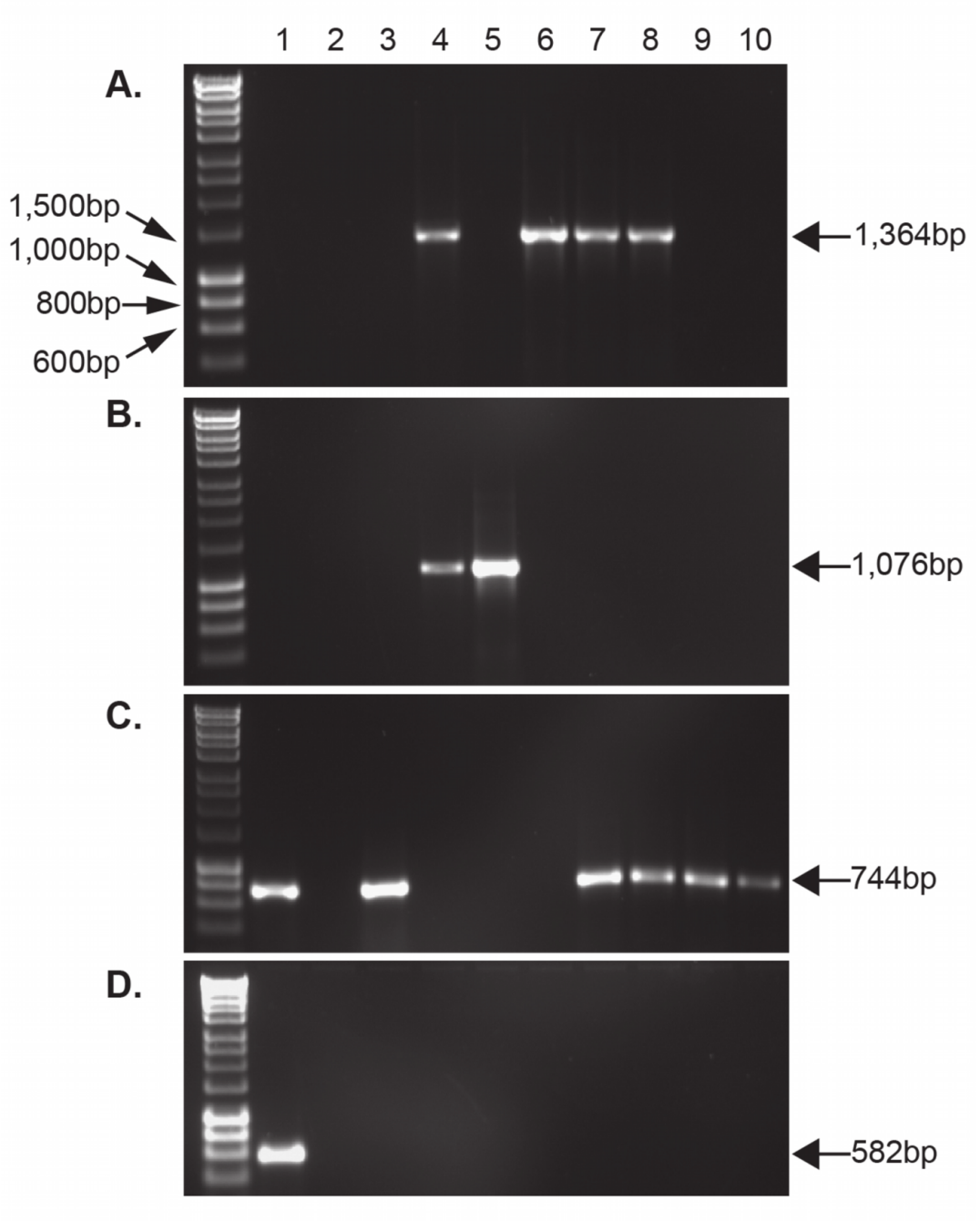


Figure S2.9. Agarose gel showing allele/locus-specific PCR products amplified for the four possible hmw alleles. Strains are listed as it follows: (1) NpNN, (2) RdS, (3) rRdS, (4) HiT, (5) HiTΔhmw1A$_{Hi375}$, (6) HiTΔhmw2A$_{Hi375}$, (7) rHiT, (8) rHiTΔhmw1A$_{86\text{-}028NP}$, (9) rHiTΔhmw1A$_{Hi375}$, (10) rHiTΔhmw1A$_{86028NP}$Δhmw1A$_{Hi375}$, and primers are in **Table S2.10**. (**A**) Primers 1456+1458 identify hmw1A$_{Hi375}$ (1,364 bp product); (**B**) primers 1456+1457 identify hmw2A$_{Hi375}$ (1,076 bp product); (**C**) primers 1459+1460 identify hmw1A$_{86\text{-}028NP}$ (744 bp product); and (**D**) primers 1461+1462 for hmw2A$_{86\text{-}028NP}$ (582 bp product). Lanes 8 and 10 rendered a correct size band upon PCR with primers 1459+1460 because mutant strains lacking hmw1A$_{86\text{-}028NP}$ were generated by partial deletion that maintains the annealing sites for the primers and product size.

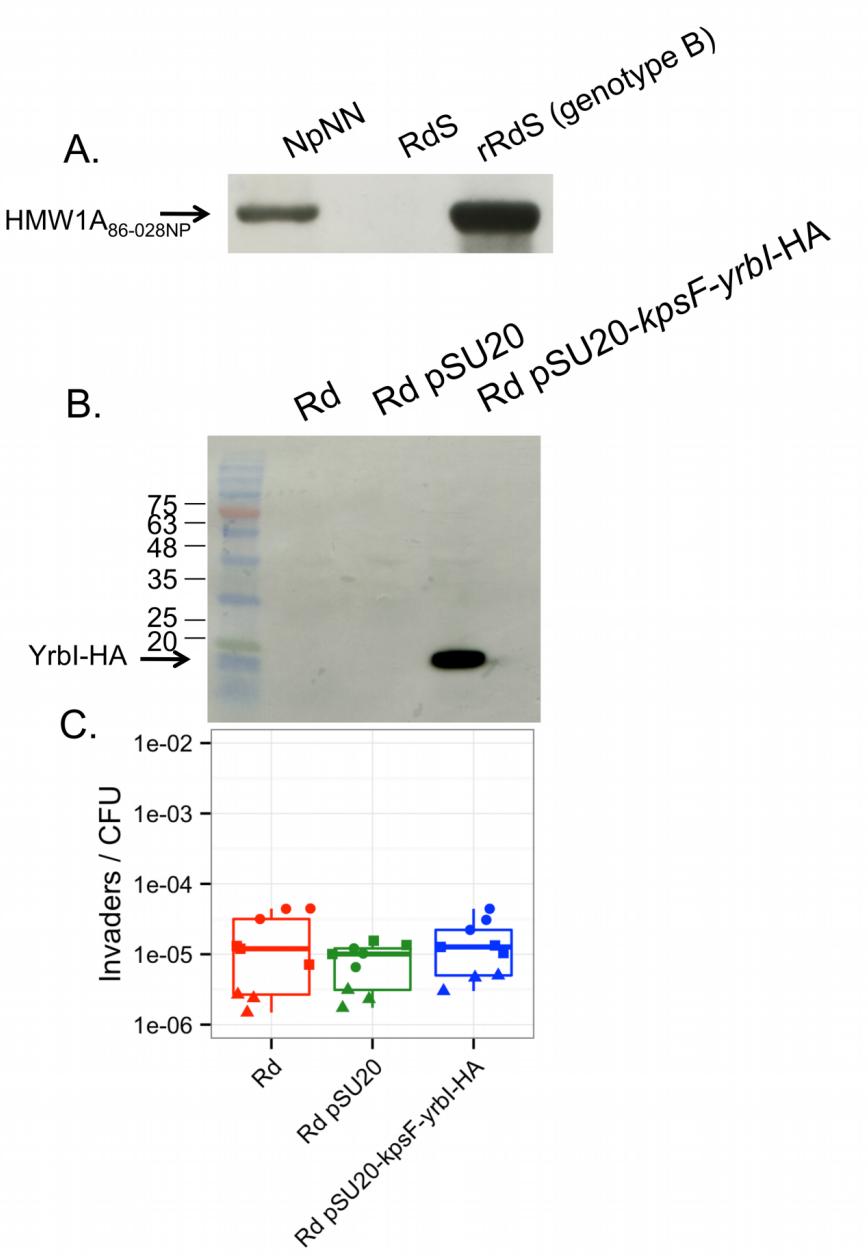


Figure S2.10. Confirmation of HMW adhesion expression and testing for a role by kpsF and yrbI. (A) Western blot showing expression of $HMW1A_{86\text{-}028NP}$ (154 KDa) adhesin. Whole cell extracts of NpNN, RdS and rRdS (P540, genotype B) were prepared and used to detect HMW by immunoblot with the guinea pig anti-HMW1A gp85 antibody. (**B** and **C**) Addition of the kpsF and yrbI alleles from 86-028NP on a plasmid does not increase intracellular invasion frequencies. (**B**) Western blot showing expression from plasmid carrying an interval carrying kpsF-yrbI from 86-028NP. Whole cell extracts of cultures (Rd, Rd pSU20, and Rd pSU20-kpsF-yrbI-HA) were prepared and used to detect Hap-HA by immunoblot with a rabbit anti-HA antibody, finding expression of the expected ~19.3-kDa protein in the expected strain. (**C**) The same strains were used to infect A549 cells and measure bacterial intracellular invasion. Experiments were performed three times in triplicate (different symbols denote independent experiments).

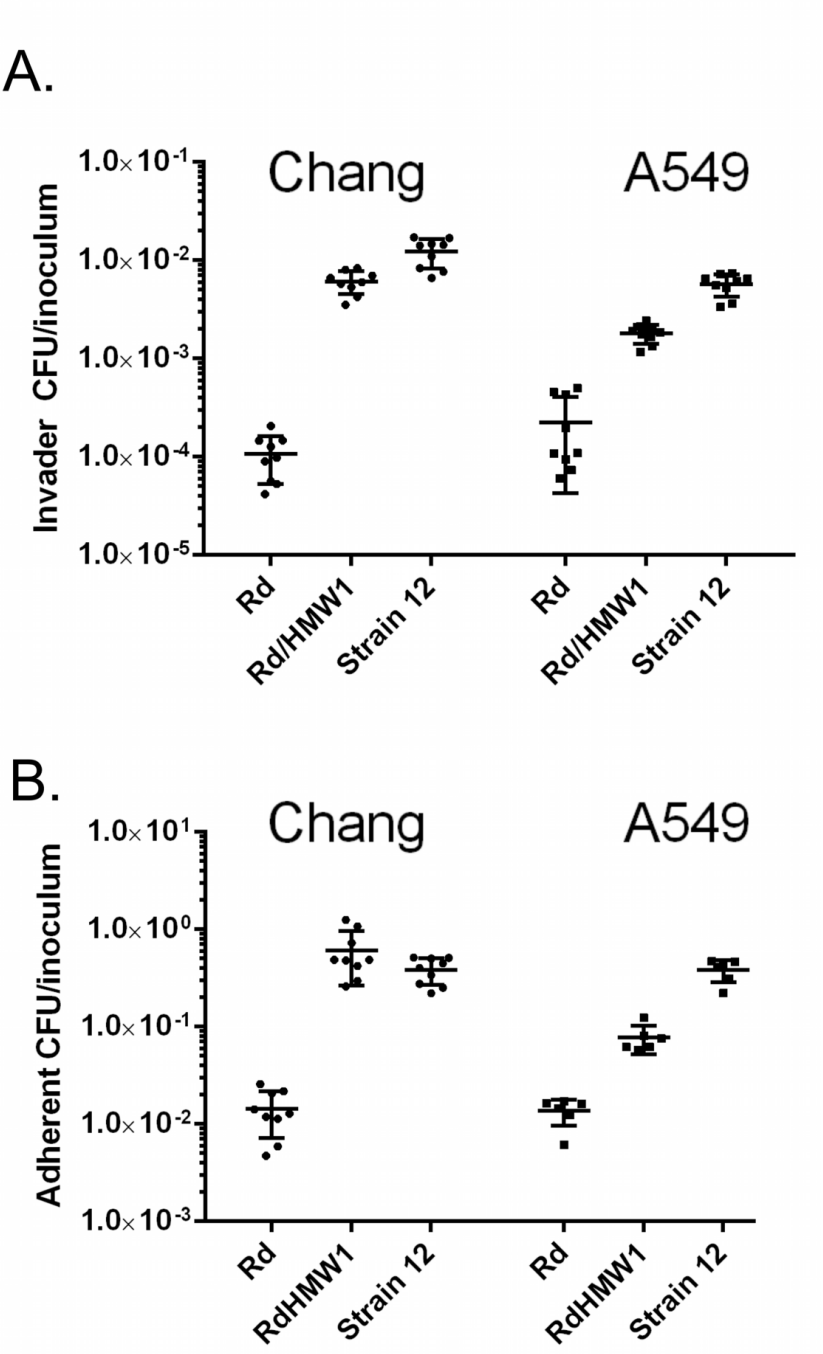


Figure S2.11. A role for HMW1 is seen for a distinct strain, for another epithelial cell type, and with an alternative protocol. Invasion (**A**) and adhesion (**B**) by Rd, Rd hmw1$_{strain12}$, and Strain 12 bacterial strains into Chang and A549 epithelial cell lines. An alternative protocol that includes centrifugation to quickly bring bacteria into contact with the cell monolayer was used for these experiments, showing that both cell type and details of the infection procedure give qualitatively similar results.

Supplementary Tables

Table S2.1. Strains and plasmids used.

Strain or plasmid	Description	Reference or source**
H. influenzae		
P189	Rd KW20, rough (unencapsulated) derivative of type d, source: R.J. Redfield (*aka* RR722)	[4]
P192 (MAP7)	Rd Nov^R Nal^R Str^R Spc^R Kan^R Rif^R (*aka* RR666)	[5]
P532 (RdS)	Rd Spc^R, made by transforming P189 with P192 DNA, selecting for Spc^R, and screening against other MAP7 resistance alleles	This study
P193	Rd Str^R, made as above but selecting for Str^R (*aka* RR514)	This study
P194	Rd Nov^R, made as above but selecting for Nov^R (*aka* RR3148)	This study
P1171	NTHi375, otitis media clinical isolate	[6]
P531 (HiT)	NTHi375 Str^R, made by transformation of the *rpsL* amplified from strain P193	This study
P190	86-028NP, otitis media clinical isolate	[1]
P195	86-028NP Nov^R (*aka* RR3129)	[2]
P351 (NpNN)	86-028NP Nal^R Nov^R (*aka* RR3131)	[2]
P809	Strain 12, otitis media clinical isolate, source: J. St. Geme (*aka* R2846)	[7]
P812	Rd KW20, source: J. St. Geme	[4]
P813	Rd/HMW1, made by addition of $hmw1ABC_{strain12}$ to Rd, source: J. St. Geme	[8]
P540 (rRdS)	Rd Spc^R, recombinant clone Nal^R s2, genotype B (rRdS)	This study
P551 (rHiT)	NTHi375 Str^R, recombinant clone Nov^R s1, genotype E (rHiT)	This study
P834	HiT rec-genotype E$\Delta hmw1A_{86\text{-}028NP}$ (rHiT $\Delta hmw1A_{86\text{-}028NP}$)	This study
P836	HiT$\Delta hmw1A_{Hi375}$	This study
P837	HiT$\Delta hmw2A_{Hi375}$	This study
P838	HiT rec-genotype E$\Delta hmw1A_{Hi375}$ (rHiT$\Delta hmw1A_{Hi375}$)	This study
P839	HiT rec-genotype E$\Delta hmw1A_{86\text{-}028NP}\Delta hmw1A_{Hi375}$ (rHiT$\Delta hmw1A_{86\text{-}028NP}\Delta hmw1A_{Hi375}$)	This study
E. coli		
Top10	Cloning strain. F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G	Thermofisher Scientific
SW102/DY380	*E. coli* strain expressing lambda recombinase	[9]
Plasmid		
pGEMT-easy	Cloning vector	Promega
pJET2.1	Cloning vector	Life Technologies
pJET2.1-(A-Erm^R-B)$_{hmw1AHi375}$	pJET2.1 with a 3,578 bp insert containing a $hmw1A_{Hi375}$ disruption cassette	This study
pJET2.1-(A-Erm^R-B)$_{hmw2AHi375}$	pJET2.1 with a 3,853 bp insert containing a $hmw2A_{Hi375}$ disruption cassette	This study
pGEMT-(A-Spc^R-B)$_{hmw1A86028NP}$	pGEMT-easy with a 4,253 bp insert containing a $hmw1A_{86\text{-}028NP}$ disruption cassette	This study
pSU20	pACYC184 derivative, Cm^R, 2,334 bp, shuttle vector for *H. influenzae* that allows for XGal screening	[10]
pSU20-(*Pr::kpsF-yrbI*-HA)	pSU20 with a 2,241 bp insert containing *kpsF* and *yrbI* genes expressed under their own promoter	This study

Table S2.2. Pool and control sequencing statistics.

SAMN0	Sample	Ref	Pairs	Est. Cov.	%Merge	%Dup	%RecipMap	%DonorMap
4392945	Hi375	Hi375	9,396,398	1879.3	55.80%	5.50%	99.90%	96.70%
4392946	HiT	Hi375	3,188,760	637.8	61.30%	7.10%	99.90%	99.10%
4392947	MAP7	Rd	11,986,693	2397.3	50.80%	8.50%	99.90%	94.00%
4392948	RdS	Rd	654,260	130.9	57.20%	2.40%	99.90%	97.80%
4392949*	NpNN	Hi375	1,376,875	275.4	34.60%	1.90%	95.70%	99.80%*
4392949*	NpNN	Rd	1,376,875	275.4	34.50%	2.00%	92.50%	99.80%*
4392950	HiT_NalR_P0	Hi375	5,323,837	1064.8	35.30%	5.40%	99.90%	98.80%
4392951	HiT_NalR_P1	Hi375	1,351,021	270.2	34.60%	2.80%	98.60%	98.10%
4392952	HiT_NalR_P2	Hi375	6,505,976	1301.2	32.50%	5.90%	99.90%	98.90%
4392953	HiT_NalR_P3	Hi375	1,173,018	234.6	31.60%	3.10%	99.00%	98.40%
4392954	HiT_NalR_P4	Hi375	2,602,566	520.5	47.10%	4.80%	98.90%	98.70%
4392955	HiT_NalR_P5	Hi375	498,668	99.7	46.90%	0.80%	96.60%	97.60%
4392956	HiT_NalR_P6	Hi375	227,374	45.5	29.80%	1.20%	98.60%	98.30%
4392957	HiT_NalR_P7	Hi375	832,148	166.4	30.40%	1.50%	98.30%	98.20%
4392958	HiT_NalR_P8	Hi375	633,131	126.6	29.10%	1.40%	99.80%	98.90%
4392959	HiT_NovR_P0	Hi375	4,399,056	879.8	56.50%	4.80%	99.90%	99.10%
4392960	HiT_NovR_P1	Hi375	627,900	125.6	31.60%	2.10%	98.30%	98.00%
4392961	HiT_NovR_P2	Hi375	1,859,651	371.9	36.50%	4.30%	98.40%	98.10%
4392962	HiT_NovR_P3	Hi375	594,415	118.9	34.70%	3.50%	98.60%	98.20%
4392963	HiT_NovR_P4	Hi375	575,382	115.1	32.70%	4.60%	99.10%	98.40%
4392964	HiT_NovR_P5	Hi375	1,259,247	251.8	39.80%	2.00%	97.90%	97.70%
4392965	HiT_NovR_P8	Hi375	489,260	97.9	64.10%	0.90%	99.80%	99.10%
4392966	RdS_NalR_P0	Rd	10,295,100	2059	51.80%	6.80%	99.90%	97.80%
4392967	RdS_NalR_P2	Rd	3,002,546	600.5	65.10%	2.20%	99.50%	98.20%
4392968	RdS_NalR_P3	Rd	1,168,294	233.7	28.90%	3.30%	99.50%	97.60%
4392969	RdS_NalR_P4	Rd	2,293,706	458.7	30.20%	2.90%	98.60%	97.30%
4392970	RdS_NalR_P5	Rd	1,494,058	298.8	41.10%	1.40%	97.60%	96.90%
4392971	RdS_NalR_P6	Rd	118,518	23.7	33.00%	0.60%	98.30%	97.10%
4392972	RdS_NalR_P7	Rd	974,181	194.8	38.00%	1.60%	98.50%	97.30%
4392973	RdS_NalR_P8	Rd	1,252,651	250.5	61.40%	1.00%	99.40%	98.00%
4392974	RdS_NovR_P0	Rd	8,521,941	1704.4	55.70%	6.40%	99.80%	98.00%
4392975	RdS_NovR_P2	Rd	3,801,127	760.2	36.30%	3.50%	99.80%	97.70%
4392976	RdS_NovR_P3	Rd	4,030,064	806	46.00%	3.50%	99.60%	97.70%
4392977	RdS_NovR_P4	Rd	1,784,374	356.9	32.80%	4.00%	99.70%	97.40%
4392978	RdS_NovR_P5	Rd	1,346,587	269.3	37.90%	1.50%	99.60%	97.70%
4392979	RdS_NovR_P6	Rd	222,740	44.5	35.90%	1.00%	97.90%	96.90%
4392980	RdS_NovR_P7	Rd	1,357,364	271.5	37.10%	1.60%	98.60%	97.30%
4392981	RdS_NovR_P8	Rd	1,930,259	386.1	52.60%	1.80%	99.50%	97.90%

* Reads were submitted as BAM alignments to the 86-028NP reference genome.

Columns are as follows:
SAMN0: the BioSample accession number for the sample as SAMN0######.
Sample: Sample listed as strain name or recipient_selectedDonorMarker_PoolOrCloneID.
Ref: Reference genome for "self"-alignment.
Pairs: Total read pairs collected by Illumina sequencing (2x101nt).
Est. Cov.: Estimated genomic coverage based on total read pairs and genome size.
% Merge: The percent of reads that were merged by COPE.
% Dup: The percent of reads whose mappings were deemed duplicates by SamBlaster.
%RMap: Percent of reads that mapped to the appropriate recipient genome (Rd or Hi375).
%DMap: Percent of reads that mapped to the donor genome (86-028NP).

Table S2.3. Recombinant clone sequencing statistics

SAMN0	Strain	Ref.	Pairs	Est. Cov.	%Merge	%Dup	%Nunmap	%RMap	%DMap
4392982	HiT_NalR_s1	Hi375	192,542	28.9	9.60%	4.40%	26.89%	99.54%	97.45%
4392983	HiT_NalR_s2	Hi375	228,743	34.3	16.00%	4.30%	23.07%	99.59%	97.63%
4392984	HiT_NalR_s3	Hi375	260,781	39.1	23.20%	3.30%	15.99%	99.64%	97.92%
4392985	HiT_NalR_s4	Hi375	156,888	23.5	6.60%	4.00%	26.42%	99.55%	97.59%
4392986	Hit_NovR_s1	Hi375	289,852	43.5	18.00%	3.70%	17.70%	99.70%	98.06%
4392987	Hit_NovR_s2	Hi375	216,142	32.4	12.80%	3.80%	21.60%	99.68%	97.96%
4392988	Hit_NovR_s3	Hi375	348,917	52.3	20.20%	3.60%	15.40%	99.46%	97.89%
4392989	Hit_NovR_s4	Hi375	436,820	65.5	2.90%	8.60%	27.81%	99.45%	97.32%
4392990	RdS_NalR_s1	Rd	142,686	21.4	3.90%	2.40%	30.95%	92.32%	89.13%
4392991	RdS_NalR_s2	Rd	244,256	36.6	7.20%	6.80%	29.69%	99.42%	95.05%
4392992	RdS_NalR_s3	Rd	248,886	37.3	25.60%	4.90%	24.57%	99.24%	94.38%
4392993	RdS_NalR_s4	Rd	301,960	45.3	23.00%	4.60%	21.23%	99.31%	95.24%
4392994	RdS_NovR_s1	Rd	201,572	30.2	27.50%	2.90%	20.58%	98.57%	94.08%
4392995	RdS_NovR_s2	Rd	177,728	26.7	23.60%	2.80%	21.50%	99.32%	95.07%
4392996	RdS_NovR_s3	Rd	414,147	62.1	4.00%	9.50%	27.68%	99.51%	95.38%
4392997	RdS_NovR_s4	Rd	317,691	47.7	8.30%	7.40%	26.77%	99.18%	95.43%

Columns as in **Table S2.2**, except **%Nunmap** estimates uninformative reads as unmapped reads with strings of 5+ Ns, and **%RMap** and **%DMap** are adjusted to exclude these uninformative reads. A high proportion of clone reads from two independent sequencing experiments (independent DNA preps, library preps, and Illumina MiSeq runs) yielded a high proportion of reads that consisted of all or mostly all Ns. The reasons for this are unknown but were likely due to problems with a particular batch of Nextera XT library kits. The low coverage data was nonetheless able to corroborate the recombination breakpoints detected in the high quality high coverage pool datasets, as well as disambiguate donor segment assignment to individual clones. Aggregating datasets by clone genotype (as in **Table S2.7**) further validated donor segment assignments in **Table S2.8**.

Table S2.4: Alignment statistics for the untransformed controls.

Reads Reference	Hi375 Hi375	HiT Hi375	MAP7 Rd KW20	RdS Rd KW20	NPNN 86-028NP
Total positions	1,850,897	1,850,897	1,830,138	1,830,138	1,914,490
Ambiguous reference[1]	-	-	115	115	-
Depth=0	22	55	116	182	1
Depth<10	100	195	156	689	56
Depth<50	233	545	261	783,986	20,256
Filtered[2]	1,850,797	1,850,702	1,829,867	1,829,334	1,914,434
% Invariant[3]	54.4%	92.0%	61.6%	97.6%	92.5%
Mean Depth	652.2	227.5	822.9	49.0	107.6
Mean Limit-of-detection (1/depth)	1.53E-03	4.40E-03	1.22E-03	2.04E-02	9.30E-03
Mean VarFreq	1.12E-03	3.92E-04	7.89E-04	6.45E-04	7.21E-04
Strand-biased positions[4]	4001	657	2811	175	643
Unbiased Mean VarFreq	1.12E-03	3.89E-04	7.84E-04	6.45E-04	7.14E-04
Biased Mean VarFreq	3.60E-03	9.36E-03	4.11E-03	2.28E-03	2.16E-02
% Unbiased with VarFreq >0.01	0.29%	0.17%	0.11%	2.37%	2.02%
% Biased with VarFreq >0.01	6.20%	17.66%	6.79%	3.43%	29.86%
Detected variants (VarFreq >0.95)[5]	0	8	326	292	48

[1] In the Rd KW20 genome, 115 positions have a non-ACGT base identity.
[2] Positions with reference base ACGT that have at least 10 reads aligned; remaining statistics refer to this set of filtered bases.
[3] % of positions for which no non-reference bases were detected (VarFreq=0)
[4] Number of positions with evidence of strand-specific sequencing errors, as determined by Fisher's exact test for strand bias between reference and alternate alleles with p-value <0.05.
[5] Among positions with no strand bias, these positions represent single-nucleotide variants between strains and reference sequences. All variants can be accounted for as due to the introduction of the antibiotic resistance alleles or due to errors in the sequence reference (in the case of Rd KW20).

Table S2.5. Transformation frequencies and estimated competence.

Recip	Rep	CFU/ml	Nov^R/CFU	Nal^R/CFU	Nov^R Nal^R /CFU	%Competence*
HiT	A	3.1E+09	4.1E-03	4.1E-03	3.3E-06	19.1%
HiT	B	3.1E+09	7.2E-03	5.7E-03	5.4E-06	13.1%
HiT	C	3.8E+09	5.3E-03	3.6E-03	3.1E-06	15.8%
HiT	Mean	3.3E+09	5.5E-03	4.5E-03	3.9E-06	16.0%
HiT	SD	4.2E+08	1.5E-03	1.1E-03	1.3E-06	3.0%
RdS	A	1.0E+09	2.3E-03	2.5E-03	8.6E-07	14.7%
RdS	B	2.8E+09	1.1E-03	2.7E-03	4.4E-07	15.1%
RdS	Mean	1.9E+09	1.7E-03	2.6E-03	6.5E-07	14.9%
RdS	SD	1.3E+09	8.9E-04	1.2E-04	3.0E-07	0.3%

* % Competence was calculated as expected doubles (Nov^RNal^R/CFU) divided by the product of the singles (Nov^R/CFU * Nal^R/CFU).

Table S2.6. Allele frequencies around antibiotic resistances in Pool 0*.

Recip	Select	Position	Recip	Donor	Depth	%recip	%donor	%other
HiT	Nal^R	1,274,373	A	C	388	5.93%	94.07%	0.00%
HiT	Nal^R	1,274,406	C	T	424	4.95%	94.81%	0.24%
HiT	Nal^R	1,274,496	T	G	546	2.20%	97.80%	0.00%
HiT	Nal^R	1,274,522	C	A	541	0.00%	99.82%	0.18%
HiT	Nal^R	1,274,532	T	G	473	3.17%	96.83%	0.00%
HiT	Nal^R	1,274,634	C	T	354	2.82%	96.61%	0.56%
HiT	Nal^R	1,274,652	T	A	342	2.92%	96.78%	0.29%
HiT	Nov^R	457,257	A	G	532	5.08%	94.92%	0.00%
HiT	Nov^R	457,263	A	T	511	6.65%	93.35%	0.00%
HiT	Nov^R	457,333	G	T	523	4.59%	95.41%	0.00%
HiT	Nov^R	457,600	C	A	434	0.00%	100.00%	0.00%
HiT	Nov^R	457,953	T	C	382	13.87%	86.13%	0.00%
HiT	Nov^R	457,956	A	C	367	13.08%	86.92%	0.00%
HiT	Nov^R	457,962	G	A	402	12.69%	87.31%	0.00%
RdS	Nal^R	1,343,717	C	G	893	28.00%	71.89%	0.11%
RdS	Nal^R	1,343,720	T	C	891	26.82%	72.73%	0.45%
RdS	Nal^R	1,343,759	T	G	952	19.75%	80.25%	0.00%
RdS	Nal^R	1,344,100	C	A	993	18.83%	81.17%**	0.00%
RdS	Nal^R	1,344,472	C	T	865	38.61%	61.39%	0.00%
RdS	Nal^R	1,344,488	A	G	703	32.43%	67.57%	0.00%
RdS	Nal^R	1,344,490	T	C	690	31.74%	66.96%	1.30%
RdS	Nov^R	586,842	A	G	635	17.17%	82.83%	0.00%
RdS	Nov^R	586,851	C	A	622	16.24%	83.76%	0.00%
RdS	Nov^R	586,854	T	A	620	16.13%	83.87%	0.00%
RdS	Nov^R	587,579	G	T	972	0.31%	99.69%	0.00%
RdS	Nov^R	587,846	C	A	1134	3.09%	96.74%	0.18%
RdS	Nov^R	587,969	C	A	983	4.27%	95.73%	0.00%
RdS	Nov^R	588,549	C	T	813	16.24%	83.64%	0.12%

* Highlighted rows indicate positions of antibiotic resistance causing donor alleles.
** The RdS Nal^R recombinant shows only ~80% Nal^R allele due to a mistake during pooling. An additional ~20% of the pool was Nov^R.

Table S2.7. Non-reference alleles at reliable SNP positions in transformed pools and untransformed controls*

DNA	Type	Total SNPs	Donor Freq>0	Novel Freq> 0	Depth (Median ± MAD)	Un-mapped to Recip	Re-mapped to donor	% Re-mapped
MAP7	Control	36,869	9,839	10,273	876 ± 268	20,955	656	3.13%
RdS	Control	36,869	406	731	52 ± 13	935	22	2.35%
RdS Nal^R	Pool 0	36,869	30,998	6,911	787 ± 142	19,861	3,054	15.38%
RdS Nov^R	Pool 0	36,869	29,838	5,405	636 ± 125	18,890	2,969	15.72%
Hi375	Control	19,520	4,027	7,762	701 ± 169	13,992	43	0.31%
HiT	Control	19,520	786	1,202	236 ± 43	3,891	8	0.21%
HiT Nal^R	Pool 0	19,520	14,882	2,119	448 ± 117	9,125	169	1.85%
HiT Nov^R	Pool 0	19,520	10,840	1,534	325 ± 70	8,477	577	6.81%

* The distribution of non-zero allele frequencies is depicted in S4 Figure part C beanplots.

Table S2.8. Clone genotype assignments.

Background	Resistance	Clone	Genotype
RdS	Nal	s1	A
RdS	Nal	s2	B (P540)
RdS	Nal	s3	A
RdS	Nal	s4	A
RdS	Nov	s1	C
RdS	Nov	s2	C
RdS	Nov	s3	C
RdS	Nov	s4	C
HiT	Nal	s1	D
HiT	Nal	s2	D
HiT	Nal	s3	D
HiT	Nal	s4	D
HiT	Nov	s1	E (P551)
HiT	Nov	s2	F
HiT	Nov	s3	E
HiT	Nov	s4	E

Table S2.9. Donor segments detected in each isolated genotype.

Reference	Genotype	Type	Begin	End
Rd_KW20	A	hitchhiker	1,056,042	1,065,007
Rd_KW20	A	Nal^R	1,338,588	1,349,925
Rd_KW20	A	invasion	1,734,959	1,766,413
Rd_KW20	B	hitchhiker	862,451	877,939
Rd_KW20	B	hitchhiker	887,142	893,453
Rd_KW20	B	Nal^R	1,334,937	1,344,101
Rd_KW20	B	invasion	1,732,805	1,750,336
Rd_KW20	C	Nov^R	585,269	588,958
Rd_KW20	C	hitchhiker	1,494,874	1,495,778
Rd_KW20	C	invasion	1,744,519	1,750,957
Rd_KW20	C	hitchhiker	1,760,431	1,760,794
Rd_KW20	**ABC**	**min. invasion**	**1,744,519**	**1,750,336**
Hi375	D	hitchhiker	1,139,838	1,143,114
Hi375	D	invasion	1,172,910	1,187,650
Hi375	D	Nal^R	1,274,284	1,274,653
Hi375	E	hitchhiker	374,543	375,508
Hi375	E	Nov^R	454,718	460,941
Hi375	E	invasion	1,178,771	1,183,728
Hi375	E	hitchhiker	1,478,173	1,482,010
Hi375	F	Nov^R	448,886	460,427
Hi375	F	invasion	1,172,475	1,185,610
Hi375	**DEF**	**min. invasion**	**1,178,771**	**1,183,728**
Hi375	D-EF	D-exclusive	1,139,838	1,142,951
Hi375	D-EF	D-exclusive	1,185,743	1,187,650
Hi375	D-EF	D-exclusive (Nov^R)	1,274,284	1,274,653

Genotypes correspond to S8 Table. Coordinates are inclusive and with reference to the recipient. Only SNP positions that were deemed reliable (see SUPPLEMENTARY RESULTS) were used to define recombination breakpoints. The spurious interval that contains the *radA*-adjacent $hmw1_{Hi375}$ interval is not reported. Donor segment type can be: "**Nov^R/Nal^R**" for the segment containing the selected donor antibiotic resistance, "**invasion**" for the segment containing the putative invasion allele/locus, or "**hitchhiker**" for segments that are neither of these. "**Min. invasion**" defines the intersection of putative invasion tracts within the same recipient background. Though "hitchhiking" segments are unlikely to be due to the selections, subtle effects on intracellular invasion by these intervals are not ruled out.

In particular, donor variation exclusive to genotype D ("**D-exclusive**") may carry alleles/loci involved in intracellular invasion (Figure 7). Genes affected by recombination in these intervals (aside from the *gyrA* allele conferring Nal^R) are NF38_05675 (mannose-6-phosphate isomerase pseudogene), NF38_05680 (hypothetical), NF38_05685 (*qseB* two-component system transcription factor), NF38_05690 (*qseC* two-component system sensor), NF38_09010 (RNase P), NF38_05855 (anaerobic NO production, iron-sulfur cluster repair), and NF38_05860 (*moaA*).

Table S2.10. Primers used in this study.

Purpose	Dir	Primer name	Primer number	Sequence (5´-3´)
RdS	F	*rpsL*-RdKW20-F	1497	GATTACGTGTAGATCGCCTTAAAACAGGTA
RdS	R	*rpsL*-RdKW20-R	1798	CCAAAATAATTTCATTTGCTAATACACGCT
Interval	F	hmw1A-7-D	13CAN	CACTCGCTAGTGGTTGTTAATGATGAA
Interval	R	hmw1A-recomb-strat2	1273	TATCCCCCAACTTTAATCGCGTCT
Deletion	F	Hmw1A-casset-F	1199	GCAAATGATAAAGTAATTTAATTGTTCAACTAACCTTAGGAGAAAATA TGATTCCGGGGATCCGTCGACC
Deletion	R	Hmw1A-casset-R-strat2	1274	TCTCTCCCCCTTCTTTACCTGAAAGGTCGATAACTGCGCCTGTTTTT ATGTAGGCTGGAGCTGCTTCG
FlankA	F	radA-hmw1A-375-FragA-Fw	1463	ATTGTTTTGAGCCCAGCAAATAACCTC
FlankA	R	radA-hmw1A-375-FragA-SmaI-Rv	1464	TCCCCCGGGTTACACCTAAAGATAGTAACATAG
FlankB	F	radA-hmw1A-375-FragB-SmaI-Fw	1465	TCCCCCGGGTATCACAAATTTCACTTTTAATGTAG
FlankB	R	radA-hmw1A-375-FragB-Rv	1466	GATGTTGCAGTTAGGGTCGCGGCTCC
FlankA	F	yrbI-hmw2A-375-FragA-Fw	1467	AGGCATAAAAATGCTGATGGATGCGGGTATT
FlankA	R	yrbI-hmw2A-375-FragA-SmaI-Rv	1468	TCCCCCGGGCCATGCCCAAGGATAGCAATATA
FlankB	F	yrbI-hmw2A-375-FragB-SmaI-Fw	1469	TCCCCCGGGGTAAAACTTGGCAATGTTACCAAT
FlankB	R	yrbI-hmw2A-375-FragB-Rv	1470	CTACGGCTGTCCATCGTCAGCAACATTGGTAC
Cloning	F	NTHI1981upstr-F1	1219	CAATTGCACGCCACCTGCAGATGCTTG
Cloning	R	NTHI1982-R1-HA	1218	TTAAGCGTAGTCTGGGACGTCGTATGGGTATTGCCCCATATTTTTCA CTGATTTTAG
Allele ID	F	HMW1A-2A-NTHi375-Fw	1456	GGAATGGATGTAGTACACGGCACA
Allele ID	R	HMW1A-NTHi375-Rv	1458	CCGCTCCGACCCTCACTCCAAAGA
Allele ID	R	HMW2A-375-Rv	1457	GCATTGCTGTCAATGGATAAATAA
Allele ID	F	HMW1A-86028NP-Fw	1459	AATAACTACAAAACTCCAGGGGTG
Allele ID	R	HMW1A-86028NP-Rv	1460	GGTAATATTGACTTTATCAGAAGA
Allele ID	F	HMW2A-86028NP-Fw	1461	ATTACGCTTGGTACGGGTTTTTTA
Allele ID	R	HMW2A-86028NP-Rv	1462	ATCACTGCTACCGGTAGCTGTAAT

Chapter 3. Antagonistic pleiotropy in a bifunctional fatty acid transporter during bacterial adaptation to chronic lung infection.

Moleres J, Fernández-Calvet A, Ehrlich RL, Martí S, Pérez-Regidor L, Euba B, Rodríguez-Arce I, Balashov S, Santos S, Liñares J, Ardanuy D, Martín-Santamaría S, Ehrlich GD, Mell JC, Garmendia J

Antagonistic pleiotropy in a bifunctional fatty acid transporter during bacterial adaptation to chronic lung infection

Javier Moleres[1+], Ariadna Fernández-Calvet[1+], Rachel L. Ehrlich[2], Sara Martí[3,4], Lucía Pérez-Regidor[5], Begoña Euba[1,3], Irene Rodríguez-Arce[1], Sergey Balashov[2], Salud Santos[3,4], Josefina Liñares[3,4], Carmen Ardanuy[3,4], Sonsoles Martín-Santamaría[5], Garth D. Ehrlich[2], Joshua C. Mell[2]*, Junkal Garmendia[1,3]*

[1]Instituto de Agrobiotecnología, CSIC-Universidad Pública Navarra-Gobierno Navarra, Mutilva, Spain; [2]Department of Microbiology and Immunology, Institute for Molecular Medicine and Infectious Diseases, Center for Genomic Sciences, Drexel University College of Medicine, Philadelphia, PA, USA; [3]Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Madrid, Spain; [4]Departamento Microbiología, Hospital Universitari Bellvitge, University of Barcelona, IDIBELL, Barcelona, Spain; [5]Department of Structural and Chemical Biology, Centro de Investigaciones Biológicas, CIB-CSIC, Madrid, Spain

[+]These authors contributed equally to this work

*Corresponding authors:

Junkal Garmendia, Instituto de Agrobiotecnología, CSIC-Universidad Pública de Navarra-Gobierno de Navarra, 31192 Mutilva, Navarra, Spain; Tel: 00-34-948168484; Fax: 00-34-948232191; e-mail: juncal.garmendia@unavarra.es;

Joshua C. Mell, Department of Microbiology and Immunology, Institute for Molecular Medicine and Infectious Diseases, Center for Genomic Sciences, Drexel University College of Medicine, Philadelphia, PA, USA; Tel: 1-215-762-1445; Fax: 1-215-762-1003; e-mail: jcm385@drexel.edu

Running title: Bacterial patho-adaptation during long-term lung colonization

Key words: *Haemophilus influenzae*, chronic obstructive pulmonary disease, genomics, adaptive evolution, convergent evolution, FadL, CEACAM-1, free fatty acids, antagonistic pleiotropy

Acknowledgments. We thank Lucía Caballero and Sergio Cuesta for their technical support. J.M. is funded by PhD studentship BES-2013-062644 from MINECO, Spain; A.F.C. was funded by a contract from MINECO, reference 20132RC947, Spain; I.R.A. is funded by a PhD studentship from Universidad Pública de Navarra, Spain; S.M. is funded by a postdoctoral contract from CIBER Enfermedades Respiratorias (CIBERES).

Funding. This work has been funded by grants from MINECO (SAF2012-31166 and SAF2015-66520-R to JG; CTQ2014-57141-R and CTQ2017-88353-R to SMS), Health Department, Regional Govern from Navarra, Spain, reference 03/2016, and SEPAR 31/2015 to J.G. CIBERES is an initiative from Instituto de Salud Carlos III (ISCIII), Madrid, Spain.

3.1. Abstract

Tracking bacterial evolution during chronic infection provides insights into how host selection pressures shape bacterial genomes. The human-restricted opportunistic pathogen nontypeable *Haemophilus influenzae* (NTHi) persists within the lower airways of chronic obstructive pulmonary disease (COPD) patients and contributes to disease progression. To identify bacterial genetic variation associated with NTHi adaptation to the COPD lung, we sequenced the genomes of 94 isolates collected from sputum samples of 13 COPD patients between 2000 and 2014. Individuals were colonized by multiple highly diverse clonal types (CT) over time, but the same CT often re-appeared and was sometimes found in different patients. Although genomes from the same CT were nearly identical, intra-CT variation due to mutation and recombination occurred. Recurrent mutations were seen in several genes likely involved in COPD lung adaptation. Prominently, nearly a third of CTs were polymorphic for null alleles of the *ompP1/fadL* gene, which encodes a bifunctional membrane protein that both binds the human hCEACAM1 receptor and imports long-chain fatty acids (LCFAs). Structural predictions provided plausible 3D models for FadL's interaction with hCEACAM1 and LFCAs. Recurrent *fadL* mutations reveal a likely case of antagonistic pleiotropy, where loss of FadL reduces NTHi's ability to infect epithelia, but also increases its resistance to bactericidal LFCAs enriched within the COPD lung. Supporting this interpretation, truncated *fadL* alleles are rare in publically available *H. influenzae* genomes, but enriched in lower airway isolates. Our results shed light on molecular mechanisms of bacterial patho-adaptation and guide future research towards developing novel COPD therapeutics.

3.2. Introduction

Bacterial pathogens evolve within long-term chronic infections, and identifying the genetic changes that arise in bacteria over time within hosts is a powerful means to decipher host selective pressures acting on bacteria that colonize new niches, as well as the specific adaptations that allow bacteria to persist (Didelot et al., 2016). Dissecting the genetics of bacterial adaptation is key to understanding and predicting the emergence of antimicrobial resistance and immune evasion traits within bacterial populations, as well as to identifying new bacterial drug targets.

Colonization of the human lower airways by opportunistic pathogens of environmental origin leads to irreversible progression of major chronic respiratory diseases such as cystic fibrosis (CF), and bacterial evolution during long-term lung infection has been extensively analyzed in this context (Didelot et al., 2016; Folkesson et al., 2012). In particular, genomic analyses of serially collected isolates of *Pseudomonas aeruginosa* and *Burkholderia* species have provided *in vivo* evidence of adaptation during long-term infection of CF lungs (Lee et al., 2017; Lieberman et al., 2011; Lieberman et al., 2014; Marvig et al., 2014; Marvig et al., 2015; Price et al., 2013; Smith et al., 2006; Winstanley et al., 2016; Yang et al., 2011). Differently, understanding how members of the human microbiome adapt to new niches on the human body and become pathogens remains scarce, limited to *Staphylococcus aureus* evolutionary dynamics during CF lung chronic infection and progression from carriage to invasive disease (Laabei et al., 2015; Lopez-Collazo et al., 2015; McAdam et al., 2011; Young et al., 2012).

Chronic obstructive pulmonary disease (COPD), the third leading cause of death globally, whose primary risk factor is smoking, is an irreversible airflow obstruction accompanied by emphysema, fibrosis, neutrophil airway infiltration, mucus hypersecretion, inflammation and, importantly, long-term lower airway colonization by opportunistic pathogens (Ahearn et al., 2017). Though nontypeable *Haemophilus influenzae* (NTHi) is typically a benign commensal of the human upper airways, it is also a common opportunistic pathogen isolated from the lower airways of COPD patients during both exacerbation and clinically stable periods, responsible for 20-30% of all acute COPD exacerbations (Bandi et al., 2001; Finney et al., 2014; King & Sharma, 2015; Sethi et al., 2002; Sethi & Murphy, 2008). NTHi persists within the COPD lung for months to years, contributing to chronic airway inflammation that results in worsening of symptoms and speeds disease progression (Anzueto, 2010; Desai et al., 2014; Finney et al., 2014; Murphy et al., 2004).

NTHi is extremely genomic and phenotypically diverse, and individuals are often colonized by multiple distinct strains, distinguished by hundreds of gene possession differences and tens of thousands of single-nucleotide variants (SNVs) (Chin et al., 2005; Cholon et al., 2008; De Chiara et al., 2014; Fernaays et al., 2006; Gilsdorf et al., 2004; Hogg et al., 2007; LaCross et al., 2013; Look et al., 2006; Murphy et al., 2004; Nakamura et al., 2011; Sethi et al., 2004; Zhang et al., 2012). In addition, NTHi are naturally competent (able to take up and recombine DNA from relatives into their chromosome), and they possess numerous phase variable genes (Gilsdorf et al., 2004; Lacross et al., 2008; Maughan & Redfield, 2009; Mell & Redfield, 2014; Power et al., 2009), traits that likely contribute to NTHi's ability to adapt and persist within the COPD lung.

To identify molecular genetic changes underlying NTHi adaptation to the COPD airway, we sampled, sequenced and compared the whole genomes of 94 sputum sample isolates longitudinally collected from 13 COPD patients at serial clinic visits over 1 to 9 years, mostly during pulmonary exacerbations. Genetic variation within clonally related isolates identified recurrent polymorphisms in several genes. Many clonal types (CT) isolated from many patients were polymorphic for null alleles of the *ompP1/fadL* gene, which we show decreases the ability of NTHi to invade human cells but also increases its resistance to the bactericidal effect of free long-chain fatty acids (LCFAs), in particular to arachidonic acid, a key inflammatory mediator in COPD. This is the first report of within-host patho-adaptive evolution by a host-restricted member of the human airway microbiome behaving as a persistent pathogen.

3.3. Materials and Methods

COPD strain isolation and PFGE typing. NTHi isolates were collected at the Hospital Universitari de Bellvitge (HUB), a tertiary-care hospital in Barcelona, Spain. NTHi were isolated from sputum samples of 13 COPD patients over a 14-year period (2000-2014) at regular check-ups or during visits requiring hospitalization due to an exacerbation of the disease (**Table S3.1**). Informed consent was not required as sputum sampling and microbial isolation process is part of the standard microbiological routine. Patient confidentiality was protected in all cases. *H. influenzae* was identified by conventional culture methods (Murray et al., 2003) and by mass spectrometry with a matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) Biotyper, version 3.0 (Bruker). NTHi isolates were PFGE genotyped as previously discussed, and considering isolates

with ≥80% relatedness were considered to belong to the same PFGE type (Puig et al., 2013).

Bacterial growth conditions and genetic manipulations. *H. influenzae* COPD isolates (**Table S3.2**), and reference strains RdKW20 and NTHi375, were grown at 37ºC with 5% CO_2 on chocolate agar (Biomérieux) or sBHI, made from brain-heart infusion supplemented with 10 μg/ml hemin and 10 μg/ml nicotinamide adenine dinucleotide (NAD), with or without agar for solid and liquid culture, respectively. Spectinomycin 30 μg/ml (Spc_{30}) was used when needed. *Escherichia coli* was grown on Luria Bertani (LB) broth or LB agar plates at 30ºC or 37ºC, supplemented with ampicillin 100 μg/ml (Amp_{100}) or spectinomycin 50 μg/ml (Spc_{50}). The *fadL* gene was knocked out by replacement of the coding sequence with a Spc^r cassette, following (Sinha et al., 2012; Tracy et al., 2008) (**Supplementary Methods**, **Table S3.6**).

Genome sequencing, assembly, annotation, data deposition. DNA was extracted from each isolate and sequenced using standard methods. Briefly, all 94 genomic DNAs were subjected to Illumina paired-end short-read sequencing, and a subset of 19 DNAs were also subjected to Pacific Biosciences RSII long-read sequencing. Genome assembly was platform specific (**Supplementary Methods**). Assemblies taxonomically classified with taxator v1.2 v1.2 (Droge et al., 2015). For MLST strain identification, the *adk*, *atpG*, *frdB*, *fucK*, *mdh*, *pgi*, and *recA* genes were extracted from assemblies and assigned STs using https://github.com/tseemann/mlst. New alleles and ST were submitted as needed (**Table S3.3**, https://pubmlst.org/hinfluenzae). To compare PacBio and Illumina assemblies of the same isolates' genomes, progressiveMauve version snapshot_2015-02-25 (Darling et al., 2010) was used to reorder Illumina contigs against the complete PacBio assembly, finding perfect agreement in all cases. To annotate strains with consistent gene names, a database of annotated *Pasteurellaceae* genomes from NCBI was created by https://github.com/rehrlich/prokka_database_maker. Assemblies were annotated using this database and Prokka (v 1.11) (Seemann, 2014).

All raw reads and assemblies were deposited at NCBI under BioProject PRJ282520; BioSample accessions for each strain are included in **Table S3.3**. Publically available genome assemblies at NCBI and Sanger Centre were downloaded (**Table S3.9**), and re-annotated with Prokka to maintain consistency. Assemblies included in the analyses were required to pass several quality control filters, including assembly quality and taxonomic identity (**Table S3.9**). The combined dataset of 186 publicly available strains

and 92 strains from this study was used for gene clustering, phylogenetic analysis and clonal typing.

Gene clustering, phylogenetic analysis, and clonal typing. Homologous protein-coding genes among all genomes were clustered together using Roary at a BLASTP threshold of 75% (Page et al., 2015), (**Supplemantary methods**). Phylogenetic tree reconstruction used the hybrid MPI/Pthreads version of RAxML (8.2.4, 10 processes, 16 cpus per process, -f a -x 1234 -m GTRGAMMA -p 7 -N autoMRE -j --set-thread-affinity) on a concatenation of all codon-aware alignments of protein-coding genes with at most one copy per strain (Stamatakis, 2014). PFGE and MLST are low-resolution strain typing methods, so we applied goeBurst via Phyloviz 2.0 (Francisco et al., 2009) to cluster isolates into clonal types (CTs), based on the allelic identity across 478 core protein-coding genes shared by all *H. influenzae* isolates studied here, as well as several *H. haemolyticus* and *H. parainfluenzae* strains.

Detecting intraclonal genetic variation in protein-coding genes. Because methods to call genetic variants by short-read alignment to reference assemblies are relatively mature, compared to methods that use whole-genome alignment, we chose a reference sequence from each CT (based on assembly quality, a Pacific Biosciences assembly when available) and then aligned short read pairs from all isolates belonging to that CT to this reference (bwa mem). Single-nucleotide and short insertion/deletion variants (SNVs and indels) were extracted using freeBayes, and the impact of coding variants was evaluated using SnpEff v4.3k (Cingolani et al., 2012). To identify homolgs across CTs, the Roary gene possession table (data not shown) was used to “liftover” gene identities between the CT reference genomes. Artifacts of gene annotation failure (for example due to frameshift and nonsense mutations), particularly for the CT references, were reconciled by considering intraclonal variation in gene possession as a “high-impact” variant.

FadL allelic identities for whole collection To extract every *fadL* gene sequence from the whole NTHi isolate collection, a local database encompassing 92 whole genome sequences was created and local blast tool was employed (Camacho et al., 2009). Results as a mean of coordinates within each genome were annotated and used to get *fadL* sequences using BEDTools getfasta (Quinlan & Hall, 2010). Nucleotide sequences were then translated to protein sequences.

Antimicrobial susceptibility testing. Susceptibility to ampicillin (AMP), amoxicillin-clavulanic acid (AMC), cefaclor (CEC), cefonicid (CID), ciprofloxacin (CIP),

chloramphenicol (CLO), cotrimoxazole (SXT), cefotaxime (CTX), tetracycline (TET) and cefuroxime (CFU) was determined by disc diffusion assay according to the criteria of the Clinical Laboratory Standards Institute (CLSI) (https://clsi.org/standards/products/microbiology/documents/m100/). For the cationic peptide polymyxin E (PxE), bacterial susceptibility was determined by e-test (Biomérieux). Strains grown on chocolate-agar were used to generate a saline solution bacterial suspensions normalized to OD_{600}=0.063 (~0.5 McFarland). A cotton swab was soaked in this suspension and used to generate a bacterial loan on Mueller Hinton Fastidious (MH-F)-agar. When dry, a PxE e-test strip was located in the middle of the plate and incubated at 37ºC for 16 h, to further determine the minimal inhibitory concentration (MIC).

Sanger sequencing of the *hmw1A* gene and predicted promoter region. Due to intrinsic technical difficulties in assembling repetitive DNA sequences through short reads generated by Illumina sequencing (Treangen & Salzberg, 2011), we lack *hmw* gene comlplete sequences in many Illumina-squenced NTHi genomes. The *hmw1A* gene and its respective promoter region in NTHi strains P651-8849, P653-8956 and P654-8983 was Sanger sequenced. Primers were design based on their clonal PacBio sequenced strain P652-8881. Primers hmw1A-Seq-F-1656, hmw2A-Seq-F-1655 and hmw12A-Seq-R-1657 were used to sequence 600 bp regions upstream of the *hmw1A* and *hmw2A* genes, respectively. Primers listed in **Table S3.6** were used to sequence the entire *hmw1A* gene coding sequence.

Gene expression analysis. Reverse transcriptase real-time quantitative PCR used total RNA extractions from late-log phase bacterial cultures, using primers against *fadL* and against *gyrA* as a control (**Table S3.6, Supplementary methods**).

Electrophoresis, Coomasie-staining and western blot. FadL presence in strains P589-8275, P590-8360, P595-8370, P631-8237, P662-7189, P614-8522, P622-7806, P607-8844 and P608-8895 was assessed by coomasie-staining of whole bacterial cell extracts separated in a 12% SDS-PAGE. HMWA protein levels in strains P651-8849, P652-8881, P653-8956 and P654-8983, were monitored by western blot. In all cases, whole cell extracts were prepared from bacterial suspensions recovered from chocolate agar plates, adjusted to OD_{600}=1 in PBS, lysed by ultrasounds sonication, two-fold diluted with 2X loading buffer (Tris-HCl 62.5mM pH 6.8, SDS 2% w/v, glycerol 10%, DTT 50 mM; Bromophenol Blue 0.01% w/v), and heated to 95ºC for 5 min. For protein monitoring by Coomasie brilliant blue staining, electrophoresis on 12% SDS-PAGE

20x20 cm gels designed for long-range protein separation with maximum resolution under reducing conditions (BioRad, PROTEAN II xi Cell) was performed at 15 mA until samples reached resolution phase (~2,5 h) and then at 30 mA for 16 h. FadL was identified in strain P589-8275 by peptide mass fingerprinting (MALDI-TOF/TOF) at the Centro Nacional de Biotecnología (CNB)-Spain proteomics facility. For HMWA antibody-detection, electrophoresis on 10% SDS-PAGE 10x7.5 cm (Bio-Rad 1653308) was performed. Proteins were then transferred to a nitrocellulose membrane. HmwA protein was monitored with a primary guinea pig anti-HMW1A (GP85) antibody diluted 1:2000 (Buscher et al., 2006) and a secondary goat anti-guinea pig IgG (Santa Cruz) antibody conjugated to horseradish peroxidase, diluted 1:5000.

Cell culture and bacterial infection invasion. HeLa and HeLa-BGP cells (Gray-Owen et al., 1997) were seeded ($4x10^5$ cells/well 24 h before infection). A549 human alveolar basal epithelial cells (ATCC CCL-185) were maintained and seeded as described (Morey et al., 2011), to $6x10^4$ cells/well for 32 h, and serum-starved 16 h before infection. Invasion assays performed as described (Euba et al., 2015a; Euba et al., 2015b; Euba et al., 2015c; Lopez-Gomez et al., 2012; Morey et al., 2011) (**Supplementary Methods**).

Free fatty acid susceptibility testing. Bacterial suspensions were incubated with varying concentrations of free fatty acids or vehicle control (ethanol), serially diluted and plated on sBHI agar. Results are expressed as percentage of bacterial survival ([c.f.u.ml$^{-1}_{LCFA}$/c.f.u.ml$^{-1}_{vehicle}$] x 100) (**Supplementary Methods**).

NTHi mouse lung infection. A CD1 mouse model of NTHi pulmonary infection was used (Euba et al., 2015a; Euba et al., 2015b; Euba et al., 2015c; Morey et al., 2013). At least five mice per treatment (genotype x time point) were intranasally inoculated with ~$2×10^8$ c.f.u./mouse, mice were euthanized at 24 or 48 hpi. Total c.f.u. in lung and BALF samples/mouse were determined (**Supplementary Methods**).

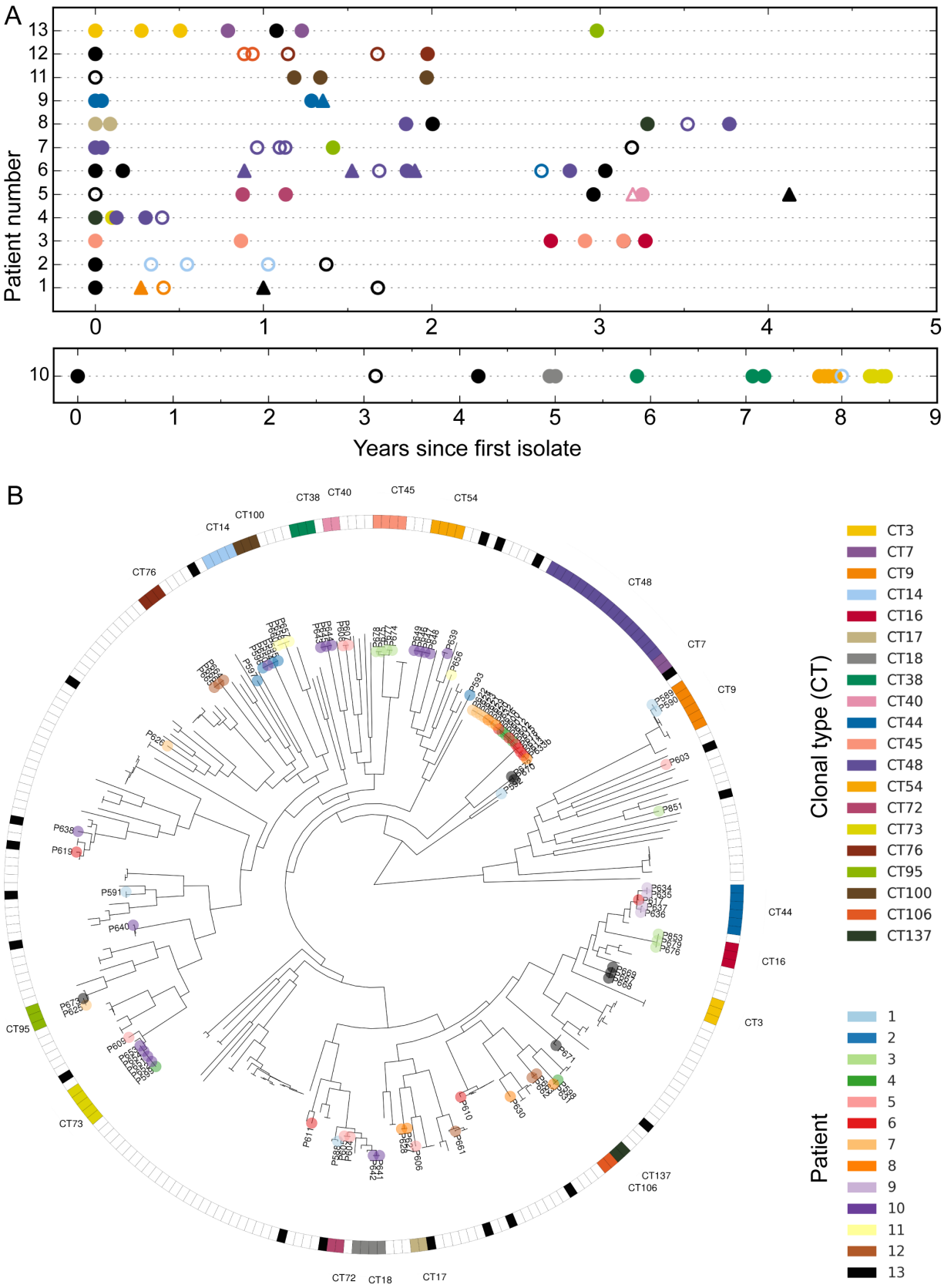
Computational modeling for FadL. See **Supplementary Methods**.

3.4. Results

3.4.1. A collection of NTHi genomes isolated from COPD patients over time

To identify natural genetic variation in bacterial genomes associated with adaptation to the COPD lung, we serially collected NTHi isolates from a set of COPD patients over time, sequenced, and analyzed their genomes. Sputum samples were collected from COPD patients during exacerbations requiring hospitalization or at primary health care routine visits, and *H. influenzae* colonies were identified and isolated. Thirteen patient series were selected on the basis of having four or more longitudinally sampled NTHi isolates, as well as having more than one independent isolate of the same pulsed-field gel electrophoresis (PFGE) type, suggestive of persistent long-term infection by the same strain. Sampled subjects were Spanish, male, born between 1912 and 1963, smokers or ex-smokers. Clinical information on each subject is summarized in **Table S3.1**. The bacterial collection consists of 94 isolates, with 4 to 18 isolates per subject, collected over a period ranging from 1 to 9 years (**Figure 3.1A**). Details on each isolate are summarized in **Tables S3.2** and **S3.3**.

Figure 3.1. (a) Overview of longitudinal sampling of NTHi from COPD patients, consisting of 92 isolates collected over 1 to 9 years from 13 COPD patients (4 to 18 isolates per patient) (**Tables S1** to **S3**). Axes indicate the patient from whom the isolate was derived and the relative time of sampling for each isolate. Symbol indicates whether isolate was collected outside an acute exacerbation (triangles) or not. Color indicates clonal type (CT), as determined by goeBurst analysis of 478 single-copy core protein-coding genes (minimum spanning tree shown in **Figure S3.3**). Black is reserved for CTs containing only a single COPD isolate from this collection. Filled symbols indicate the presence of a full-length *fadL* gene; open symbols indicate presence of a truncated *fadL* gene. (**b**) Species-level phylogenetic tree reconstruction using RAxML, based on a concatenation of codon-aware alignments of 478 single-copy protein-coding homologous gene clusters. Patient ID is indicated by node color, whereas CT is indicated by the color in the outer ring, as in **Figure 3.1A**. Only COPD isolates sequenced for this study are labeled; **Figures S3.4** and **S3.5** show an unrooted phylogeny including *H. haemolyticus* and *H. parainfluenzae* outgroups and a fully labeled rectangular tree that includes bootstrap results.



The genomes of all 94 isolates were subjected to Illumina shotgun sequencing, and a subset of 19 were also sequenced using the Pacific Biosciences RSII (PacBio) to act as complete reference sequences. Genome assembly, gene annotation, taxonomic classification, and *in silico* multilocus sequencing typing (MLST) were performed. Results are summarized in **Table S3.7**; results of DNA modification motif analysis from a subset of PacBio assemblies are shown in **Table S3.8**. Poor quality genomes were removed from subsequent analysis, which used 17 PacBio. Assemblies from the two sequencing methods showed high agreement overall (**Figure S3.1, Methods**).

3.4.2. Gene clustering and phylogenomic analysis

To provide context to these new COPD genomes for comparative genomic analyses, we combined our genome collection with 186 publicly available *H. influenzae* assemblies, of which only 27 were from COPD or other lower airway infections (**Table S3.9**). Pan-genome analysis was conducted by clustering together homologous protein-coding genes using the pan-genome analysis pipeline Roary (**Methods** and (Lee et al., 2017)), resulting in 4,837 total homologous gene clusters, of which 1,386 were present in >=95% of *H. influenzae* assemblies, and 1,844 were only present in <15% of assemblies (**Figure S3.1B**).

For molecular typing and clustering of closely-related isolates into clonal types (CTs), we extended beyond PFGE and MLST results by applying the goeBurst algorithm in Phyloviz to allele assignments made from 478 core protein-coding genes (**Methods, Figure S3.2**). Overall, 146 clonal types were detected across all genomes, 141 in *H. influenzae* genomes, 40 in our COPD collection, 20 CTs contained >1 isolate from our COPD collection. The 72 COPD isolates belonging to these 20 CTs form the basis for the analysis of recurrent genomic changes described below. Consistent with results from clonal typing, the species-level phylogenetic tree built from alignments of all protein-coding homologs with at most one copy per strain illustrated several findings (**Figure 3.1B,** unrooted tree with outgroups, **Figure S3.4;** fully labeled rectangular version, **Figure S3.5**):

(1) Individual patients were colonized by numerous genetically diverse strains over time, though because only one isolate was sequenced at each time point, polyclonal diversity at any given time point remains unclear (Berrens et al., 2007; Farjo et al., 2004; Hiltke et al., 2003; Lacross et al., 2008; Murphy et al., 1999; Sethi et al., 2002). High polyclonality and high average degree of divergence among CTs contrasts with previous studies of

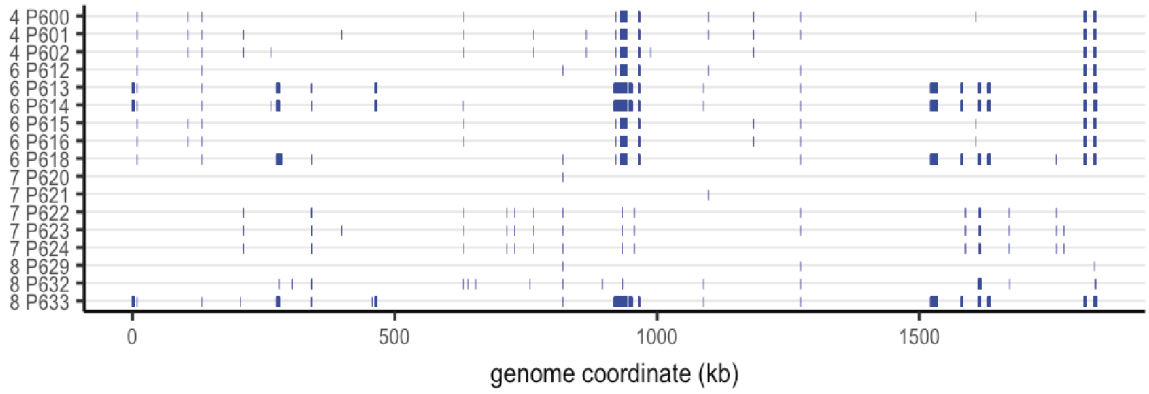
environmentally acquired opportunistic pathogens, which have typically supported colonization by single CTs followed by subsequent diversification (Lee et al., 2017; Lieberman et al., 2014).

(2) Groups of closely related isolates were collected both within and among COPD patient series, and they sometimes included isolates from public databases with diverse clinical and geographical origin (**Figure S3.2**).

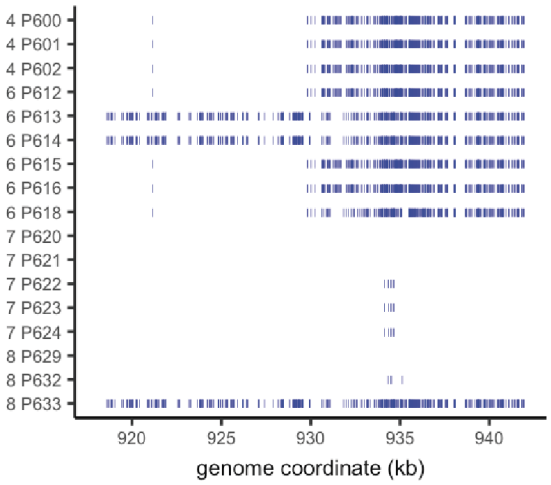
(3) Several instances of probable inter-patient transmission were observed. Most notably, for CT48, seventeen isolates were found across four patients that all live within a 400 meter radius. Isolation dates, along with SNVs among this subset of strains, suggest a probable transmission model (**Figure 3.2D**).

Figure 3.2. Genetic variation within CT48. (**a**) Whole-genome view of all short variants (SNVs and short indels) arising within the 17 isolates in CT48, which were found in 4 patients. Variants were detected by alignment of Illumina short read pairs to the PacBio reference sequence for strain P621-7028, and fixed differences from this reference were excluded. Blocks are due to numerous closely-spaced variants, which resemble natural transformation events seen in the laboratory. (**b**) Zoomed-in view of recombination tracts near center of the genome, showing that recombination tracts in this region likely had multiple independent donor sources. (**c**) Heatmap of pairwise variant counts among CT48 isolates, clearly showing 4 distinct haplotypes. Most differences are due to the recombination tracts distinguishing the isolates. P618-9380 in haplotype A is more similar to haplotypes B and C than they are to each other, and as seen above, are the result of recombination events in two different lineages combining in this strain. (**d**) Probable transmission of CT48 strains among the 4 subjects (who all live within 800 meters of each other). The initially observed strain acquired two recombination tracts in independent lineages, which then combined into the lineages observed with P618-9380.

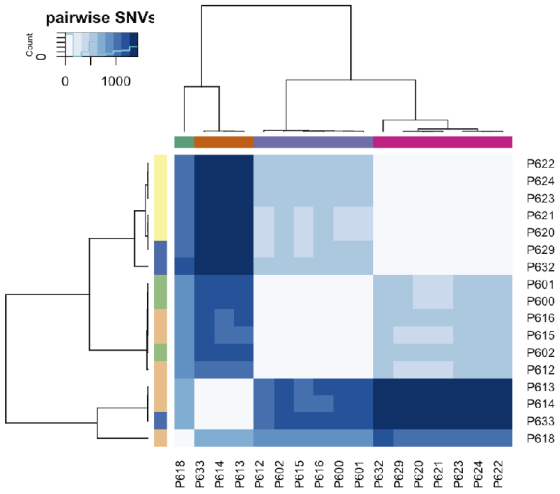
A



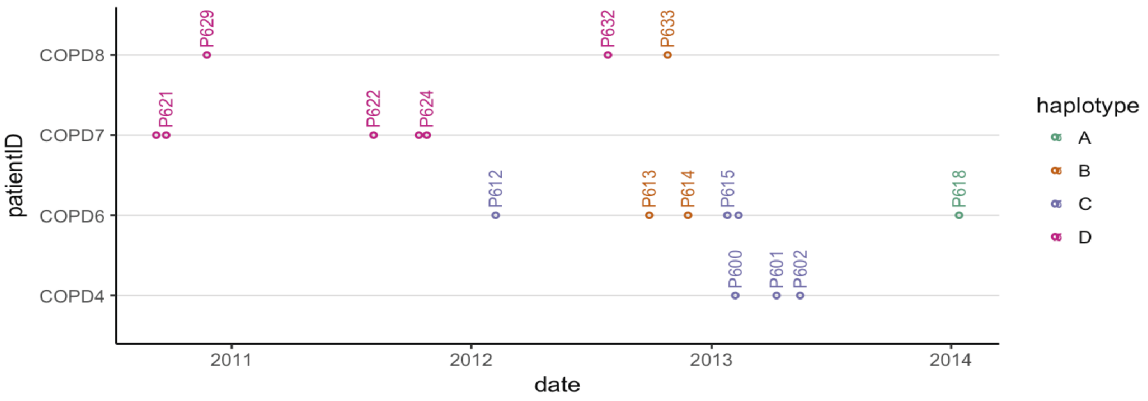
B



C



D



3.4.3. Recurrent genetic changes among clonal types, especially in genes encoding membrane-associated functions

Polymorphisms among isolates of the same CT—especially when collected from the same subject—point to recent mutation or recombination events that may contribute to the bacterium adapting to the COPD lung environment. Because only a single isolate was collected at each sampling point from what are most likely polyclonal populations, we were unable to polarize changes, in which we would treat earlier isolates as having parental genotypes and later ones as having derived genotypes (Ankrum & Hall, 2017; Guttman et al., 2006; Lee et al., 2017). However, independent polymorphisms at the same locus in different lineages and different patients could indicate recent adaptive evolution (or alternatively, hotspots of mutation or recombination).

To rank genes by evidence of recurrent changes in distinct NTHi lineages, we first identified genetic variants (SNVs, short indel, and gene possession) that occurred within each of the 20 multi-isolate CTs. This showed that intra-clonal genetic variation was characterized by only few genetic differences affecting only a few loci, either as isolated point mutations or clusters of closely-spaced SNVs characteristic of natural transformation events (Croucher et al., 2012; Mell et al., 2011; Mell & Redfield, 2014) (**Figure 3.2A and 3.2B**). We next classified variants based on their impact on protein coding and ranked homologous protein-coding gene clusters by the number of CTs affected by intra-CT polymorphism. Of the total gene clusters (data not shown), only 299 were polymorphic for short variants in one or more of the 20 multi-isolate CTs. Of these, only 15 gene clusters had intra-CT polymorphisms in three or more clonal types (**Table 3.1**), and only 5 clusters had high-impact short variants in three or more CTs (bold in **Table 3.1**).

The top hits in **Table 3.1** affected by high-impact variants encoded membrane proteins, enzymes determining lipooligosaccharide (LOS) structure, and a DNA methyltransferase subunit. Of these, several are known to be phase-variable due to simple sequence repeats in the open-reading frame (*lic2A, oafA, losA*, and a phase-variable DNA methyltransferase *hsdM3* (Deadman et al., 2009; Erwin et al., 2006; Fox et al., 2007)). The frequent frame-shift variants arising within CTs in these genes are due to copy number changes in these repeats. Also among top hits, were several genes affected by dozens of lower impact missense and synonymous intra-CT variants affecting the same cluster (*hgpB*, *hgpC*, *tbp2*), which were due to overlapping recombination tracts in

different CTs (**Figure 3.2A**), potentially indicating loci undergoing diversifying selection (Morton et al., 2012; Whitby et al., 2012)

Table 3.1. Ranked list of gene clusters with intra-clonal polymorphism, ranked by number of CT affected by short variant (SNVs or short indels) **or gene presence/absence polymorphisms within the CT.** Variants occuring in different CTs were counted independently.

Cluster ID	Gene name	Annotation	Intra-CT variation	Intra-CT gene presence variation	Short Variants	Frame-shifts	Nonsense	Inframe Indels	Missense	Synonymous
cluster1166	***lic2A***	**UDP-Gal--lipooligosaccharide galactosyltransferase**	**8**	0	11	8	0	2	1	0
cluster1300	***hgpB***	**Hemoglobin-haptoglobin binding protein B**	**8**	0	161	5	0	3	68	85
cluster1540	*hgpC*	Hemoglobin-haptoglobin binding protein C	**7**	4	118	0	0	2	54	62
cluster0819	*ompP2*	Outer membrane protein P2 precursor	**6**	0	25	0	0	8	16	1
cluster1230	***ompP1/ fadL***	**Long-chain fatty acid outer membrane transporter**	**6**	1	40	4	3	0	20	13
cluster1107	***tbp2***	**Transferrin-binding protein 2 precursor**	**5**	0	107	8	0	3	73	23
cluster1542	*tuf*	Elongation factor Tu	**4**	6	7	0	0	0	2	5
cluster0965	*spoT*	Guanosine-3',5'-bis 3'-pyrophosphohydrolase	**4**	0	5	0	0	0	4	1
cluster1529	*group_1017*	Glycosyl transferase family 8 protein	**3**	1	3	2	0	1	0	0
cluster0617	*hup*	Heme-utilization protein Hup	**3**	0	76	0	0	1	31	44
cluster1786	*group_315*	Putative uncharacterized protein	**3**	4	20	2	0	0	13	5
cluster1467	*hsdM3*	Putative type I restriction enzyme HindVIIP M protein	**3**	0	77	2	0	2	11	62
cluster1659	*hmw1C/hmw2C*	Putative glycosyltransferase involved in glycosylation of HMW1A and HMW2A	**3**	0	3	0	0	0	0	3
cluster1245	***oafA***	**O-antigen acetylase**	**3**	0	3	3	0	0	0	0
cluster1679	*hmw1A/ hmw2A*	High molecular weight adhesin 1/2	**3**	1	5	0	0	0	5	0

3.4.4. Genomic traits associated to acquisition of resistance to cotrimoxazole by NTHi within the COPD lung

Bacterial resistance to a panel of ten conventional antibiotics was performed for the entire COPD isolate collection, aiming to identify intra-clonal or intra-patient genetic traits associated to changes in antimicrobial resistance (**Table 3.2**). We focused on cotrimoxazole (SXT) because we found a significant proportion of resistance strains, and up to four CTs containing both sensitive and resistant strains (CT44, 76, 96 and 100). In NTHi, resistance to SXT is associated to polymorphisms and/or short insertions in the *folH* and *folP* genes, DHFR overproduction, or acquisition of the sulfonamide (SUL) genes *sul1* and *sul2* (de Groot et al., 1988; de Groot et al., 1996; Enne et al., 2002). Existing evidence also relates *H. influenzae* resistance to thymidine auxotrophy, due to loss-of-function mutations in the thymidylate synthase encoding gene *thyA* (Platt et al., 1983; Rodriguez-Arce et al., 2017). Genomic analysis showed that *sul1* and *sul2* were absent in the NTHi SXT resistance.

The *folH* gene was shown to be present and contain full-length variable sequences among the 92 new genomes. The *folH* gene allelic variation rendered 23 different alleles, being allele 10 the most frequent one, present in 17 isolates (**Table 3.2**). Although specific alleles were in most cases associated to specific CT, *folH* variants 1, 4, 9 and 14 were present in strains belonging to more than one CT (allele 1, in 11 strains distributed in 8 CTs; allele 4, in 19 strains distributed in 9 CTs; allele 9, in 2 strains distributed in 2 CTs; allele 14, in 9 strains distributed in 3 CTs) (**Table 3.2**). An alignment of FolH variants associated to either SXT sensitivity (encoded by *folH* alleles 11, 19, 13, 1, 16, 20, 7, 18, 14, 6, 3) or resistance (encoded by *folH* alleles 22, 8, 2, 5, 15, 10, 9, 12) revealed nine amino acid changes potentially associated to acquisition of SXT resistance (M21I, I95L, T118A, T124I, E132D, K143E, R149H and F154V/S) (**File S3.2**). Given that CTs 44, 76, 96 and 100 contained both SXT^{s} and SXT^{R} isolates, we specifically compared their FolH sequences. Strains in CT76, 95 and 100 showed the same *folH* variant per CT, excluding a possible link between *folH* gene mutations and acquisition of SXT resistance. By contrast, SXT^{s} and SXT^{R} isolates contained different *folH* variants in CT44 (alleles 9, 12 and 13). Isolates from patient 9 belonging to CT44 (P634 and P635 are SXT^{R} and contain *folH* allele 12; P636 and P637 are SXT^{S} and contain *folH* allele 13) allowed tracking *folH* gene transitions occurring within the same host niche over time, from ancestor to derived isolates, maybe linked to the observed changes in SXT resistance. Indeed, two of the identified changes, M21I and I95L, were observed while comparing FolH proteins encoded by alleles 13 and 12 (**SFile S3.2**). Amino acid

substitutions M21I, I95L and F154V were previously linked to acquisition of resistance to trimethoprim (de Groot et al., 1996), and our data further support such observations. In addition, amino acid substitutions N13S, P54A, A56P, I74V, F79L, E135K and R142H were previously suggested to be associated to trimethoprim resistance (de Groot et al., 1996). From those, N13S, I74V, F79L and E135K were also found in our strain collection, but in both SXT sensitive and resistant strains, therefore limiting their involvement in acquisition of this antibiotic resistance (**File S3.2**).

We also analyzed the *folP* gene distribution and variation. The *folP* gene was present and containing full-length variable sequences in our COPD isolate collection. *folP* gene allelic variation rendered 27 different alleles, being allele 1 the most frequent one, present in 20 isolates (**Table 3.2**). Specific alleles were mostly associated to specific clonal types, but *folP* allelic variants 1, 4, 7, 8, 9 and 11 were present in strains belonging to more than one CT (allele 1, in 20 strains distributed in 11 CTs; allele 4, in 3 strains distributed in 2 CTs; allele 7, in 7 strains distributed in 2 CTs; allele 8, in 4 strains distributed in 2 CTs; allele 9, in 18 strains distributed in 2 CTs; allele 11, in 3 strains distributed in 2 CTs) (**Table 3.2**). Overall, the *folP* gene variant 1 was present in 19 SXT^S isolates; however, strain P617-9224 (CT44) contained such *folP* variant while being SXT^R. Given that such resistance could relate to the observed *folH* gene variation or, alternatively, to other unexplored mechanism, we considered the *folP* gene variant 1 as linked to SXT^S for further analysis. An alignment of FolP variants associated to either SXT sensitivity (encoded by *folP* alleles 20, 8, 22, 14, 4, 18, 6, 16, 21, 3, 26, 12, 19, 24) or resistance (encoded by *folH* alleles 11, 10, 25, 27, 13, 2, 15, 5) revealed eight amino acid changes potentially associated to SXT resistance (S31R, T62A, A129T, G190C, G212V, D230N, P232A, A262S) (**Figure S3.2**). A 15-bp insertion corresponding to the addition of codons for Ser-Phe-Leu-Tyr-Asn after Pro64 and replacement of Asn65 by Asp was shown to be present in NTHi SXT^R strain containing FolP variant 10; moreover, FolP variants 11, 25 and 27, also present in SXT^R strains, have a 3-bp insertion corresponding to the addition of a codon for Asp after Asn65. Both insertion types have been previously linked to NTHi resistance to SXT (Enne et al., 2002). We also compared the *folP* gene sequences in isolates belonging to CTs 44, 76, 96 and 100. Strains in CT76, 96 and 100 showed the same *folP* variant, excluding a possible link between *folP* gene mutations and acquisition of SXT resistance. SXT^S and SXT^R isolates contained different *folP* variants in CT44 (alleles 1 and 11). Same as before, strains from patient 9 belonging to CT44 (P634 and P635 are SXT^R and contain *folP* allele 11; P636 and P637 are SXT^S and contain *folP* allele 1) allowed tracking *folP* gene transitions occurring within

the same host niche over time, from ancestor to derived isolates. In this case, one of the identified amino acid changes, G189C, was observed while comparing FolP proteins encoded by the *folP* alleles 1 and 11 (**File S3.2**). In fact, the substitution of conserved residue Gly189 by Cys has been suggested to be involved in raising the sulfamethoxazole MIC (Enne et al., 2002).

Table 3.2. **Antimicrobial susceptibility testing by disc difussion assay.** The *folP* and *folH* gene alleleic variation within this NTHi collection is indicated in columns 5 and 6, and supported by **Supplementary File S3.2**.

Clonal Type	Strain	Patient	SXT		*folP* allele	*folH* allele	PxE	Amp	AMC	CEC	CID	CIP	Cm	CTX	TET	CFU
3	P667-4462	13	28	S	1	14	0.032	26	28	25	28	28	35	37	35	29
	P668-6062	13	30	S	1	14	0.047	27	26	26	25	32	33	30	NA	28
	P669-6977	13	30	S	1	14	0.047	25	27	25	27	36	32	32	30	29
6	P626-9221	7	30	S	18	18	0.032	27	28	26	29	32	33	36	29	19
7	P670-7113	13	31	S	4	4	0.064	22	28	24	29	35	31	31	29	33
	P672-7661	13	34	S	4	4	0.064	24	30	25	28	40	34	32	30	31
8	P619-9590	6	**6**	**R**	15	15	0.032	23	22	23	24	30	30	28	29	26
9	P589-8275	1	26	S	3	3	**0.023**	22	22	20	21	30	30	28	29	22
	P590-8360	1	32	S	3	3	**0.064**	25	25	24	25	38	33	33	32	26
14	P594-8239	2	26	S	7	6	0,064	22	23	24	25	34	29	29	29	24
	P595-8370	2	30	S	7	6	0,047	28	25	27	25	40	36	32	31	27
	P596-8591	2	32	S	7	6	0,064	29	29	31	30	36	36	32	30	30
	P650-8603	10	31	S	7	6	0.064	28	28	33	31	37	40	39	29	28
16	P676-2514	3	34	S	1	14	**0.032**	28	30	27	30	42	32	32	35	32
	P679-2791	3	32	S	1	14	**0.125**	NA	NA	NA	NA	NA	NA	NA	NA	NA
	P853-2792	3	32	S	1	14	**0.094**	NA	NA	NA	NA	NA	NA	NA	NA	NA
17	P627-3766	8	28	S	1	1	0.032	33	32	30	30	38	NA	30	35	35
	P628-3847	8	29	S	1	1	0.032	28	32	26	26	33	30	30	30	33
18	P641-4342	10	**6**	**R**	2	2	0.094	20	20	20	21	33	31	26	29	18
	P642-4396	10	**6**	**R**	2	2	0.094	23	23	24	26	40	33	38	33	26
27	P588-8079	1	29	S	1	1	0.094	25	25	22	25	37	31	31	29	26
28	P671-7552	13	**6**	**R**	25	22	0.032	14	28	24	26	37	32	33	29	32
35	P610-7579	6	30	S	1	1	0.032	22	23	22	24	35	32	33	29	24
38	P643-7915	10	23	S	8	14	0,023	22	24	20	22	32	31	27	29	27
	P644-8083	10	34	S	8	14	0,047	30	30	33	29	38	38	36	33	32
	P645-8193	10	33	S	8	14	0,032	27	28	30	32	30	30	30	30	31
40	P607-8844	5	40	S	14	4	0.064	30	32	30	30	40	30	30	30	33
	P608-8895	5	34	S	14	4	0.125	28	28	26	31	35	36	30	25	26
44	P617-9224	6	**6**	**R**	1	9	0.064	22	20	10	20	29	29	29	19	18
	P634-7311	9	**6**	**R**	11	12	0.125	25	26	25	29	39	39	34	36	27
	P635-7537	9	**6**	**R**	11	12	0.094	26	28	30	28	37	30	30	30	28
	P636-8296	9	18	S	1	13	0.125	24	23	22	22	32	31	28	29	21
	P637-8346	9	19	S	1	13	0.094	28	28	32	27	25	32	38	32	24

45	P674-1509	3	32	S	26	4	0.125	26	27	26	26	21	32	31	30	NA
	P675-1755	3	29	S	26	4	0.125	23	21	23	25	21	35	32	31	NA
	P677-2580	3	32	S	26	4	0.125	24	29	25	26	21	35	29	30	27
	P678-2720	3	32	S	26	4	0.094	25	24	23	25	20	36	27	29	23
47	P661-4460	12	30	S	24	1	0.047	30	35	30	30	40	30	30	30	35
48	P600-8643	4	**6**	**R**	9	10	0,023	24	24	21	22	40	32	32	30	23
	P601-8794	4	**6**	**R**	9	10	0,023	28	27	20	26	41	37	28	30	26
	P602-8883	4	**6**	**R**	9	10	0,023	27	26	23	23	40	35	34	31	25
	P612-8066	6	**6**	**R**	9	10	0,016	24	27	20	24	40	31	29	29	21
	P613-8456	6	**6**	**R**	9	10	0,032	30	30	30	30	40	30	30	30	30
	P614-8522	6	**6**	**R**	9	10	0,032	32	28	27	31	43	25	36	31	29
	P615-8618	6	**6**	**R**	9	10	0,016	27	30	22	24	40	32	36	30	28
	P616-8647	6	**6**	**R**	9	10	0,016	24	24	20	24	40	32	32	31	24
	P618-9380	6	**6**	**R**	9	10	0,016	24	26	20	23	42	38	36	33	26
	P620-6998	7	**6**	**R**	9	10	0,023	22	24	20	21	40	32	30	29	21
	P621-7028	7	**6**	**R**	9	10	0,016	22	26	20	24	40	34	32	31	24
	P622-7806	7	**6**	**R**	9	10	0,016	24	22	20	22	39	31	29	29	21
	P623-7854	7	**6**	**R**	9	10	0,016	30	30	30	30	30	30	30	30	30
	P624-7870	7	**6**	**R**	9	10	0,016	27	24	6	24	27	23	28	20	24
	P629-7106	8	**6**	**R**	9	10	0,016	24	24	20	23	36	31	32	32	21
	P632-8384	8	**6**	**R**	9	10	0,016	25	27	26	27	41	34	34	32	27
	P633-8486	8	**6**	**R**	9	10	0,016	30	30	25	30	40	30	30	30	28
52	P609-9724	5	**6**	**R**	11	9	0.047	27	26	23	26	33	30	35	30	26
54	P646-8484	10	29	S	21	4	**0.032**	25	29	24	26	37	31	29	30	29
	P647-8506	10	35	S	21	4	**0.094**	35	33	32	37	40	37	44	31	32
	P648-8526	10	25	S	21	4	**0.047**	24	26	23	25	38	32	32	29	26
	P649-8561	10	24	S	21	4	**0.064**	23	24	24	24	31	29	27	29	25
59	P597-8879	2	34	S	8	4	0.032	27	25	27	28	42	33	31	32	30
72	P604-7629	5	33	S	1	1	0.032	30	30	26	32	40	35	37	39	30
	P605-7719	5	32	S	1	1	0.047	25	25	28	24	36	30	35	30	32
73	P599-8624	4	**6**	**R**	10	8	0.016	23	27	20	23	41	32	34	25	24
	P651-8849	10	**6**	**R**	10	8	**0.023**	25	26	24	27	40	30	30	30	27
	P652-8881	10	**6**	**R**	10	8	**0.023**	30	33	30	30	37	30	30	30	35
	P653-8956	10	**6**	**R**	10	8	**0.064**	30	22	30	30	22	30	30	30	22
	P654-8983	10	**6**	**R**	10	8	**0.047**	30	35	22	30	40	30	30	30	34
76	P664-7614	12	27	S	7	4	0,023	25	23	32	30	32	32	32	32	25
	P665-7858	12	26	S	7	4	0,023	25	22	22	25	30	30	30	30	28
	P666-8053	12	**6**	**R**	7	4	0,023	27	27	27	26	31	29	30	23	26
83	P640-3960	10	29	S	20	19	0.064	16	25	28	29	40	35	35	35	33

87	P638-3144	10	30	S	16	16	0.047	25	25	23	25	38	33	28	30	27
91	P639-3649	10	30	S	19	3	0.125	26	28	30	27	32	32	35	30	30
95	P625-8065	7	**6**	**R**	17	17	0.064	27	24	18	25	33	29	34	29	24
	P673-8639	13	20	S	17	17	0.032	22	23	25	21	20	29	31	29	20
100	P657-8759	11	26	S	23	21	0,032	28	30	26	29	35	36	30	31	21
	P658-8889	11	**10**	**R**	23	21	0,032	30	22	20	24	26	30	29	29	22
	P660-9367	11	20	S	23	21	0,032	28	32	25	29	28	29	30	29	26
105	P656-8020	11	30	S	22	20	0.032	22	27	22	23	26	30	33	30	27
106	P662-7189	12	32	S	9	7	**0.023**	26	35	29	27	28	34	33	31	34
	P663-7216	12	31	S	1	7	**0.064**	26	29	25	28	39	30	31	30	29
107	P591-8654	1	32	S	4	4	0.047	27	31	28	28	40	30	34	29	28
119	P603-4482	5	30	S	12	11	0.064	28	25	26	30	40	35	32	35	27
124	P630-7232	8	35	S	1	1	0.032	27	28	25	27	40	20	30	30	30
135	P593-7974	2	27	S	6	4	0.064	30	30	30	30	30	30	30	30	33
137	P598-8560	4	34	S	1	1	0.047	24	28	24	24	40	32	28	29	24
	P631-8237	8	23	S	1	1	0.047	26	25	24	25	34	31	31	29	27
138	P611-7680	6	38	S	1	1	0.032	28	33	27	35	40	35	35	30	33
140	P851-2713	3	**6**	**R**	23	0.047	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
145	P606-8650	5	**6**	**R**	13	4	0.064	25	24	27	26	32	30	31	29	26
146	P592-9166	1	**6**	**R**	5	5	0.064	22	20	19	20	38	32	28	30	20

AMP: Ampicillin, AMC: Amoxicilin-Clavulanic, CEC: Cefaclor, CID: Cefonicid, CIP:Ciprofloxacin, CLO: Chloramphenicol, SXT: Cotrimoxazole (Sulfametoxazole-Trimetropim), CTX: Cefotaxime, TET: Tetracycline, CFU: Cefuroxime. CLSI criteria for disc diffusion assay breakpoints (mm): AMP: S (≥ 22), I (19-22), R (≤ 18); AMC: S (>20), R (<19); CEC: S (≥20), I (17-19), R (<16); CID: S (≥20), I (17-19), R (<16); CIP: S (>21); CLO: S (≥ 29), I (26-28), R (≤ 25); SXT: S (≥ 16), I (11-15), R (≤ 10); CTX: S (≥ 26); TET: S (≥ 29), I (26-28), R (≤ 25); CFU: S (≥ 20), I (17-19), R (≤ 16). S: Sensitive, I: Intermediate, R: Resistant.

3.4.5. Increased resistance to polymyxin E as pathoadaptive signature within the COPD lung

AMPs are positively charged innate immunity soluble peptides whose action is initiated through electrostatic interaction with the bacterial surface, the anionic lipid A moiety of the LPS in Gram-negative bacteria (Brogden, 2005; Hancock, 1999; Nicolas & Mor, 1995; Vaara, 1992). Polymyxins are pentacationic amphipathic lipopeptide antibiotics active against Gram negative bacteria and, similar to AMPs, they interact with the anionic LPS (Evans et al., 1999; Hermsen et al., 2003). Increasing resistance to the synthetic AMP polymyxin B (PxB) and to the natural AMP human β-defensin-1 (hBD-1) among NTHi clonal isolates has been reported (Garmendia et al., 2014). Following this notion, we determined polymyxin E (PxE) MIC for all NTHi strains (**Table 3.2**), and assessed phenotypic variation among clonal isolates from the same patient over time. We observed increasing resistance to PxE in several isolate sets, and next focused on sets containing strains displaying MIC differences ≥3 lanes in the PxE e-test strip (labeled in red bold in **Table 3.2**). SNVs were identified that distinguished the isolates contained within each set. Variants in coding sequences causing dN changes are shown in **Table 3.3**. These dN changes may be selectively beneficial. These genes encode proteins involved in heme acquisition, LOS biosynthesis, or outer membrane transport.

Table 3.3. Non-synonymous SNPs in NTHi clonal sets isolated from the same patient and displaying MIC differences ≥3 lanes in the PxE e-test strip.

ORF	REF	ALT	P651-8849	P652-8881	P653-8956	P654-8983	Aa Ref	Aa Alt	Aa pos	Protein length	Gene name	Function
309		92 del. A	0	0	1	0	Lys	fs	33	619	*selB*	selenocysteine-specific elongation factor
405	T	C	0	0	1	0	Leu	Pro	139	189	*rpoE*	RNA polymerase sigma factor
429		45-48del. CCAA	0	1	0	1	Asn	fs	15	966	*hgpB*	Hemoglobin and hemoglobin-haptoglobin binding protein B
434		454-455 ins. G	0	1	0	0	Ter	fs	156	155	*cpdB*	O-methyltransferase

ORF	REF	ALT	P676-2514	P679-2791	P853-2792	Aa Ref	Aa Alt	Aa pos	Protein length	Gene name	Function
12		61-64 del. TCAA	0	0	1	Lys	fs	22	316	*lic3A*	UDP-Gal-LOS galactosyltransferase
435		120-123 dup. CCAA	0	0	1	Asn	fs	42	1045	*hgpC*	Hemoglobin and hemoglobin-haptoglobin binding protein nB
441		454-455 ins. G	1	0	1	Ter	fs	156	155	*cpdB*	O-methyltransferase
656	A	T	1	1	0	Trp	Arg	267	439		Hypothetical protein
739		82-85 del. TCAA	0	0	1	Lys	fs	29	322	*lic3B*	UDP-Gal-LOS galactosyltransferase
1351		39-45 del. ACCAACA	0	1	0	Thr	fs	23	796	*Fe upt*	TonB-dependent transport protein, Fe transport

ORF	REF	ALT	P662-7189	P663-7216	Aa Ref	Aa Alt	Aa pos	Protein length	Gene name	Function
331	A	C	0	1	Asn	Thr	251	1068	*hgpC*	Hemoglobin and hemoglobin-haptoglobin binding protein C
414	T	A	1	0	Ile	Phe	132	394		Elongation factor Tu
1647	T	A	1	0	Val	Asp	275	1477	*hmw2A*	High molecular weight adhesion-invasin

ORF	REF	ALT	P589-8275	P590-8360	Aa Ref	Aa Alt	Aa pos	Protein length	Gene name	Function
407	A	C	1	0	Ser	Arg	22	470		Putative ATPase, hypothetical protein
407	A	C	1	0	Asn	Thr	282	470		Putative ATPase, hypothetical protein
640	G	T	1	0	Pro	Thr	513	844	*infB*	translation initiation factor IF-2
728	A	C	1	0	Ter	Ser	229	228	*fadL*	Long-chain fatty acid transporter
1000	T	C	1	0	Asn	Asp	59	137		phage GP46 family protein,hypothetical protein
1011	A	G	1	0	Ser	Pro	91	101		Hypothetical protein
1014	C	A	1	0	Glu	Asp	131	161		phage virion morphogenesis protein
1016	T	C	1	0	Asn	Asp	512	546		Hypothetical protein
1016	TG	CA	1	0	Lys	Glu	514	546		Hypothetical protein

ORF	REF	ALT	P646-8484	P647-8506	P648-8526	P649-8561	Aa Ref	Aa Alt	Aa pos	Protein length	Gene name	Function
936		112 del. G	0	1	1	0	Ser	fs	41	85	*ptsH*	phosphocarrier protein HPr

del.: deletion, ins.: insertion, fs.: frameshift, dup.: duplication

3.4.6. Phase variation in the *hmw1A* gene promoter region modulates NTHi epithelial invasion during COPD colonization

We have previously determined that specific *hmw1A* allelic variants favor NTHi self-aggregation and subsequent airway epithelial cell adhesion and invasion (Chapter 2 and (Mell et al., 2016)). Moreover, HMW1 is known to undergo graded phase variation controlled by the number of 7-bp repeats (5´-ATCTTTC)$_n$ upstream of the *hmw1A* gene (Cholon et al., 2008). We asked about the frequency of the NTHi epithelial hyperinvasive phenotype (Mell et al., 2016) in the COPD isolate collection under study, and its role in adaptation. Bacterial self-aggregation as a means of possible hyperinvasion (data not shown) was screened on the entire strain collection to select highly self-aggregative strains. Airway epithelial invasion was tested for a representative set of highly self-agreggative NTHi strains, including P651-8849, P652-8881, P653-8956 and P654-8983 (CT73, patient 10). Isolate P652-8881 invasion rate was even higher than that of strain 86-028NP (Mell et al., 2016).

HMW1A$_{P652}$ and HMW2A$_{P652}$ alleles were identified based on binding domains referred to those previously assigned to NTHi$_{strain12}$ (**Table 3.4**). Moreover, HMW1A$_{P652}$ and HMW1A$_{86\text{-}028NP}$ share 90.7% aa identity, thus pointing to the hyperinvasive phenotype shown by strain 86-028NP (Mell et al., 2016). Of note, clonal strains P651-8849, P652-8881, P653-8956 and P654-8983 showed progressively decreasing epithelial invasion rates (**Figure 3.3A**). Genomic variation between these four strains are 4 SNVs, and none of them are located at any *loci* potentially related to interaction with eukaryotic cells (**Table 3**). The *hmw1A* gene sequence was shown to be identical for these four strains (**File S3.3**). However, we observed a progressive increase in the number of 7-bp repeats upstream of the *hmw1A* gene, ranging from 11 to 18 repeats, which in turn correlated with a progressive reduction of HMW1A protein levels (**Figures 3.3B, 3.3C, 3.4**). NTHi strains harboring HMW1 are also likely to present the HMW2 operon (Buscher et al., 2004). We did not observe significant changes in the number of 7-bp repeats upstream of the *hmw2A* gene (**Figure 3.3B**); HMW2A protein levels were comparable among strains (**Figure 3.3C**).

Thus, we identified a second *hmw1A* allelic variant (together with the one described in Chapter 2) likely to be associated to NTHi hyperinvasion. Increasing number of 7-bp repeats upstream of the *hmw1A* gene in a given clonal strain reflected a progressive decrease of HMW1A protein amount and NTHi airway epithelial invasion. A previous study on *H. influenzae* isolates collected serially from COPD patients showed that HMW1

and HMW2 expression in a given strain also decreased over time reflecting a progressive increase in the numbers of 7-bp repeats, associated with high serum titers of HMW1/HMW2-specific antibodies (Cholon et al., 2008). Overall, HMW phase variation is likely to be a genomic trait of NTHi patho-adaptation within the COPD lung.

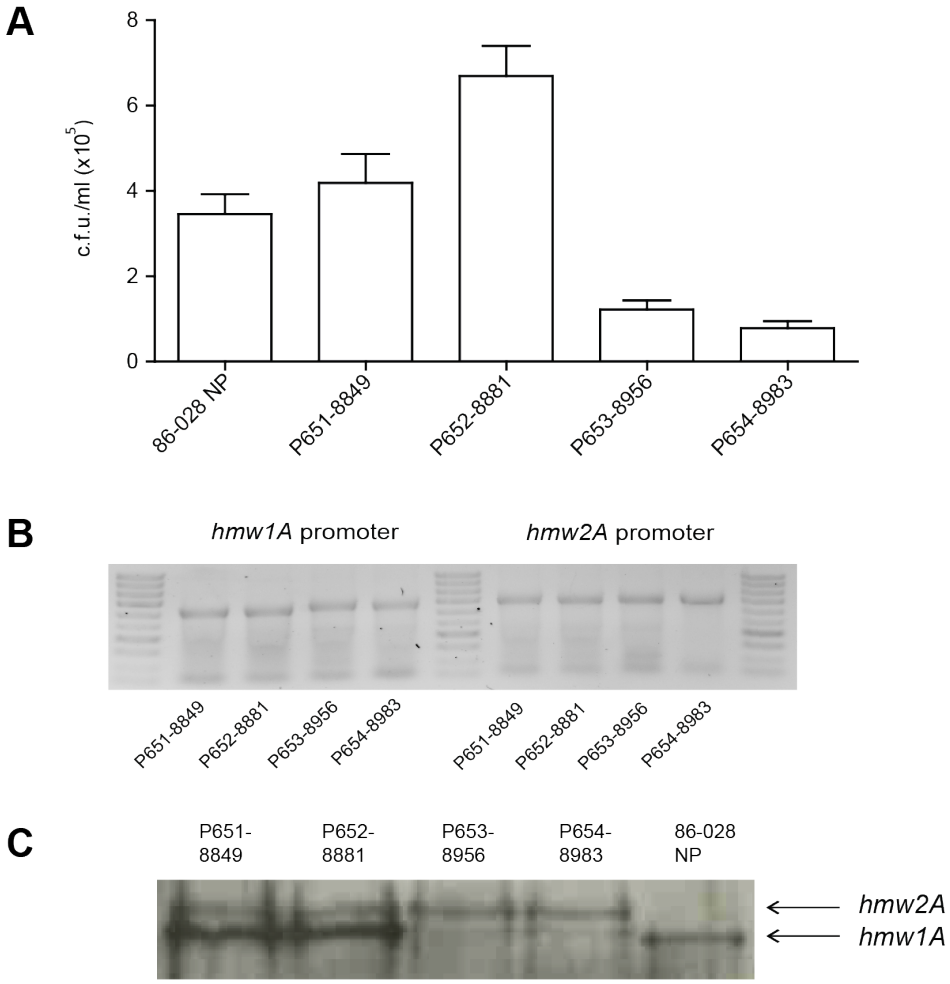


Figure 3.3. The *hmw1A* gene phase variation is a genomic trait from NTHi evolution within the COPD lung. (**a**) Invasion rates of NTHi isolates P651-8849, P652-8881, P653-8956 and P654-8983, including 86-028NP as reference strain, in A549 cells. Invasion rate increases is high for strains P651-8849 and P652-8881, and then decreases for strains P653-8956 and P654-8983. (**b**) PCR amplification of a DNA region upstream the *hmw1A* (left) and *hmw2A* (right) genes encompassing a variable number of the heptanucleotide repeat $(5'\text{-ATCTTTC})_n$. Precise number of 7-bp repeats in the *hmw1A* gene promoter region: P651-8849: 11, P652-8881: 12 , P653-8956: 18; P654-8983: 18 (**Figure 3.4**). (**c**) Immunodetection of HMW with a guinea pig anti-HMW1A gp85 antibody. More intense bands correspond to HMW1A protein in strains P651-8849 and P652-8881.

Table 3.4. HMW1A and HMW2A binding region identity for four NTHi strains. Three of them, Strain 12, 86-028NP and Hi375 have been described in Chapter 2 and (Mell et al., 2016). P652-8881 is presented in this Chapter 3.

Locus	Strain	HMW1A$_{st12}$	HMW2A$_{st12}$	Best Hit
yrbA-adjacent	Strain12	*	33.8%	HMW1A
	86-028NP	39.8%	34.8%	HMW1A
	Hi375	35.8%	50.7%	HMW2A
	P652-8881	**36.8**	**64.8%**	**HMW2A**
radA-adjacent	Strain12	33.8%	*	HMW2A
	86-028NP	32.8%	55.4%	HMW2A
	Hi375	100%	33.8%	HMW1A
	P652-8881	**41.6%**	**36%**	**HMW1A**

```
P651-8849       TCTTTCATCTTTCATCTTTCATCTTTCATCTTTC--------------------------
P654-8983       TCTTTCATCTTTCATCTTTCATCTTTCATCTTTC-TCTTTCATCTTTCATCTTTCATCTT
P652-8881       TCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTC-------------------
P653-8956       TCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTT
                **********************************

P651-8849       ----------------------ATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT
P654-8983       TCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT
P652-8881       ----------------------ATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT
P653-8956       TCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCAT
                                       ************************************

P651-8849       CTTTCACATGAAATGATGAACCGAGGGGAGGGAGGGGCAAGAATGAAGAGGGAGCTGAAC
P654-8983       CTTTCACATGAAATGATGAACCGAGGGGAGGGAGGGGCAAGAATGAAGAGGGAGCTGAAC
P652-8881       CTTTCACATGAAATGATGAACCGAGGGGAGGGAGGGGCAAGAATGAAGAGGGAGCTGAAC
P653-8956       CTTTCACATGAAATGATGAACCGAGGGGAGGGAGGGGCAAGAATGAAGAGGGAGCTGAAC
                ************************************************************

P651-8849       GAACGCAAATGATAAAGTAATTTAATTGTTCAACTAACCTTAGGAGAAAATATG
P654-8983       GAACGCAAATGATAAAGTAATTTAATTGTTCAACTAACCTTAGGAGAAAATATG
P652-8881       GAACGCAAATGATAAAGTAATTTAATTGTTCAACTAACCTTAGGAGAAAATATG
P653-8956       GAACGCAAATGATAAAGTAATTTAATTGTTCAACTAACCTTAGGAGAAAATATG
                ******************************************************
```

Figure 3.4. Alignment of the *hmw1A* promoter region sequences in NTHi strains P651-8849, P652-8881, P653-8956 and P654-8983. A putative translation start codon for the *hmw1A* gene is shown in green; 7-bp repeats (5´-ATCTTTC) are shown alternatively in bold. Precise number of 7-bp repeats in the *hmw1A* gene deduced promoter region: P651-8849: 11; P652-8881:12; P653-8956:18; P654-8983: 18.

3.4.7. Frequent independent loss-of-function mutations in *fadL* arise across COPD clonal types

Aside from changes in phase-variable genes, the most striking observation from the convergent evolution analysis above was that of recurrent independent predicted loss-of-function changes in the *fadL* gene (**Figure 3.1A and 3.5; Table 3.1; Tables S3.3, S3.4, S3.5 and File S3.1**). Of the 20 multi-isolate CTs, seven were affected by polymorphisms in *fadL*, and this included nine independent null alleles induced by frameshift or nonsense mutations. Since the above analysis only considered intra-CT polymorphisms and was sometimes affected by annotation problems at truncated alleles, we extended the analysis by extracting the *fadL* allele from each of our new genome assemblies using tblastx (**Methods**). Multiple alignment of the resulting 92 amino acid sequences identified thirty-four distinct *fadL* alleles encoding full-length protein, as well as sixteen distinct alleles with mutations predicted to encode truncated proteins (**Tables S3.3, S3.4 and S3.5**). Amino acid variation among the thirty-four FadL full-length protein variants is represented in **Figure 3.5A**, showing higher variability in predicted surface exposed loops. Predicted full-length and truncated variants are illustrated in **Figure 5B**, showing clear evidence for multiple independent null mutations at different positions in the gene, including both frameshift and nonsense mutations (**Table S3.5** and **File S3.1**). Predicted truncation alleles were found in 24 isolates collected from 11 out of the 13 patients and spanning 14 CTs. Where possible, we identified full-length alleles that were the likely ancestors of specific truncated variants, identifying sets of isolates from the same patient and CT with no variation in *fadL* outside the putative null mutation (**Figure 3.5B**, **Table S3.5**). We further provide direct evidence that the FadL protein is not detected in strains carrying predicted null alleles by SDS-PAGE of bacterial whole cell lysates from representative isolates. A protein band consistent with FadL predicted molecular weight was present in strains P589, P631, and P608, but absent from P590, P595, P662, P614, P622, and P607 (**Figure 3.5C**). Band identity as FadL was confirmed for strain P589 by peptide mass fingerprinting MALDI-TOF/TOF. These results strongly suggest *fadL* loss-of-function are selected for during NTHi adaptation to the COPD lung.

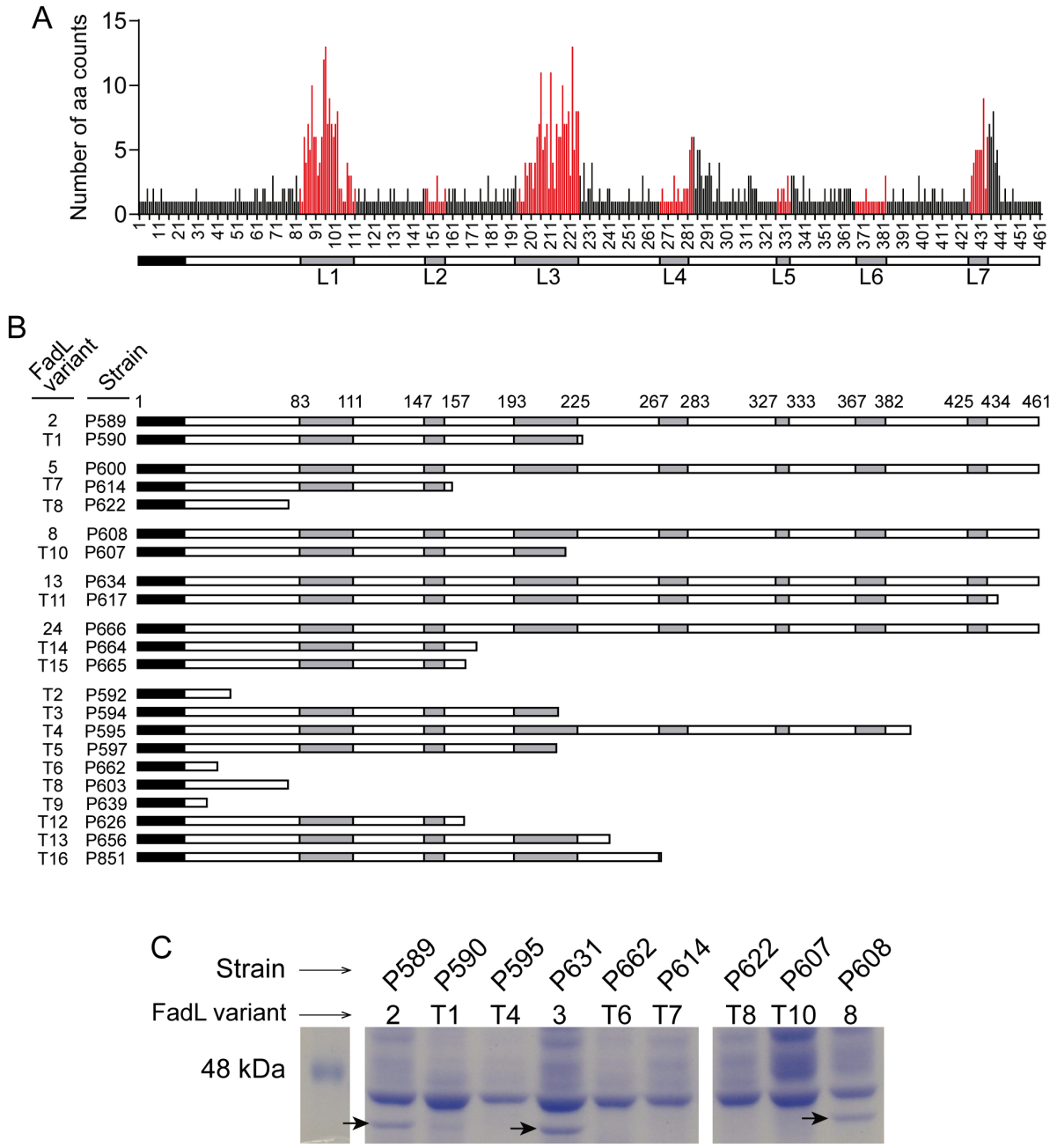
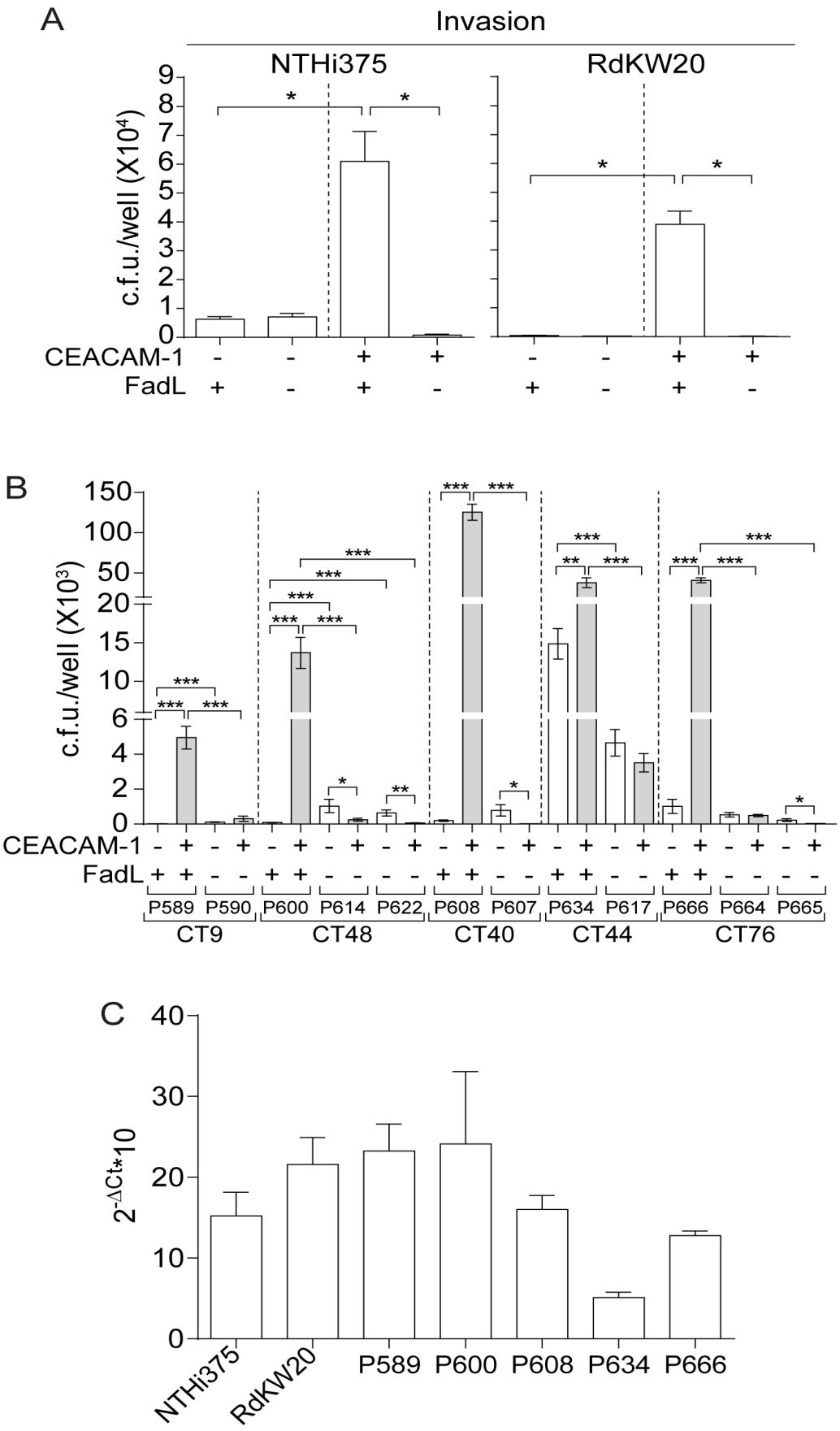


Figure 3.5. FadL allelic variation and loss-of-function is a main genomic trait from NTHi evolution within the COPD lung. (**a**) Count of distinct amino acid variants at each position in an alignment of 34 full-length FadL protein sequences seen in the set of 92 new COPD genomes, distributed through a total of 68 isolates. Colored bins correspond to residues in the predicted extracellular loops. Nineteen full-length variants were uniquely found in one strain, and fifteen were found in more than one isolate. FadL variant 5 was the most frequently found. Thirteen CTs contained strains with only one FadL full-length variant (CT 3, 7, 16, 17, 18, 38, 45, 48, 54, 72, 73, 95, 137) and two CTs contained strains with more than one FadL full-length variant (CT 44, 100) (**Tables S3.3** and **S3.4**). (**b**) Multiple independent *fadL* truncated alleles were observed throughout the COPD collection, shown schematically here to indicate surface-exposed loops and pairing of truncated alleles with full-length alleles where possible. (**c**) Putative null alleles do not express FadL protein. Whole protein extracts from bacterial cultures were analysed on 12% SDS-PAGE 20x20 cm gels followed by Coomassie Brilliant Blue staining. A protein band compatible with FadL predicted molecular weight (approximately 48 KDa) was present in strains P589-8079, P631-8237 and P608-8895, containing FadL full length variants 2, 3 and 8, respectively, and absent in strains P590-8360, P595-8370, P662-7189, P614-8522, P622-7806 and P607-8844, containing truncated variants T1, T4, T6, T7, T8 and T10, respectively (**Tables S3.4** and **S3.5**).

3.4.8. FadL variation affects hCEACAM1-dependent NTHi interactions with host cells

$FadL_{NTHi}$ has been reported to be a bacterial ligand for the human carcinoembryonic antigen-related cell adhesion molecule 1 (hCEACAM1), facilitating NTHi entry into epithelial cells (Tchoupa et al., 2015; Virji, 2000). We next assessed the impact of naturally occurring variation in the *fadL* gene on the interplay between NTHi and host cells. By comparing bacterial invasion of HeLa cells (which do not normally express hCEACAM1) to that of a stably transfected HeLa derivative cell line (expressing hCEACAM1, HeLa-BGP) (Gray-Owen et al., 1997), we first tested whether bacterial entry cells depends upon the interaction between FadL and hCEACAM1. As controls, we generated isogenic *fadL* knockout mutants in reference strains NTHi375 and RdKW20 (Fleischmann et al., 1995; Mell et al., 2014). Gentamicin protection assays showed strong defects in bacterial entry into HeLa cells in the absence of either FadL or hCEACAM1 in both strain backgrounds (**Figure 3.6A**).

Figure 3.6. FadL is a bacterial ligand of the hCEACAM1 receptor. (**a**) HeLa (non-expressing hCEACAM1, white bars) and HeLa-BGP (stably transfected derivative expressing hCEACAM1, gray bars) cells were used to quantify epithelial invasion by NTHi375 and RdKW20, both WT and Δ*fadL* derivatives. WT strains invaded HeLa-BGP cells at significantly higher rates than those obtained for HeLa cells (for NTHi375 and RdKW20, $p<0.0001$), in contrast to their respective Δ*fadL* mutants. Inactivation of *fadL* decreased invasion of HeLa-BGP cells, when compared to that of their respective isogenic WT strains ($p<0.0001$). (**b**) HeLa (white bars) and HeLa-BGP (grey bars) cells were also used to quantify invasion by COPD isolates representative of all available FadL full-length ancestor and derived truncated allele pairs (strains belonging to CT 9, 48, 40, 44 and 76; FadL full-length variants 2, 5, 8, 13, 24; FadL truncated variants T1, T7, T8, T10, T11, T14 and T15). Strains containing FadL full-length variants invaded HeLa-BGP cells at significantly higher rates than HeLa cells (for P589, P600, P608 and P666, $p<0.0005$; for P634, $p<0.005$), in contrast to strains containing FadL truncated variants. Moreover, natural *fadL* truncation alleles significantly lowered bacterial entry into HeLa-BGP cells compared to their full-length ancestors ($p<0.0005$ when comparing P590 to P589; P614 to P600; P622 to P600; P617 to P634; P607 to P608; P664 to P666; P665 to P666). (**c**) Expression of the *fadL* gene upon bacterial exponential growth in sBHI as measured by quantitative real-time PCR. Data are shown for strains containing FadL full-length variants assayed in panels (**a**) and (**b**), NTHi375, RdKW20, P589, P600, P608, P634 and P666. Expression of the *fadL* gene was comparable for all tested isolates.



As expected, COPD clinical isolates also showed FadL-dependent invasion of hCEACAM1-expressing cells. Pairs of clonal isolates carrying ancestral full-length and derived null alleles of FadL were tested (**Tables S3.4** and **S3.5**). In all cases, invasion of hCEACAM1-expressing cells was substantially decreased in the absence of full-length FadL (**Figure 3.6**). Moreover, significant differences in hCEACAM1-dependent invasion rates were observed among isolates carrying distinct full-length FadL alleles. These differences could not solely be explained by differences in *fadL* gene expression; for example, isolate P608-8895 showed the highest invasion rate, but *fadL* gene expression comparable to the other strains tested (**Figure 3.6C**).

We next examined the interaction of $FadL_{NTHi}$ with hCEACAM1 and the potential effect of variation in FadL's surface-exposed loops by using computational molecular modeling. This interaction has been suggested for $FadL_{NTHi}$ and not for $FadL_{E.\ coli}$ (Tchoupa et al., 2015), but a 3D perspective of the molecular recognition events is lacking. Homology models of FadL variants from RdKW20, NTHi375, and P608-8895, an (selected as a representative COPD isolate with high rates of HeLa-BGP cell invasion) were computed and refined by molecular dynamics (MD) simulations. All three computed 3D structures were similar to each other and to $FadL_{E.\ coli}$, though the mobile variable surface-exposed loops had distinct conformations between species and some differences were seen within the β-barrel (**Figures 3.7A, S3.4, S3.5, S3.6, S3.7** and **Supplementary Text**).

Next, we performed protein-protein docking simulations (Pierce et al., 2014) between these three $FadL_{NTHi}$ models and hCEACAM1 (PDB-ID: 4QXW). The resulting protein-protein complexes were submitted to MD simulations, rendering high stability along the simulation. Protein-protein interactions analysis of the three $FadL_{NTHi}$/hCEACAM1 complexes identified several relevant polar, hydrophobic and CH-π interactions that remained stable along the MD simulation, which varied depending on the allele used (**Figure 3.7B**, **Figure S3.10, Supplementary Methods and Results**). Among the residues involved in the interactions, we found strong interactions with hCEACAM1 residues Tyr34 and Ile91, previously reported as necessary for the interaction with *H. influenzae* strains from site-directed mutational analyses (Virji, 2000). Overall, these predicted protein-protein complexes account for the interaction between FadL and hCEACAM1. They further show that natural variation in FadL surface-exposed loops may directly affect NTHi's ability to interact with the epithelial cell surface.

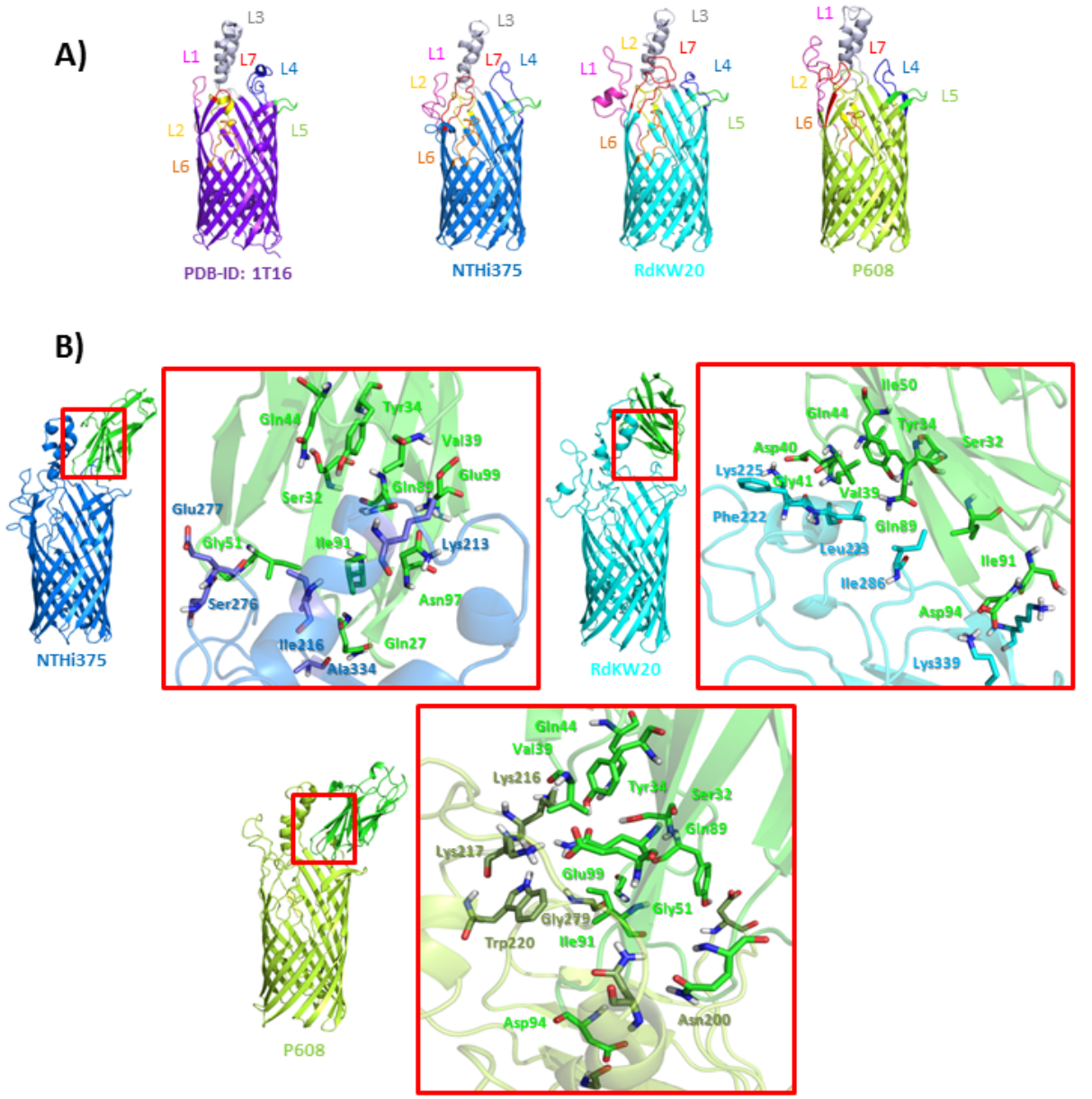


Figure 3.7. (**a**) **Computed homology models for *H. influenzae* FadL: $FadL_{NTHi375}$ in blue, $FadL_{RdKW20}$ in cyan and $FadL_{P608}$ in lime green.** On the left (purple), the X-ray crystallographic structure from $FadL_{E.\ coli}$ is shown (PDB-ID: 1T16). The resulting modelled 3D structures correspond to a 14-stranded β-barrel, in agreement with other similar proteins embedded in the *H.influenzae* outer membrane, with seven extracellular loops (L1 to L7), of which L1, L3, L4 and L7 are mobile and exposed to the outside. (**b**) Protein-protein docking between the amino-terminal domain of hCEACAM1 (PDB-ID: 4QXW) (green) and the modeled structures of $FadL_{NTHi375}$ (blue), $FadL_{RdKW20}$ (cyan) and $FadL_{P608}$ (lime green). Detail of the protein-protein interactions is depicted. These interactions remained stable along the MD simulation of each hCEACAM1/FadL complex and a more complete analysis can be found at the Supporting Information. Interestingly, the residues from hCEACAM1 (residues Ser32, Tyr34, Val39, Gln44, Gln89 and Ile91) that interact with FadL were previously reported to be crucial for the interaction with pathogen components (Villullas et al., 2007). Our computational studies clearly identify these particular residues as directly involved in the interaction with FadL, suggesting they shape a region relevant for the molecular recognition.

3.4.9. Loss of *fadL* function confers resistance to the bactericidal effects of arachidonic acid

$FadL_{NTHi}$ is also a predicted transporter of exogenous long-chain fatty acids (LCFAs). In *E. coli* and *Sinorhizobium meliloti*, the *fadL* gene is required for using free fatty acids as a sole carbon source (Hearn et al., 2009; Krol & Becker, 2014; Lepore et al., 2011), and key amino acid residues required for fatty acid binding and lateral diffusion have been elucidated for $FadL_{E.\ coli}$ (Hearn et al., 2009; Lepore et al., 2011). Docking calculations of three LCFAs were performed to the three homology models of $FadL_{NTHi375}$, $FadL_{RdKW20}$ and $FadL_{P608}$. Additional normal mode analysis (NMA) of the three $FadL_{NTHi}$ 3D models provided distinct stable conformations representing LCFA motions through the channel. Tested LCFAs included arachidonic acid (AA), oleic acid (OA), and lauryl dimethylamine-N-oxide (LDA) as a control (since it was previously crystallized with $FadL_{E.\ coli}$). All the calculations predicted several binding poses for the three fatty acids at the entrance to- and within the $FadL_{NTHi}$ β-barrel (**Figures 3.8**, **S3.11, S3.12 and S3.13**, **Supplementary Results**). All three fatty acids were predicted to interact with $FadL_{NTHi}$ at equivalent sites to those for LDA in $FadL_{E.\ coli}$, but other distinct binding sites were also identified inside the β-barrel. The docking calculations provide reasonable binding poses to $FadL_{NTHi}$ for the studied LCFAs, and support its role as a fatty acid transporter.

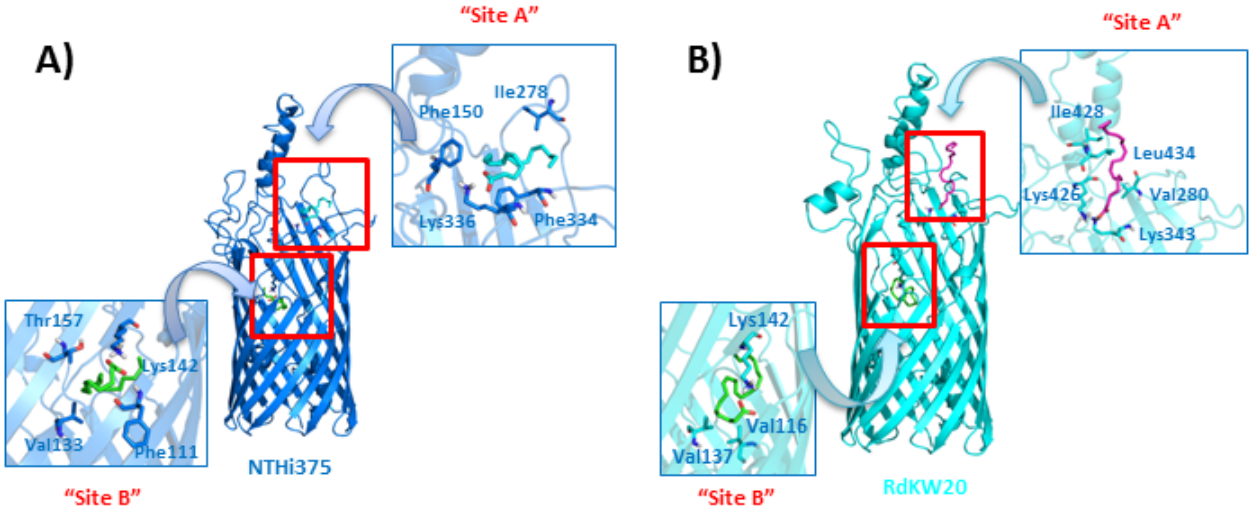


Figure 3.8. Docked binding poses of AA in $FadL_{NTHi375}$ (A, blue) and $FadL_{RdKW20}$ (B, cyan). Detail of the interactions is depicted in sticks. Calculations predicted two favourable binding modes of AA in both FadL structures ($FadL_{NTHi375}$ and $FadL_{RdKW20}$): a first one at the entrance of the β-barrel (site A), in a hydrophobic groove located between L3 and L4. Interactions with polar residues, specifically with Lys at L5 (Lys336 $FadL_{NTHi375}$ and Lys426 $FadL_{RdKW20}$), lipophilic and CH-π interactions with aliphatic and aromatic residues were identified; and a second binding mode at the depth of the β-barrel (site B), establishing polar contact with Lys142 present in L2. Interestingly, these binding poses were also found for $FadL_{P608}$ together with an additional binding site (see **Supplementary Results** for more details).

However, further use of LCFAs as a sole carbon source is unlikely due to the absence of a complete β-oxidation pathway in NTHi (Fleischmann et al., 1995). Indeed, no bacterial growth was observed in a defined minimal medium free of fatty acids (MM-FFA) when supplemented with AA as carbon source (**Figure S3.14A**).

Free fatty acids act as natural detergents, and a bactericidal effect on *H. influenzae* has been described for AA (Knapp & Melly, 1986). We next examined the viability of wildtype (WT) and isogenic *fadL* mutant strains when incubated with AA. WT strains had reduced viability in MM+AA, in a dose dependent manner; susceptibility was higher in RdKW20 than in NTHi375, maybe due to their distinct interactions with AA (**Figure 3.8**). By contrast, both *fadL* knockout mutants were more resistant to AA than their respective WT strains (**Figure 3.9A**). The same effect was seen for OA, though only at substantially higher concentrations (**Figure S3.14B**). Resistance to AA was also assayed for two representative pairs of COPD isolates containing predicted FadL ancestral full-length and derived null alleles (**Table S3.5**). In both cases, loss of FadL function was associated with increased bacterial resistance to AA (**Figure 3.9B**). Although one or more of the few additional genetic variants that distinguish the two pairs of isolates may be responsible

for this increased AA resistance, *fadL* loss-of-function was the only common one in to both isolate pairs (**Table 3.5**). Given that AA metabolites are key players in COPD-related airway inflammation (Malhotra et al., 2012), these molecules may in turn be a selective pressure within the COPD lung, and *fadL* loss-of-function bacterial pathoadaptive trait within this niche.

Figure 3.9. FadL loss-of-function confers resistance to arachidonic acid (AA) within the COPD lung. (**a**) NTHi375 and RdKW20 WT and *fadL* mutant strains grown on chocolate agar were used to generate normalized bacterial suspensions in MM-FFA, for further incubation with AA. A fatty acid dose dependent bactericidal effect was observed for both WT strains. Both *fadL* mutants were more resistant to AA than their respective parental strains (NTHi375, AA 25 μM, $p<0.01$; AA 50 μM, $p<0.005$; RdKW20, AA 25 μM, $p<0.005$). (**b**) FadL natural loss-of-function increases NTHi resistance to AA. Fatty acid resistance was assayed for representative clonal isolates pairs containing two predicted FadL full-length ancestor and derived truncated allele pairs. A fatty acid dose dependent bactericidal effect was observed. Strains P590 and P666, containing FadL truncated variants, were more resistant to AA than their respective clonal counterparts P589 and P665, respectively (P590, AA 25 μM, $p<0.0005$; AA 50 μM, $p<0.01$; P666, AA 25 μM, $p<0.0005$, AA 50 μM, $p<0.05$). In panels (**a**) and (**b**), results are expressed as % bacterial survival, referred to that in the presence of vehicle solution. (**c**-**d**) Inactivation of the *fadL* gene does not have an effect in a murine lung acute infection model. CD1 mice were intranasally infected with ~$2x10^8$ bacteria/mouse of WT (circle) or Δ*fadL* (square) strains. Mice were euthanized at 24 and 48 hpi, and bacterial loads were quantified in lungs (left) and BALF (right) samples (log_{10} c.f.u./sample). Data are shown for NTHi375 (panel c) and RdKW20 (panel d) backgrounds. In both cases, comparable bacterial loads in lungs and BALFs were observed for WT and mutant strains. (**e**) Model showing AA as a selective pressure for FadL loss-of-function within the COPD lung. AA metabolites are key players in COPD-related airway inflammation. PLA_2-catalyzed hydrolysis of membrane phospholipids results in production of a free fatty acids, most importantly AA. This environment at the COPD lung exerts a specific selection pressure to lose functional *fadL*, and this is likely by providing NTHi resistance to the bactericidal effect of those free fatty acids.

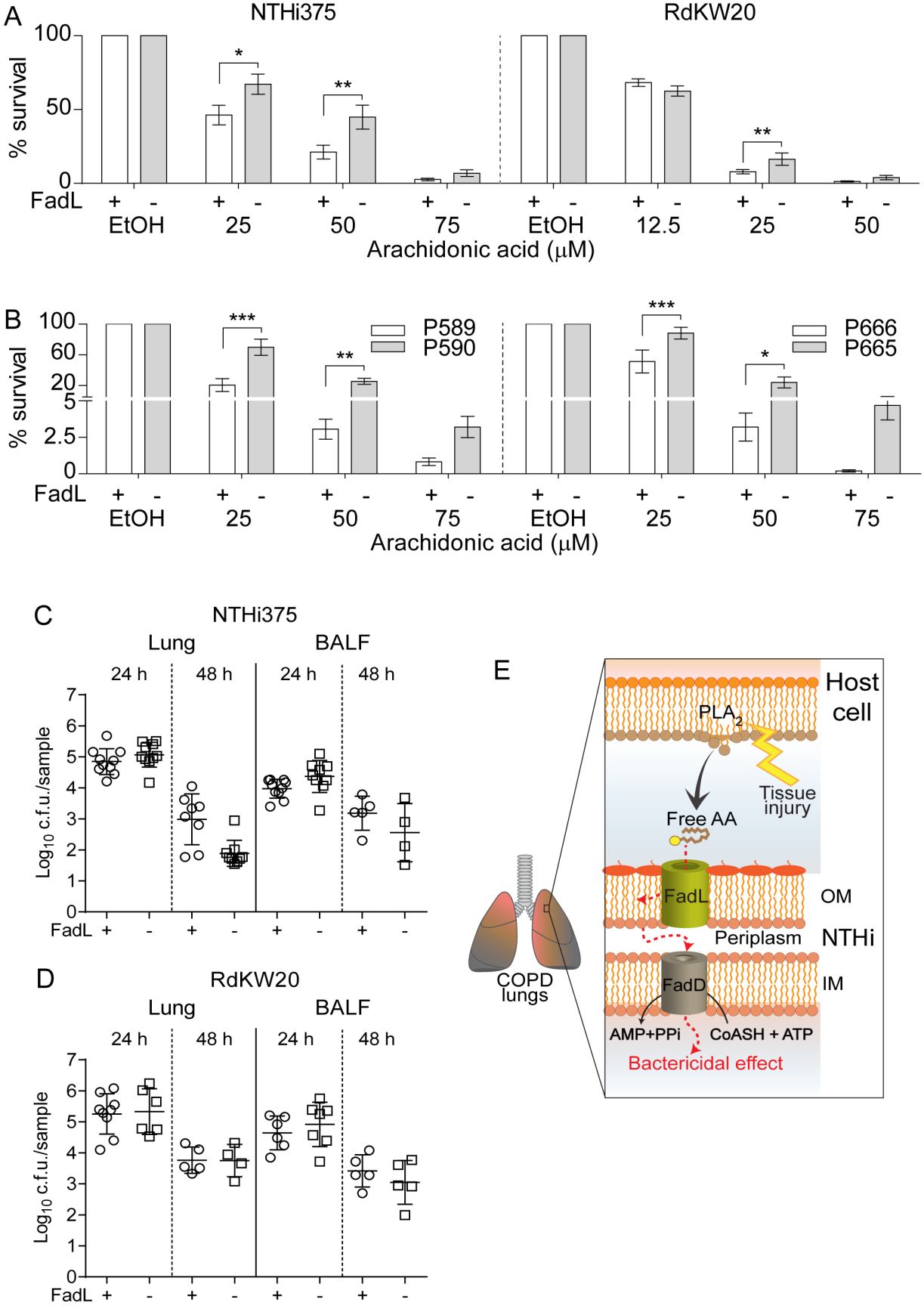


Table 3.5. Coding region SNVs present in strains belonging to CTs 9 (isolates P590 and P589) **and 76** (isolates P665 and P666)**.**

		CT9									
nt REF	**nt ALT**	**P590**	**P589**	**Mutation type**	**Impact**	**Gene name**	**aa REF**	**aa ALT**	**aa position**	**Protein length**	**Function**
A	C	0	1	missense	MODERATE	L_00407	Ser	Arg	22	470	Putative ATPase, hypothetical protein
A	C	0	1	missense	MODERATE		Asn	Thr	282	470	
G	T	0	1	missense	MODERATE	L_00640	Pro	Thr	513	844	Translation initiation factor IF-2
A	C	0	1	stop lost & splice region	HIGH	L_00728	Ter	Ser	229	228	Outer membrane protein P1 precursor, long-chain fatty acid outer membrane transporter
T	C	0	1	missense	MODERATE	L_01000	Asn	Asp	59	137	Phage GP46 family protein, hypothetical protein
A	G	0	1	missense	MODERATE	L_01011	Ser	Pro	91	101	Putative uncharacterised protein, putative protein
C	A	0	1	missense	MODERATE	L_01014	Glu	Asp	131	161	Phage virion morphogenesis protein
T	C	0	1	missense	MODERATE	L_01016	Asn	Asp	521	546	Putative uncharacterized protein, hypothetical protein, hypothetical protein
TG	CA	0	1	missense	MODERATE		Lys	Glu	514	546	

		CT76									
nt REF	**nt ALT**	**P665**	**P666**	**Mutation type**	**Impact**	**Gene name**	**aa REF**	**aa ALT**	**aa position**	**Protein length**	**Function**
AGGA	AA	0	1	frameshift	HIGH	L_00010	Gly	fs	140	258	Hypothetical protein
A	C	0	1	missense	MODERATE	L_00404	Lys	Thr	413	1046	Outer membrane receptor protein, putative Fe transport, hemoglobin/hemoglobin-haptoglobin binding protein B
CT	CTCGAT	0	1	frameshift & stop lost & splice region	HIGH	L_00626	Ter	fs	168	167	Outer membrane protein P1 precursor, long-chain fatty acid outer membrane transporter
T	C	0	1	non coding transcript	MODIFIER	L_00890					rRNA
A	C	0	1	missense	MODERATE	L_00973	Tyr	Asp	345	366	Outer membrane protein P2 precursor
C	T	0	1	missense	MODERATE		Gly	Ser	93	366	

A C	C T	0 0	1 1	missense missense	MODERATE MODERATE	L_01164	Gln Pro	Pro Leu	288 402	704 704	Guanosine-3', 5'-bis 3'-pyrophosphohydrolase
T A	A T	0 0	1 1	non coding transcript non coding transcript	MODIFIER MODIFIER	L_01171					rRNA
T	G	0	1	missense	MODERATE	L_01757	Ser	Ala	322	366	Histidinol-phosphate aminotransferase 2

fs.: frameshift

3.4.10. FadL inactivation has no effect in a murine lung acute infection model

A previously used NTHi mouse respiratory infection model (Euba et al., 2015a; Euba et al., 2015b; Euba et al., 2015c; Euba et al., 2017) rendered a fatty acyl residues profile including AA (GC-MS analysis of lung fatty acid methyl ester composition, Fernández-Calvet et al. submitted), and human-restricted bacterial pathogens are known to selectively interact with hCEACAM1 (Voges et al., 2010). NTHi murine pulmonary infection may uncouple FadL functional analysis as a LCFA transporter and a hCEACAM1 ligand, the latter being absent from the mouse. CD1 mice were next infected with NTHi375 and RdKW20, WT and *fadL* mutant strains, and bacterial loads quantified in lung and bronchoalveolar lavage fluid (BALF) samples. Bacterial counts were comparable for each WT and mutant strain pair (**Figure 3.9C** and **3.9D**), showing that *fadL* gene inactivation gives no substantial benefit in this infection model. We acknowledge that this model system may be suboptimal in this case since NTHi fails to establish a long-term infection, showing only variation in time-to-clearance, which occurs in <3 days for the WT strains.

3.4.11. Truncated *fadL* alleles are rare in most NTHi but common in isolates from lower airways infections

Because the subjects in this study all had COPD for years prior to sampling NTHi, and because we do not know the relative abundance of different strains and sub-strains in these patients over time, we could not to conclusively demonstrate that null alleles of *fadL* arose within the COPD lung, despite 5 of 20 multi-isolate CTs being polymorphic for null alleles. To further test whether loss of *fadL* was specifically associated with lower airways infections, we examined the distribution of *fadL* annotations across 186 publicly available genome assemblies derived from isolates of diverse origin (**Figure S3.5, Table S3.9**, (De Chiara et al., 2014; Price et al., 2015; Staples et al., 2017)). Of 159 genomes isolated collected outside the context of lower respiratory infection (including nasopharyngeal carriage, middle ear from pediatric otitis media, invasive bacteremia and meningitis isolates), only three isolates lacked a single full-length *fadL* annotation (1.9%, one from otitis media, one from bacteremia, and one other). By contrast, out of 27 isolates from the lower airways (isolates from sputum of COPD or other pulmonary infections), 9 had truncated *fadL* genes (33.3%). This shows that an intact *fadL* gene is the norm among NTHi as a whole, but loss-of-function mutations are common for lung isolates (Fisher's exact test p-value = 1.3×10^{-6}). Including the COPD isolates from this study gave consistent results: only 3 of 159 non-lung isolates lacked a predicted full-

length *fadL*, whereas 26 of 119 lung/sputum isolates lacked it (21.8%, Fisher's exact test p-value = $7.2x10^{-8}$). These results suggest that intact *fadL* function is usually beneficial for *H. influenzae*, except in the context of lower respiratory infections when intact *fadL* function becomes deleterious.

3.5. Discussion

Our comparative genomic analyses of longitudinally collected NTHi isolates from COPD patients has provided insights into the evolutionary dynamics of a bacterium as it adapts from being a nasopharyngeal commensal to become a chronically infecting opportunistic pathogen, as well as revealed novel therapeutic approaches for this and other COPD infections. To our knowledge, this is the first genome-scale analysis of longitudinally collected bacterial isolates from COPD patients, as well as the largest and most diverse longitudinal collection of genomes from a human respiratory opportunistic pathogen of commensal instead of environmental origin.

We acknowledge that our strain collection presented several limitations. First, we are unlikely to have sampled the initial colonizing bacterial isolates, therefore we were unable to polarize genetic changes to a recent common ancestor and/or initial period of rapid adaptation (Yang et al., 2011). Second, we only analyzed single isolates from each sampling point. Evidence in CF patients on environmental pathogens suggests the existence of multiple divergent lineages or sub-lineages of chronically infecting pathogens that are present simultaneously (Chung et al., 2017; Jorth et al., 2015; Lieberman et al., 2014; Lieberman et al., 2016; Markussen et al., 2014). By contrast, previous studies of NTHi used low-resolution genotyping methods (MLST and PFGE) to show its much higher polyclonality in both health and disease, within single subjects over time or co-existing at the same time (Berrens et al., 2007; Farjo et al., 2004; Hiltke et al., 2003; Lacross et al., 2008; Murphy et al., 1999; Sethi et al., 2002).

However, the low-resolution genotyping methods used in studies defining NTHi's polyclonality are insufficient for identifying intra-clonal genomic variation, whether caused by mutation or recombination with other isolates. We previously conducted a limited genomic analysis of three COPD isolates collected from the same patient, with identical PFGE typing and behavior in all but one of a series of phenotypic assays. The later isolates had increased resistance to AMPs and—among only a few SNVs—also carried non-conserved amino acid replacement substitutions in the SapABCDZ, known to confer AMP resistance to NTHi (Garmendia et al., 2014; Mason et al., 2005). Following the

notion that intra-clonal variation coupled to phenotyping may reveal NTHi genes undergoing adaptation to the COPD lung, we examined a larger set of longitudinally collected COPD isolates for intra-clonal polymorphisms. Yet we still found evidence for parallel molecular evolution in several genes, in which several CTs were polymorphic for variant alleles predicted to affect protein function. Among other interesting candidate patho-adaptive genes, the *ompP1/fadL* locus stood out because of the high frequency of independent loss-of-function alleles seen in different CTs and different patients. High amino acid variation in the FadL protein has been reported in *H. influenzae* and *Neisseria meningitidis* (Mes & van Putten, 2007; Tchoupa et al., 2015; Yero et al., 2010), in agreement with reports on the key role of OMP for within-host adaptive strategies (Cullen & McClean, 2015; Wu et al., 2016).

Null mutations seen in *fadL* within COPD CTs suggest a strong selective pressure on this locus to be lost during adaptation to the COPD lung, consistent with NTHi's reductive evolution towards being a long-term resident of the lower airways (Andersson & Kurland, 1998). In support of this, our survey of nearly 200 publically available *H. influenzae* genomes found that putative null mutations in *fadL* were rare in isolates collected outside the context of lower respiratory infections, compared to their high frequency in isolates from lung infections. Given that the $FadL_{NTHi}$/hCEACAM1 interaction was known (Tchoupa et al., 2015), and supported by our new 3D models, we were initially puzzled that losing this interaction could be beneficial to bacteria in COPD. But finding that loss of FadL gave resistance to microbicidal fatty acids suggested instead a case of antagonistic pleiotropy, wherein a single gene controls multiple phenotypes, some of which may be beneficial and others deleterious depending on conditions. Antagonistic pleiotropy can be modulated by gene inactivation, such that loss-of-function mutants can outcompete the WT parental strain once a trait becomes deleterious (Bliven & Maurelli, 2016). This may be the case for *fadL*, encoding a bifunctional protein that acts as both a ligand for a human cell surface receptor and a free fatty acid uptake system, where loss of the gene avoids triggering bacterial cell death by the combined fatty acid detergent effect and NTHi inability to metabolize this type of molecule (lacking β-oxidation). Independently, *fadL* frameshift mutations have also been shown to confer resistance to the antibacterial compound A-344583 (Lerner et al., 2005). Other explanations for recurrent *fadL* loss-of-function mutations are possible: as a surface protein, FadL may be subjected to immune system pressure (as potentially indicated by the increased variability in the exposed surface loops) and so cells may benefit from its loss, if other functions are not compromised. This would require either that immune recognition of FadL is distinct in the

COPD lung compared to other environments, or that normal benefits of expressing FadL are absent in the COPD lung.

Several lipids are molecular mediators of respiratory disease (Petrache & Petrusca, 2013). In particular, AA metabolites are key players in COPD airway inflammation (Malhotra et al., 2012). Phospholipase A_2 (PLA_2)-catalyzed hydrolysis of membrane phospholipids results in production of free fatty acids, most importantly AA, which serves as precursor for inflammatory mediators such as platelet-activating factor (PAF) (Meyer et al., 2005). In turn, lipoprotein-associated PLA_2 mediates PAF hydrolysis and dampens PAF-mediated inflammation (McIntyre et al., 2009). *H. influenzae* exploits molecular mimicry to evade host inflammation through GlpQ, an enzyme that hydrolyzes phosphorylcholine (ChoP) moieties shared by both PAF and the bacterial LOS molecule, therefore removing PAF from the airways and suppressing inflammation. Such bacterially driven regulation of PAF signaling could be overcome by elevating PAF levels at the site of infection through PLA_2 targeting (Hergott et al., 2015) which, in turn, may also dampen free AA release at the COPD lung, therefore limiting FadL negative selection during NTHi patho-adaptation (**Figure 3.9E**). The PLA_2 antagonist darapladib has been shown to be safe in humans (Investigators, 2014).

Collectively, our results bring to light the utility of comparing the genomes of bacterial isolates collected over time from long-term chronic infections. They further show the existence of genetic polymorphisms within clonally related NTHi isolates as a result of point mutations and recombination events likely mediated by natural transformation, and this intra-clonal variation can help tease apart transmission of strains between subjects and genetic changes resulting in changes in pathogenic traits. Parallel changes affecting protein function were seen for a small number of genes, and follow-up studies support a bifunctional role for the FadL protein in both interaction with host epithelia and uptake of free fatty acids. We conclude that the environment of the COPD lung exerts a specific selection pressure to lose functional *fadL*, and this is likely by providing resistance to bactericidal fatty acids, in spite of gene truncation also compromising the ability of NTHi to colonize or invade epithelial cell surfaces. We speculated that FadL-hCEACAM1 binding may be essential for infection establishment but, once established, persistence may favor loss of FadL to overcome free fatty acids pressure. This hypothesis aligns with observed low hCEACAM1 expression in lung tissue (Fagerberg et al., 2014). Finally, we point out PLA_2 as a potential target of evolutionary medicine for host-directed therapeutics against NTHi chronic infections in the setting of COPD.

3.6. Supplementary Material

3.6.1. Supplementary Methods

DNA extractions. A single colony from each clinical isolate was collected from chocolate-agar and inoculated into 3 ml sBHi cultures. DNA was extracted after overnight growth with the DNeasy Blood and Tissue Kit (Qiagen, Cat. 69506) following manufacturer´s instructions. Purity, quality and concentration were measured by Nanodrop, agarose gel electrophoresis, and Qbit fluorometry.

Illlumina sequencing and assembly. Nextera XT (Illumina) libraries were produced following manufacturer´s recommendations, and paired-end sequencing (2x151nt) was conducted on an Illumina NextSeq 500. Raw base call data (bcl) was converted into FastQ format (Illumina version 1.8) using the bcl2fastq conversion software (Illumina, version 1.8.3, setting-no-eamss). *De novo* assembly of Illumina datasets used read pairs processed as follows: (1) reads were trimmed of adapters with Trimmomatic (v0.33) (Bolger et al., 2014); (2) overlapping read pairs were merged with COPE (Liu et al., 2012); (3) reads were error corrected using ErrorCorrectReads.pl from allpathlg and (4) finally, reads were assembled with SPAdes (Bankevich et al., 2012).

Pacific Biosciences sequencing and assembly. SMRTbell libraries with mean insert size of ~6 kb were prepared according to manufacturer's specifications. PacBio sequencing (v 2.1.0) was performed with 1 to 2 genomes per SMRTcell with P4-C2 chemistry. *De novo* assembly of PacBio data was done using SMRT pipe (v 2.3.0). SMRT cells with two genomes were demultiplexed based on barcodes. Initial assembly with the HGAP assembler yielded single contigs with mean coverage from 54 to 66-fold. Genome were circularized and and permuted to the *dnaA* gene using Circlator (1.0.2), followed by Quiver-based error correction (Chin et al., 2013) and identification of DNA modification motifs using RS Modification and Motif Analysis.

Reference assemblies from public databases. *H. influenzae* assemblies available at NCBI were downloaded on June 23, 2016, and these were reconciled with a partially overlapping set of genomes available at the Sanger Institute (http://www.pnas.org/content/111/14/5439), keeping the assembly with the fewest contigs. In total 105 assemblies from NCBI and 81 assemblies from the Sanger Institute were used in this analysis. Assemblies were taxonomically classified using Taxator-tk, which found that several genomes from NCBI were more accurately classified as *H. haemolyticus* or *H. parainfluenzae*. Strain provenance was encoded as "lower

respiratory" if isolated from a sputum sample and annotated as from either COPD or pneumonia (**Table S3.9**).

Homologous gene clustering. Roary v3.5.1 was used to cluster homologous protein-coding genes from across the full set of assemblies and annotations (**Methods**). After testing blastp thresholds from 50 to 95% in increments of 5%, 75% amino acid identity was chosen for clustering because it was the highest threshold before the rapid increase in the number of rare genes. Contig breaks and chromosomal rearrangements cause Roary to sometimes over-split clusters into separate paralogous clusters based on synteny in adjacent genes, so following Lee *et al*. (Lee et al., 2017), we generated both "split" and "merged" gene possession tables; analyses presented all used "merged" gene possession tables, in which paralog splitting was reversed to generate homologous sets. Included in this set were genomes from outgroup sister species *H. haemolyticus* and *H. parainfluenzae* to provide outgroups. Thus, for clonal typing using goeBurst, a set of 478 protein-coding gene clusters were used, since these were present once and only once in all assemblies.

Clonal typing. For each single copy core gene, unique nucleotide sequences were defined as distinct alleles, and strings of allele IDs for each isolate were used as MLST-like input. The goeBurst full MST algorithm was used to produce an MST where the edge weights were distances between strains. A clustering threshold was chosen such that phylogenetically related strains clustered together, so edges with weights >25 were dropped from the graph and each resulting connected component was assigned an arbitrary CT number.

Knockouts of *fadL*. To disrupt the *fadL* gene, a DNA fragment containing the *fadL* gene and its respective flanks (2,222 bp in total) was PCR amplified with *Phusion* DNA Polymerase (Thermofisher) using genomic DNA from NTHi375 as a template and primers fadL-F1/1302 and fadL-R1/1303 (**Table S3.6**). The gene containing PCR product was cloned into pJET1.2/blunt (Thermofisher), generating pJET1.2-*fadL*. A Spc^r cassette was PCR amplified from pRSM2832 using gene-specific mutagenic primers fadL-F2/1358 and fadL-R2/1359 (Tracy et al., 2008). Primers were designed to delete sequences between the start codon and the last 7 codons of *fadL*. *E. coli* SW102 cells were prepared for recombineering, co-electroporated with pJET1.2-*fadL* (Amp^r) (50 ng) and the *fadL*-specific mutagenic cassette (Spc^r) (200 ng), as previously described (Sinha et al., 2012). Mutagenized clones containing pJET1.2-*fadL::spc* were selected on LB agar with Amp_{100}, Spc_{50}. This plasmid was used as template to amplify the *fadL::spc*

disruption cassette with primers fadL-F1/1302 and fadL-R1/1303, which was used to transform NTHi375 and RdKW20 using the MIV method. Transformants were selected on sBHI agar with Spc_{30}, to obtain NTHi375Δ*fadL* and RdKW20Δ*fadL* mutant strains. Disruption was confirmed by PCR.

Cell culture and bacterial infection/invasion assays. For infection, PBS-normalized bacterial suspensions (OD_{600}=1) were prepared by using NTHi strains grown on chocolate agar. A multiplicity of infection (MOI) of ~100:1 was used. To monitor invasion, cells were incubated with bacteria for 2 h in 1 ml EBSS (Earle's Balanced Salt Solution, Gibco), washed 3 times with PBS, incubated for 1 h with RPMI 1640 medium containing 10% FCS, Hepes 10 mM, and gentamicin 200 µg/ml to kill extracellular bacteria. Cells were washed 3 times with PBS, lysed with 300 µl PBS-saponin 0.025% for 10 min at room temperature, and serial dilutions plated on sBHI agar. Infections were performed in triplicate at least three independent times ($n \geq 9$). Results are expressed as c.f.u./well.

RNA extraction and real-time quantitative PCR (RT-qPCR) analysis. NTHi strains were grown on chocolate agar. Bacteria (2 to 5 colonies) were inoculated into 20 ml sBHI, grown for 11 h, diluted into 40 ml fresh sBHI to OD_{600}=0.05 and grown to OD_{600}=0.6. Bacterial total RNA was isolated using TRIzol reagent (Invitrogen). Total RNA quality was evaluated using RNA 6000 Nano LabChips (Agilent 2100 Bioanalyzer). All samples had intact 16S and 23S ribosomal RNA. Complementary DNA (cDNA) was synthesized from total RNA (1 µg) using PrimerScript RT Reagent kit (Takara). Real-time quantitative PCR was performed using 1X SYBR Premix Ex Taq II (Tli RNaseH Plus, Takara) and primer mix; fluorescence data were analyzed with AriaMx Real-Time PCR System (Agilent Technologies). Relative quantities of mRNAs were calculated using the comparative threshold cycle (Ct) method and normalized using the *gyrA* gene as an endogenous control for each strain. Specific primers were designed with Primer Express software: for strain, P589-8275, fadL-qPCR-F3/1659 and fadL-qPCR-R1/1607; for P600, fadL-qPCR-F1/1605 and fadL-qPCR-R2/1608; for P608, fadL-qPCR-F2/1606 and fadL-qPCR-R1/1607 (**Table S3.6**); for strains NTHi375, Hi RdKW20, P634 and P666, fadL-qPCR-F1/1605 and fadL-qPCR-R1/1607. Measures were performed in duplicate and at least three times ($n \geq 6$).

Fatty acid susceptibility testing. NTHi strains grown on chocolate agar for 16 h were resuspended to OD_{600}=0.1 in a defined minimal medium free of fatty acids (MM-FFA), consisting of MIV solutions S21, S22, S23 and S24 (100:1:1:1) (Herriott et al., 1970), 10 µg/ml hemin, and 10 µg/ml NAD. As needed, MM-FFA was supplemented with 20 mM

glucose, vehicle control (96% ethanol), or free fatty acids (dissolved in ethanol to 100 mM stocks) at varying concentrations (OA, from 0.8 to 2 mM; AA, from 12.5 to 75 μM) (Sigma-Aldrich). OA and AA stock solutions (100 mM) were prepared and diluted to the required working concentrations in ethanol. MM-FFA with fatty acid (160 μl) were transferred to individual wells in 96-well microtiter plates (Sarstedt); 40 μl of the previously prepared bacterial suspensions were added to each well, and incubated for 20 h at 37ºC with 5% CO_2 in static conditions. Vehicle solution, consisting of an ethanol volume equivalent to that used for the highest fatty acid concentration tested, and controls with MM-FFA only were performed in parallel. After incubation, bacteria were serially ten-fold diluted in PBS and plated on sBHI agar. Results are expressed as percentage of bacterial survival ([c.f.u.ml$^{-1}_{LCFA}$/c.f.u.ml$^{-1}_{vehicle}$] x 100). Experiments were performed in triplicate on at least four independent occasions ($n \geq 12$).

NTHi mouse lung infection. CD1 female mice (18-20 g) aged 4 to 5 weeks were purchased from Charles River Laboratories (France), housed under pathogen-free conditions at the Institute of Agrobiotechnology facilities (registration number ES/31-2016-000002-CR-SU-US), and used at 22-25 g. Animal handling and procedures were in accordance with the current European (Directive 86/609/EEC) and National (Real Decreto 53/2013) legislations, following the FELASA and ARRIVE guidelines, and with the approval of the Universidad Pública de Navarra (UPNa) Animal Experimentation Committee (Comité de Ética, Experimentación Animal y Bioseguridad) and the local Government authorization. NTHi375 and RdKW20, WT and Δ*fadL* strains, were used for lung infection. Mice were randomly divided into groups (n=5). Infecting bacteria were previously grown on chocolate agar. For intranasal infection, 20 μl of a NTHi suspension containing ~2×10^8 c.f.u. were placed at the entrance of the nostrils of each mouse until complete inhalation, in mice previously anesthetized with ketamine-xylacine (3:1). At 24 hpi, mice were euthanized using cervical dislocation. Lungs were homogenized with PBS (1:10 w/v), serially ten-fold diluted in PBS and plated on sBHi agar for viable bacterial counts. Results are expressed as mean±SD of individual $\log_{10}$ c.f.u./lung. BALF samples were obtained by perfusion and collection of 0.7 ml of PBS, with help of a sterile 20G (1.1 mm diameter) Vialon™ intravenous catheter (Becton-Dickinson) inserted into the trachea. An aliquot of each recovered BALF was serially ten-fold diluted in PBS, and plated on sBHI agar to determine the number of viable bacteria. Results are expressed as mean±SD of individual $\log_{10}$ c.f.u./ml BALF.

Molecular modeling methods

Molecular dynamics simulations

We assessed the stability of the homology models, the mobility of the loops and the stability of the binding poses, by means of MD simulations as implemented in AMBER 14 suite of programs (Case, 2014). Missing hydrogen atoms were added and protonation state of ionisable groups was computed by using Maestro Protein Preparation Wizard (Schrodinger, 2012a; Schrodinger, 2012b; Schrodinger, 2012c; Schrodinger, 2012d). Atom types and charges were assigned according to AMBER ff10 force field. The homology models, together with the structure from PDB 1IT16, were hydrated by using cubic boxes containing explicit TIP3P water molecules extending 10 Å away from any protein atom for simulating the aqueous environment with the help of AmberTools (Case, 2014) with added counter ions to neutralize the system. Minimization was performed using Sander and MD simulations were run using the pmemd, which are distributed within the AMBER 14 package. A 1 fs integration step and the shake algorithm on every hydrogen-containing bond. The smooth particle mesh Ewald method was used to represent the electrostatic attractions in the system while each simulation was under periodic boundary conditions, and the grid spacing was 1 Å. Initial annealing of the system occurred steadily and lightly from 100°K to 300°K over 25 ps. Temperature was kept constant at 300°K during 50 ps with progressive energy minimizations and also a solute restraint. The restraints were gradually released by the solute, which was closely followed by a 20 ps heating period which went from 100°K to 300°K, once completed the restraints were removed. Each of the simulations lasted 10 ns. The systems then advanced in an isothermal-isobaric ensemble. Long-range electrostatic interactions were accounted for by means of the particle mesh Ewald approach applying periodic boundary conditions. The root mean square deviation (RMSD) as a function of time with respect to the starting structure for the Cα atoms was computed using CPPTRAJ.

Normal mode analysis

To compute the low frequency normal modes of $FadL_{NTHi}$ models, the elastic network model was used with the Web interface ElNemo (http://www.sciences.un-nantes.fr/elnemo/). The starting geometries were those from the MD simulations. From the resulting normal mode (NM) conformations, we selected three structures for docking purposes: the one corresponding to the crystallographic structure (NM-6), the one corresponding to the last structure of the displacement in one of the directions of the normal mode (NM-1), and the last located structure in the opposite direction (NM-11).

The three structures were submitted to 100 000 steps of steepest descent minimization with MacroModel and optimized with AMBER force field, before being used for docking calculations.

Ligand-protein docking calculations

Preparation of the macromolecules. 3D coordinates of hCEACAM1 and the FadL homology models were used for docking purposes. Quality of the minimized models was verified using the PDBSUM online resource, which displays Ramachandran plot of the structures. Models were prepared for docking calculations by adding Kollman charges with the help of AutoDockTools (Case, 2014).

Building of ligands. The 3D structure of LDA was extracted from the crystallographic structure PDB-ID 1T16. The 3D coordinates of AA and OA were built with Maestro (Schrodinger, 2012a; Schrodinger, 2012b; Schrodinger, 2012c; Schrodinger, 2012d). The geometries of the three ligands were optimized with MMFFs force field by using MacroModel. Ligands were prepared for docking calculations using AutoDockTools, setting all rotatable bonds free to move during the docking calculation.

Docking calculations. Docking calculations of all compounds were performed by using AutoDock 4.2.2. (Morris, 1998). Analysis was performed with AutoDockTools. The grid point spacing was set at 0.375 Å, and a hexahedral box was built with x, y, z dimensions FadL: 122 Å, 108 Å, 92 Å centered in the centroid position among residues Arg161, Lys336 and Ala378 for $FadL_{NTHi375,}$ Arg194, Lys343 and Ala385 for $FadL_{RdKW20}$ and Arg190, Lys339 and Ala381 for $FadL_{P608}$. A total of 200 runs using Lamarckian Genetic algorithm was performed, with a population size of 100, and 250000 energy evaluations.

Protein-protein docking calculations

Protein-protein docking was performed with ZDOCK Server (Pierce et al., 2014), between each of the FadL models ($FadL_{RdKW20}$, $FadL_{NTHi375}$ and $FadL_{P608}$) and the *h*CEACAM1 structure. The FadL structures were previously modeled by us as explained above. The hCEACAM1 structure was obtained from the X-ray crystallographic structure (PDB-ID: 4QXW), refined (missing atoms were added, and all the structure was minimized) and optimized by MD simulations. Protein-protein docking was performed and the different bound complexed were further optimized by MD simulations.

Statistical analysis. For cell infection and bacterial loads in lungs and BALF samples, mean values were calculated and statistical comparison of means performed using the two-tail *t* test. A two-way ANOVA was used for statistical comparison, within rows, of

column mean (Tukey's multiple comparisons test) for fatty acid bactericidal assays. Mean±SEM is represented and a $p<0.05$ value was considered statistically significant. Analyses were performed using Prism software, version 7 for Mac (GraphPad Software) statistical package.

3.6.2. Supplementary Results

Computational Modeling

Homology modeling

Three $FadL_{NTHi}$ variants, $FadL_{NTHi375}$ $FadL_{RdKW20}$ and $FadL_{P608}$, were considered for computational 3D structures calculation by homology modeling. FadL coding sequences of strains RdKW20 and NTHi375 were extracted from NCBI (http://www.ncbi.nlm.nih.gov). FadL from *E. coli* (PDB-ID: 1T16) was used as a template, the alignment of the sequences is shown at **Figure S3.6**. For each sequence, four homology models were generated using four different servers: SWISS-MODEL (Arnold et al., 2006; Biasini et al., 2014; Guex et al., 2009; Kiefer et al., 2009), I-TASSER (Roy et al., 2010; Yang & Zhang, 2015; Zhang, 2008), PHYRE2 (Kelley et al., 2015) and RAPTOR X (Kallberg et al., 2012). The models derived from I-TASSER, PHYRE-2, and RAPTOR X were discarded because the structures did not have coherence with other experimental 3D structures from homologous proteins. The 3D structures of the computed structures are very similar between them, and also in comparison to *E. coli* FadL. Detailed analysis was performed, focusing on the inner part of the β-barrel and the flaking loops. Regarding the interior of the β-barrel, some different sequence amino acids in comparison to *E. coli* FadL were identified, pointing to a putative role in the recognition of the FA ligands (**Figure S3.7**).

Thus, computational studies were performed using the 3D structures obtained from SWISS-MODEL, which were submitted to molecular dynamics (MD) simulations to optimize the geometry. The three modeled structures showed high stability (**Figure S3.8**). Normal mode analysis (NMA) was also performed to study the flexibility of the loops and the three models exhibited high coherence among them thus confirming the stability (**Figure S3.9**).

Docking of FadI and hCEACAM1

We have performed protein-protein docking with ZDOCK Server (Pierce et al., 2014), between each of the $FadL_{NTHi375,}$ $FadL_{RdKW20}$, and $FadL_{P608}$ models with hCEACAM1 (PDB-ID: 4QXW). The resulting protein-protein complexes were submitted to 100ns MD simulations being possible to observe high stability for the two complexes along the simulation (**Figure S3.10**).

Protein-protein interaction analysis of the $FadL_{NTHi375}$/hCEACAM1 complex identified relevant polar interactions between Lys213($FadL_{NTHi375}$)-Glu99(hCEACAM1), Ala334-

Gln27, Ser276-Gly51, and also hydrophobic interactions between Ile216-Ile91. These interactions remained stable along the MD simulation. In the case of Lys213($FadL_{NTHi375}$), new interactions were established with residues Gln89 and Asn97 from *h*CEACAM1, strengthening the protein-protein complex. Also remain stable along the MD simulation the CH-π interaction between Tyr34-Val39 both from hCEACAM1, hydrogen bond between Gln44 and OH from Tyr34, Gln44 with O from Ser32, Gln89-Ser32 and Gln89-Tyr34. The MD simulation of the $FadL_{RdKW20}$/*h*CEACAM1 complex also led to a stable protein-protein complex where the main interactions were polar contacts between Lys225($FadL_{RdKW20}$)-Asp40(*h*CEACAM1), Phe222-Gly41, Lys290-Ser93 and Lys339-Asp94, and hydrophobic interactions between Ile286-Ile91 side chains. Also remain stable along the MD simulation the CH-π interaction between Tyr34-Val39 both from *h*CEACAM1, hydrogen bond between Gln44 and OH from Tyr34, Gln44 with O from Ser32, and Gln89-Tyr34. Protein-protein interaction analysis of the $FadL_{P608}$/hCEACAM1 complex identified relevant polar interactions between Lys216($FadL_{P608}$)-Gln89(hCEACAM1), Lys217-Glu99, Asn200-Gln1, Gly193-Asp94, Asp332-Tyr48 and Gly279-Gly51; also CH-π interaction between Trp220-Ile91 and hydrophobic interactions between Val196-Val96 and Leu214-Val39. These interactions remained stable along the MD simulation. Also remain stable along the MD simulation the CH-π interaction between Tyr34-Val39 both from hCEACAM1, hydrogen bond between Gln44 and OH from Tyr34, Gln44 with O from Ser32, and Gln89-Tyr34. Interestingly, the residues from *h*CEACAM1 (residues Ser32, Tyr34, Val39, Gln44, Gln89 and Ile91) that interact with FadL were previously reported to be crucial for the interaction with pathogen components (Villullas et al., 2007). Our computational studies clearly identify these particular residues as directly involved in the interaction with FadL, suggesting they shape a region relevant for the molecular recognition.

Docking calculations of AA, LDA and OA

Docking calculations of AA, OA and LDA were performed in the three modeled structures $FAdL_{NTHI375}$ and $FadL_{RdKW20}$, and $FadL_{P608}$. Results of the docking studies of AA inside $FAdL_{NTHI375}$ and $FadL_{RdKW20}$ are reported in the manuscript main text. We here detail the description of the different predicted binding poses and the interactions. Calculations predicted two AA binding modes with good theoretical binding energies for the $FadL_{NTHi375}$ model. One binding mode places AA at the entrance of the pocket, in a hydrophobic groove located between L3 and L4 (site A), where AA is bound and solvent accessible. Interactions with polar residues, specifically with Lys336 (L5), lipophilic interactions with Ile278, and CH-π interactions with Phe150 and Phe334 side chains

were identified. A second binding mode for AA was predicted in the $FadL_{NTHi375}$ model at the deep of the β-barrel, establishing polar contact with Lys142 present in L2 (site B), which has a kink that points inward. Interestingly, these binding poses were also found for $FadL_{RdKW20}$ model. Docking of AA was also performed inside the $FadL_{P608}$ structure. In this case, apart from the docked poses in the binding sites A and B, an additional binding pose was found in a third binding site (site C) within a hydrophobic pocket of the β-barrel, where the carboxylate group establishes a hydrogen bond with the Lys339, and hydrophobic interactions are established with Phe135, Leu341 and Ile382 side chains (**Figure S3.11**).

We also performed docking calculations with LDA in three modeled structures $FadL_{NTHi375,}$ $FadL_{RdKW20}$, and $FadL_{P608}$. For the $FadL_{NTHi375}$ model, two LDA binding modes were predicted with good theoretical binding energies. One binding mode places LDA at the entrance of the pocket, in a hydrophobic groove located between L3 and L4 (site A), where LDA is bound and solvent accessible. Interactions with polar residues, specifically with Lys336 (L5), lipophilic interactions with the hydrophobic pocket delimited by Ile149, Ile195, Val227, Leu273 and Ile278, and CH-π interactions with Phe150 were identified (**Figure S3.12**). A second binding mode for LDA was predicted in the $FadL_{NTHi375}$ model at the deep of the β-barrel, establishing polar contact with Lys138 present in L2 (site B), which has a kink that points inward (**Figure S3.12**). This binding pose was also found for LDA in the $FadL_{RdKW20}$ model. In that case, an additional binding pose was found in a third binding site (site C) within a hydrophobic pocket of the β-barrel, where the N-oxide group of LDA establishes a hydrogen bond polar interaction with the Asp238 NH group, and hydrophobic interactions are established with Phe25, Leu160, Val192, Leu236 and Leu295 side chains (**Figure S3.12**). These three binding poses (sites A, B and C) were found also in the case of $FadL_{P608}$. Our docked poses for LDA in the FadL models and the X-ray crystallographic pose structure (PDB-ID: 1T16) present some differences. LDA binds in two sites in the $FadL_{E.\ coli}$ crystal complex, at the hydrophobic pocket delimited by the positively charged residues Arg182 and Lys342, and another site located close to Arg391, where LDA establishes polar interactions. None of the docked solutions obtained for the FadL models were predicted to bind to these sites.

We were also prompted to perform docking calculations of OA in the 3D models of $FadL_{NTHi375}$ and $FadL_{RdKW20}$, leading to predicted binding poses at sites A and B (**Figure S3.13**). At site C we did not find any docked pose. The main interactions involved polar

residues between L3 and L4 (site A), such as Lys336 (for $FadL_{NTHi375}$) and Lys343 (for $FadL_{RdKW20}$), and polar interactions with Lys138 (for $FadL_{NTHi375}$) and Lys142 (for $FadL_{RdKW20}$) and present in L2 (site B). Overall, the docked poses for OA, at both sites A and B, were very similar to those obtained for LDA. However, docking calculations of OA in the 3D model $FadL_{P608}$ led to predicted binding poses also at site C within a hydrophobic pocket of the β-barrel involved main polar interactions with Lys339 (for $FadL_{P608}$).

3.6.3. Supplementary Figures

Figure S3.1. PacBio vs. Illumina. Each of the 17 PacBio genomes were also sequenced on Illumina and assembled as described in methods, so the 17 pairs of assemblies were compared. (**a**) Each subplot compares the feature counts for the annotated genomes. Strains are on the x axis, feature counts are on the y axis, Illumina genomes are indicated by green dots, and PacBio genomes are indicated by blue dots. (**b**) Roary computed the pangenome for the 17 pairs of assemblies as described in the methods. In each subgraph, the x axis is the number of strain pairs and the y axis is the number of gene clusters from roary. (1) only_present_in_illumina – genes that were present in the Illumina assembly but not the PacBio assembly. Most genes were present in both assemblies as indicated by the bar for zero strain pairs. (2) only_present_in_pacbio – genes that were present in the PacBio assembly but not the Illumina assembly. (3) pacbio_has_more_copies – in many cases, multiple genes from a genome belong to the same gene cluster. This subplot counts the number of strains where the PacBio assembly had a higher copy number for the gene cluster than the Illumina assembly. (4) This subplot counts the number of strains where the Illumina assembly had a higher copy number for the gene cluster than the PacBio assembly. (5) Genes that were absent in both assemblies of some strains. These are more likely to be true absences instead of assembly artifcats. (6) Genes with the same copy number in both assemblies.

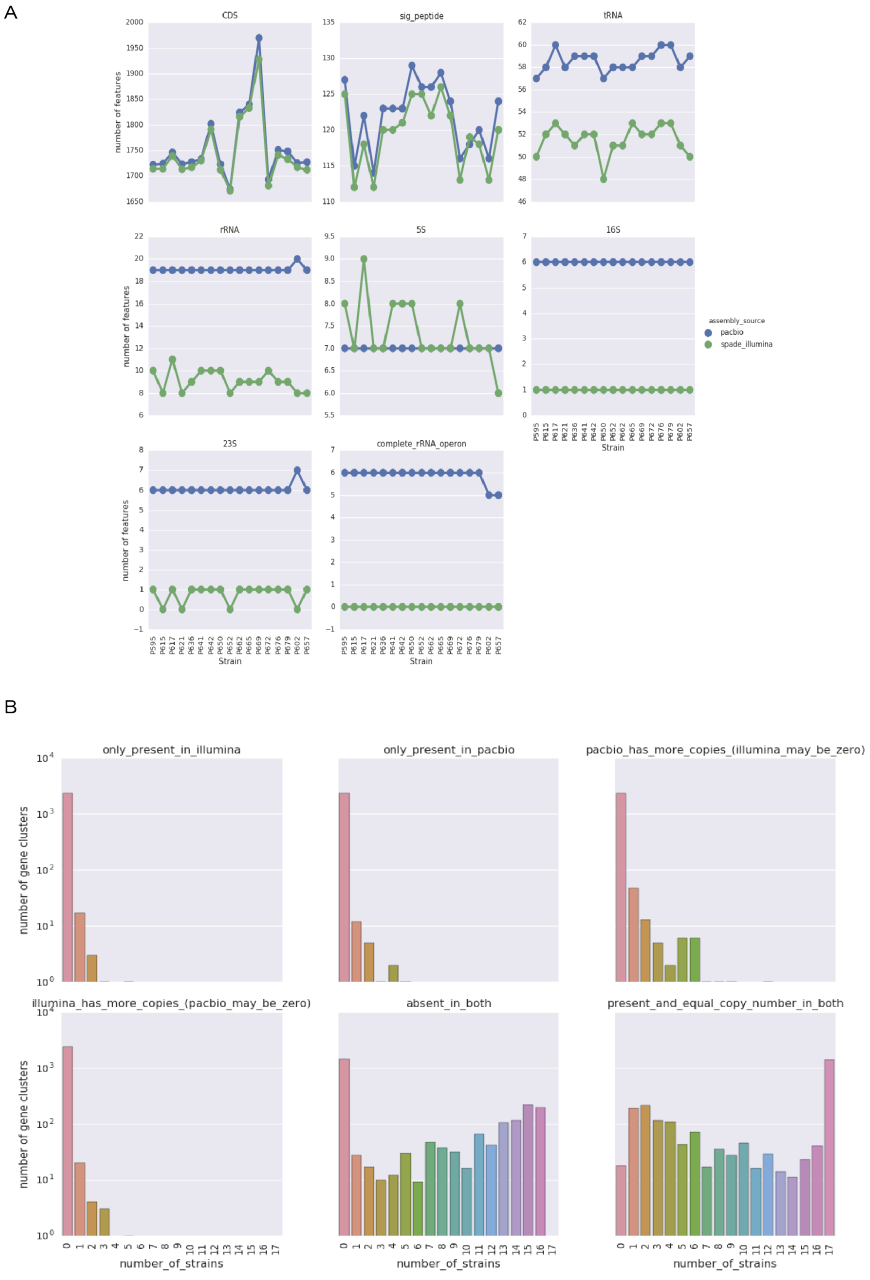


Figure S3.2. Heatmap of gene content of *Haemophilus influenzae* strains. Each row in the heatmap is a strain. Each column is a gene from the pan genome. Dark blue indicates that the gene was annoted in the strain, and light blue indicates that the gene was not annotated in the strain. Strains (rows) are ordered based on the phylogram on the left, and genes (columns) are ordered based on hierarchical clustering.

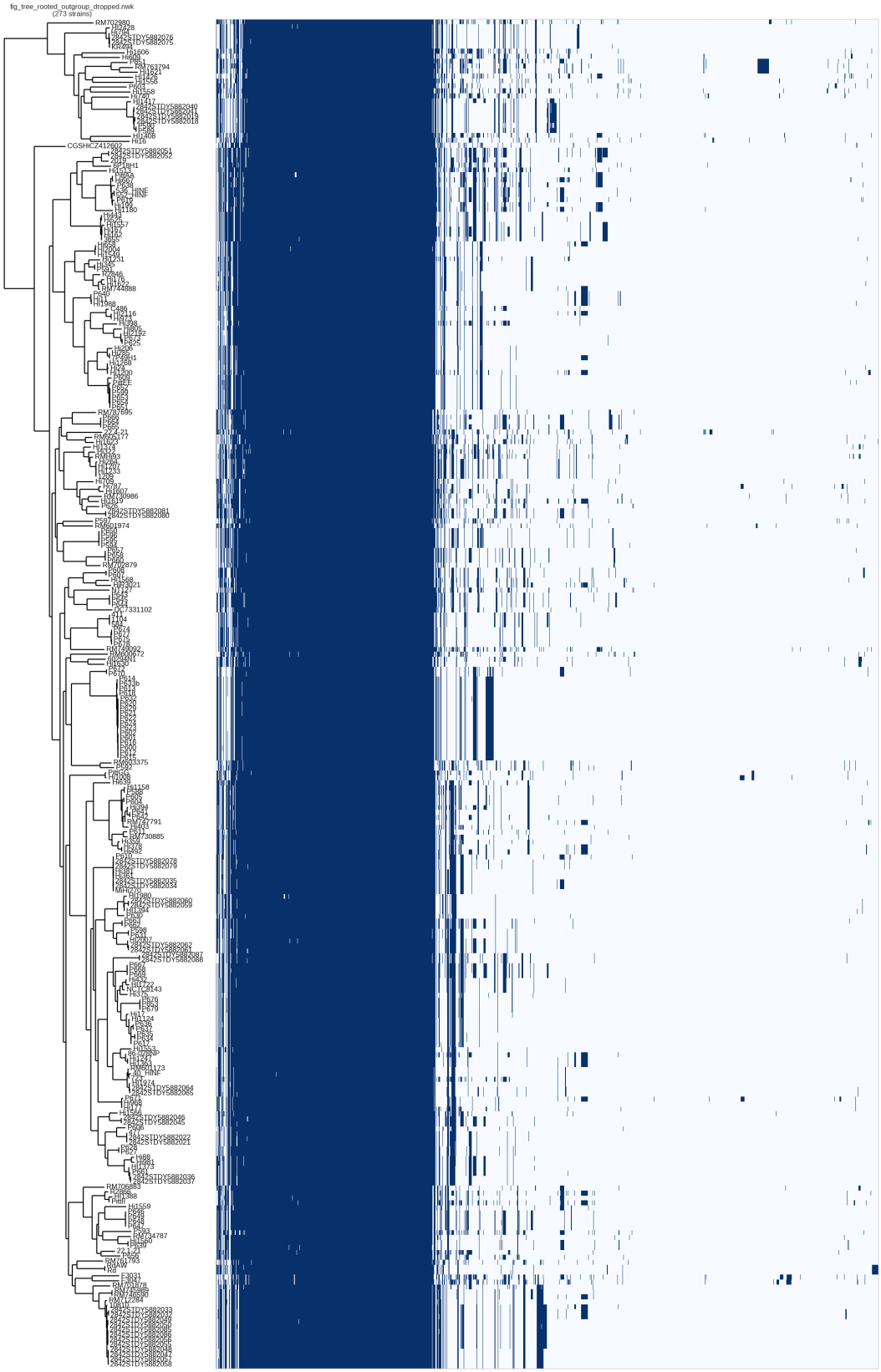


Figure S3.3: MST of database strains and strains sequenced for this study generated by goeBurst. Yellow nodes are group founders and blue nodes are common nodes.

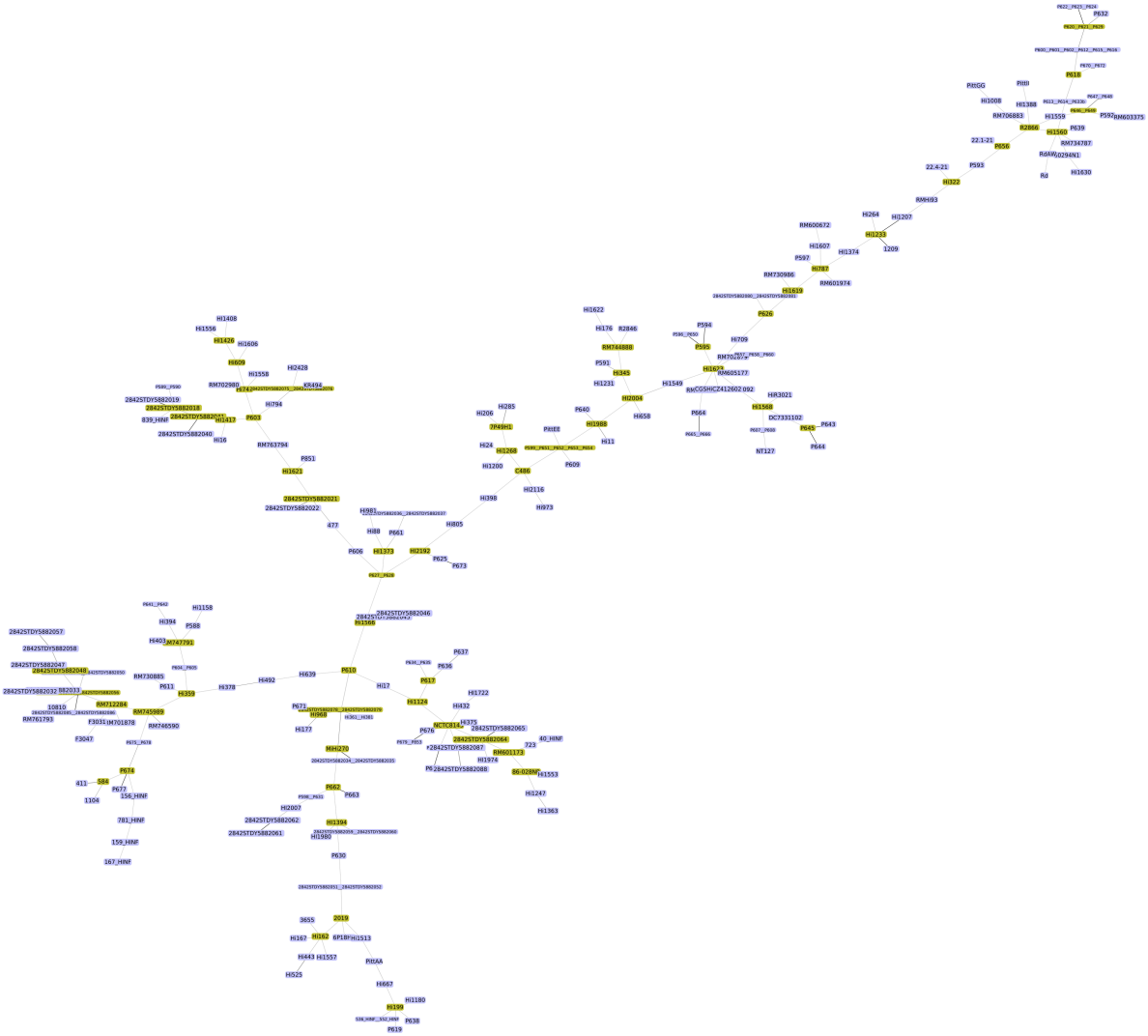


Figure S3.4: An unrooted phylogeny of *Haemophilus influenzae, Haemophilus haemolyticus,* and *Haemophilus parainfluenzae* strains. Twelve *Haemophilus haemolyticus* and 18 *Haemophilus parainfluenzae* genomes were downloaded from NCBI on October 13, 2016, annotated with Prokka, and combined with the 186 database strains and the 92 strains sequenced for this study. Genes were clustered using Roary, condon aware alignments of genes with at most one copy per strain were concatenated, and a phylogeny was created using RAxML. (METHODS) The 92 strains sequenced for this study are indicated by red dots and are distributed among the database *Haemophilus influenzae* strains.

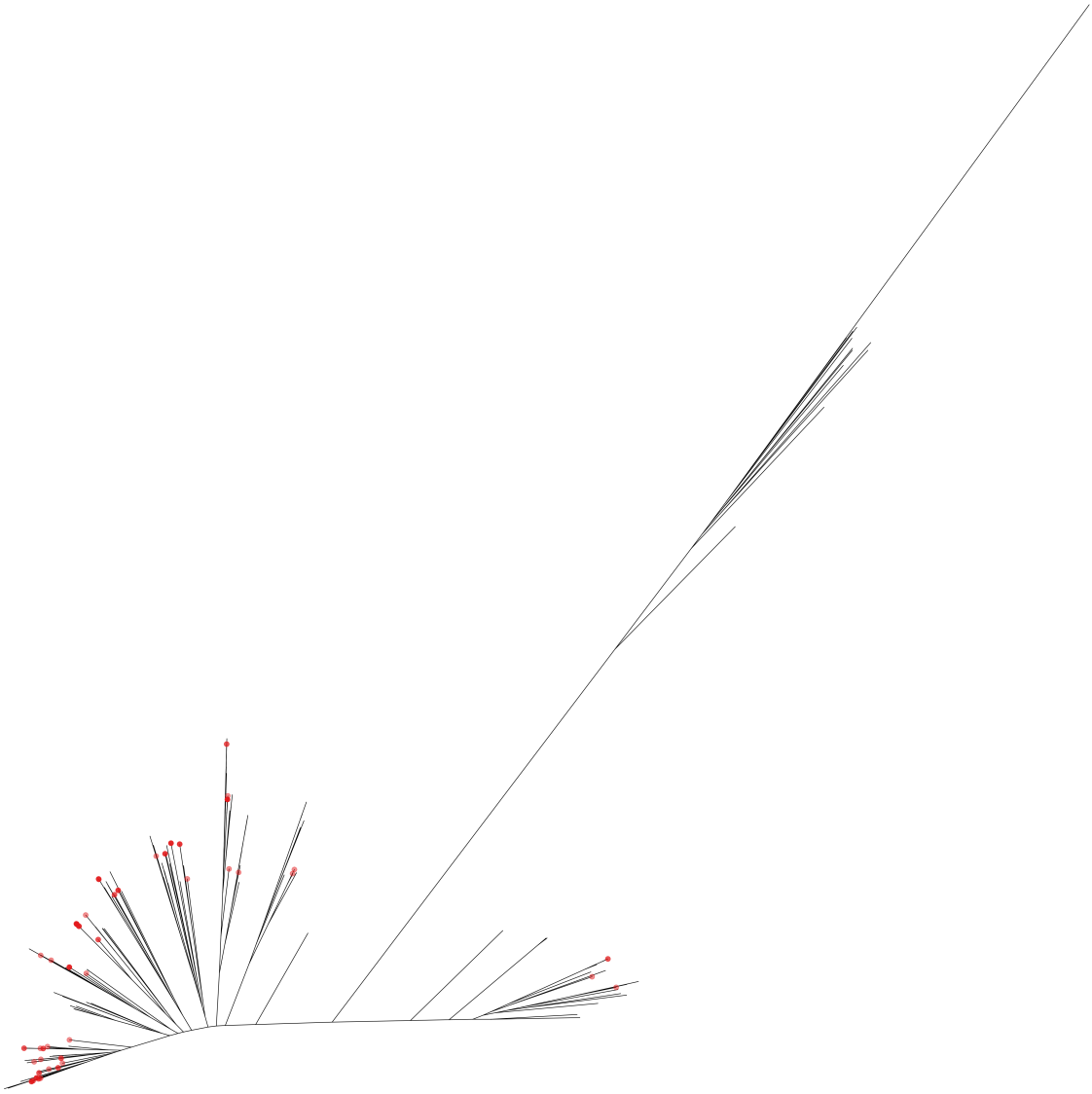


Figure S3.5. A rooted phylogram of the *Haemophilus influenzae* database strains and strains sequenced for this study. The phylogram was rooted based on the location of the *Haemophilus haemolyticus* and *Haemophilus parainfluenzae* strains in **Figure S3.4**. Terminal nodes corresponding to strains from this study are colored by patient number. Interior nodes are labeled with their bootstrap values if the values are <90. The four colorbars to the right of the phylogram describe properties of the strains. (a) no_AECOPD: Strains sequenced for this study have a red bar if they were collected during a visit in which the patient did not have an acute exacerbation of COPD. (b) broken_fadL: Strains have a green bar if the fadL locus has two annotated fadL genes. (c) lower_respiratory: Strains have a green bar if they were isolted from the lower respiratory system. (d) clonal_type: Each clonal type containing two or more strains sequenced for this study was assigned a color. All strains (this study and database) belonging to these clonal types were colored and the clonal types were labeled to the right of this colorbar.

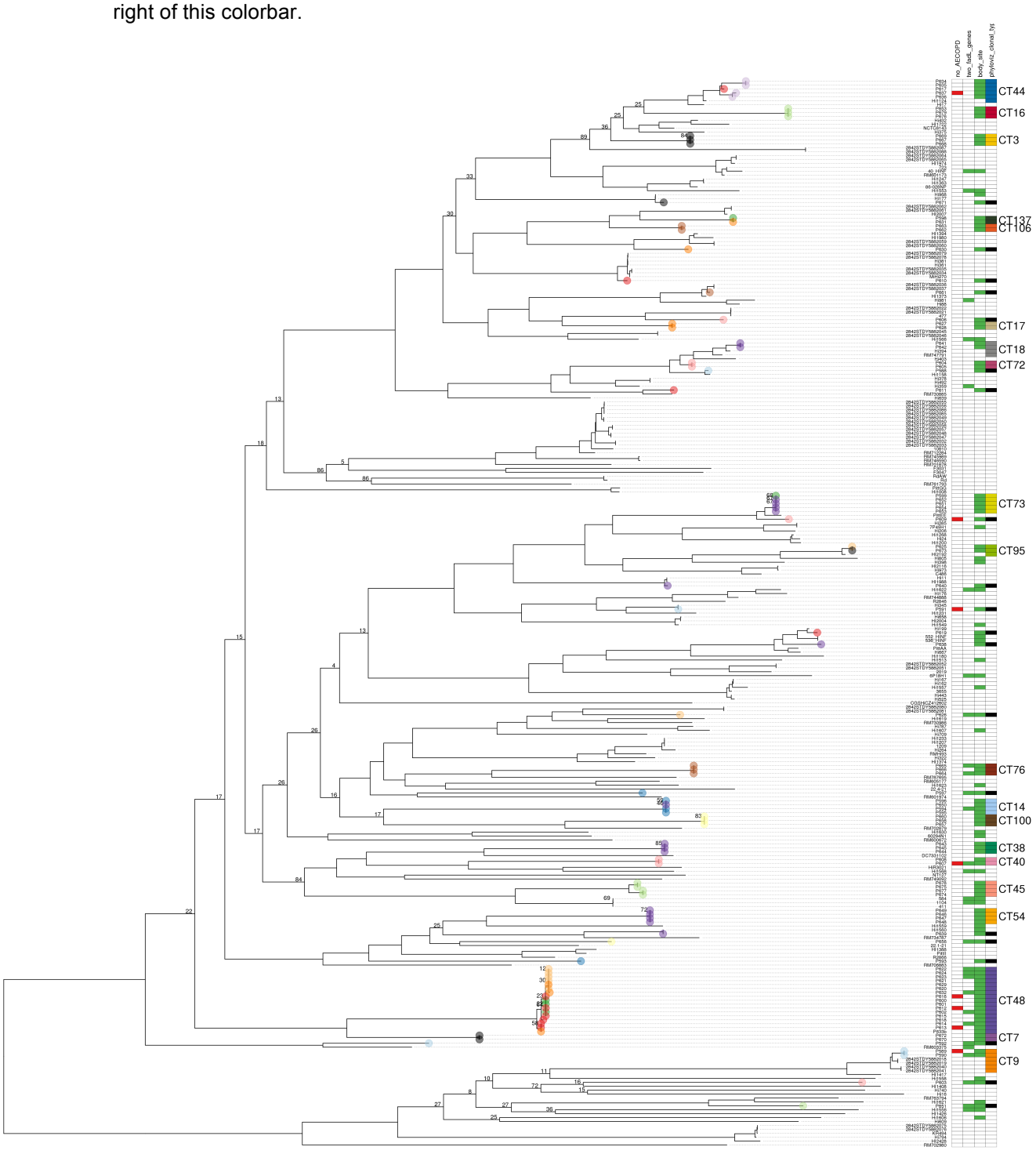


Figure S3.6: Alignment of the amino acids sequences from $FadL_{E.\ coli}$ (PDB-ID: 1T16), **$FadL_{RdKW20}$, $FadL_{NTHi375}$ and $FadL_{P608}$** (loops showed in different colors) obtained from T-Coffee multiple sequence alignment server (Notredame et al., 2000).

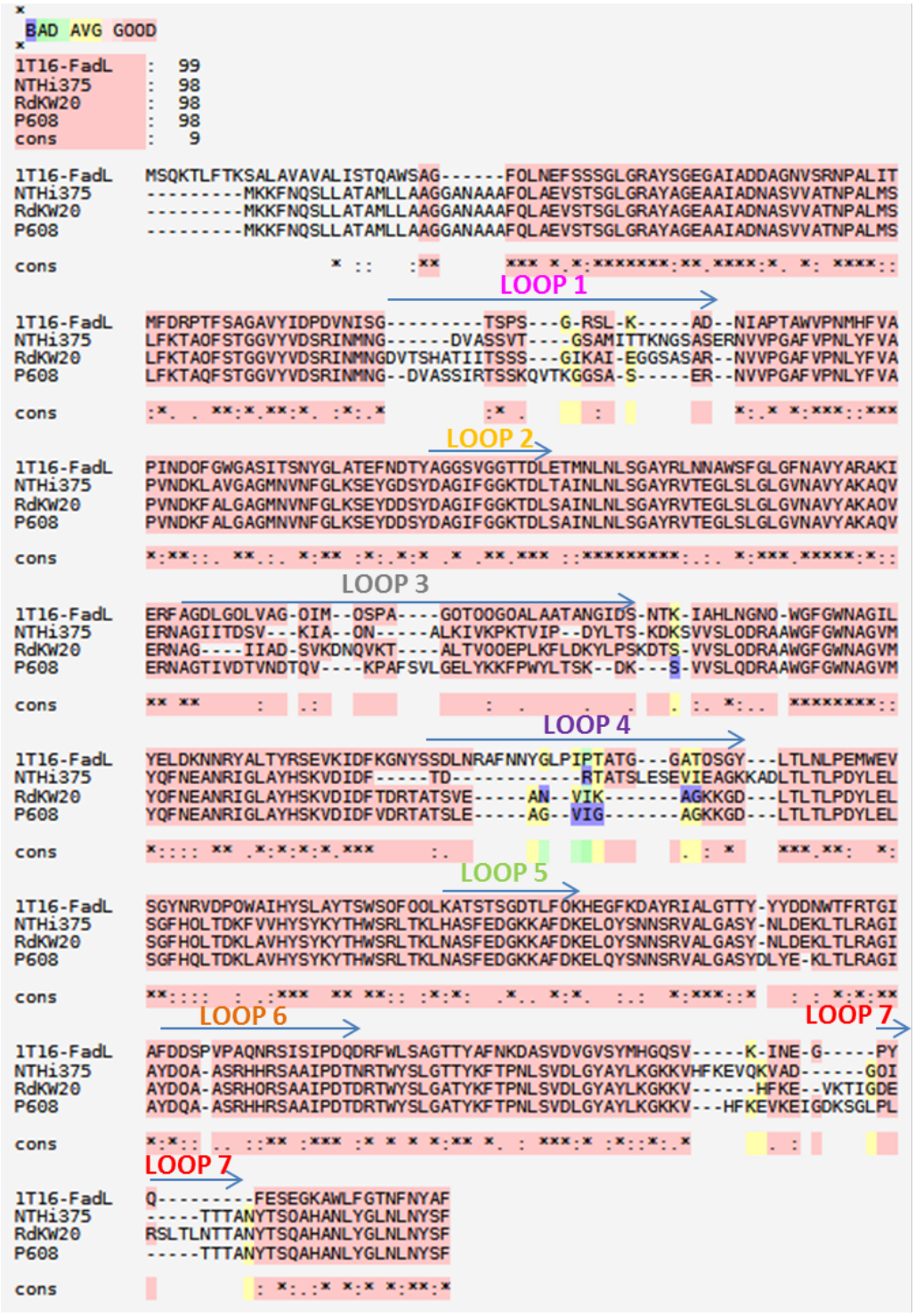


Figure S3.7: Superimposition of the three homology-based models ($FadL_{RdKW20}$ in cyan; $FadL_{NTHi375}$ in blue and $FadL_{P608}$ in lime green) **and the X-ray crystal structure from $FadL_{E.\ coli}$** (PDB-ID: 1T16, in purple). Main different and common amino acids in the inner β-barrel are shown in sticks. Regarding the inside of the β-barrel, species-specific differences in specific residues were identified: in $FadL_{E.\ coli}$ position Ala319, there is Lys in the FadL models; in $FadL_{E.\ coli}$ positions Phe33 and Phe346, Val and Tyr are respectively found in the FadL models; in $FadL_{E.\ coli}$ position His108, there is Tyr; in $FadL_{E.\ coli}$ position Glu344, there is Leu; in $FadL_{E.\ coli}$ position Lys347, there is Ser, and in $FadL_{E.\ coli}$ position Ala349, there is Asn in the FadL models. Regarding the loops, some differences were found in the predicted 3D folding, and consequently in their mobility along the MD simulations. In particular, loops L1, L3, L4 and L7 exhibited high mobility along the MD.

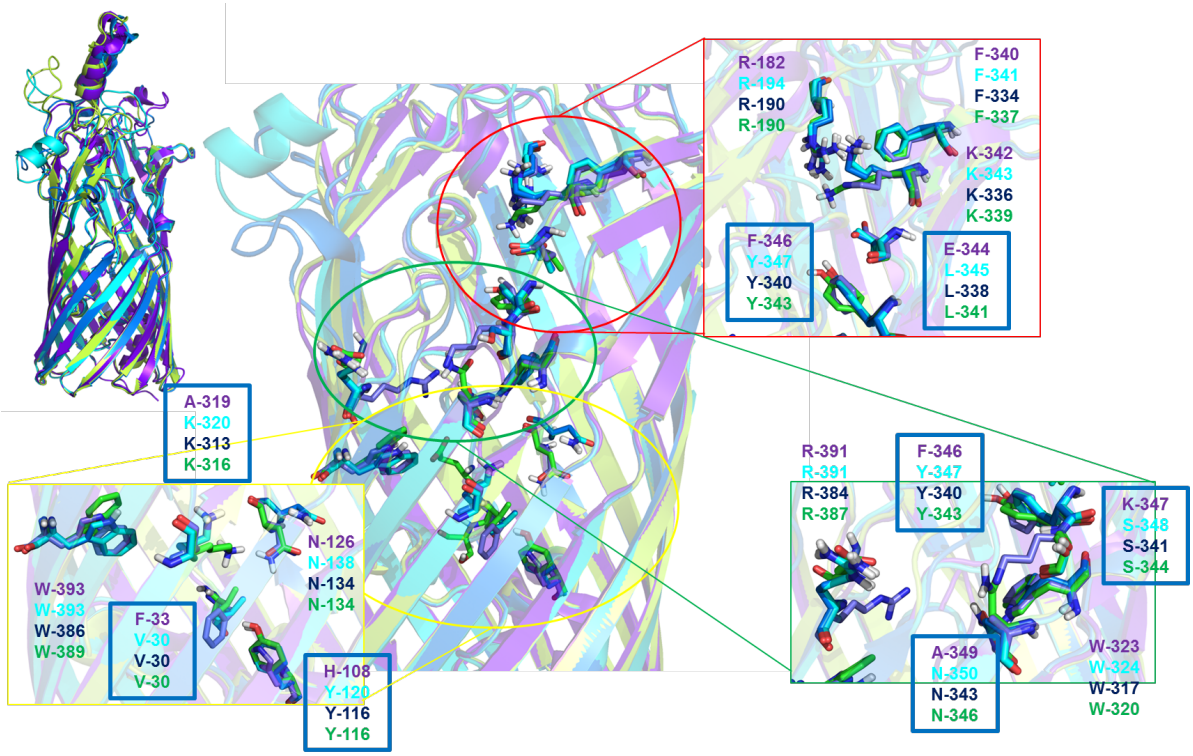


Figure S3.8: RMSD of the backbone of the different models of FadL on the left, and RMSF of the backbone of $FadL_{NTHi375}$ (blue), $FadL_{RdKW20}$ (cyan), and $FadL_{P608}$ (green) on the right.

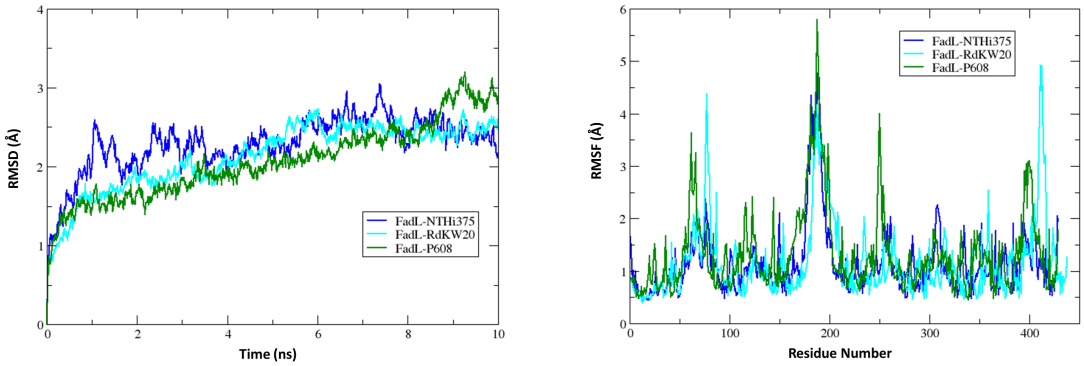


Figure S3.9: Three different conformations of $FadL_{NTHi375}$(blue, left), **$FadL_{RdKW20}$**(cyan, middle) **and $FadL_{P608}$** (green, right) **from NMA are superimposed.**

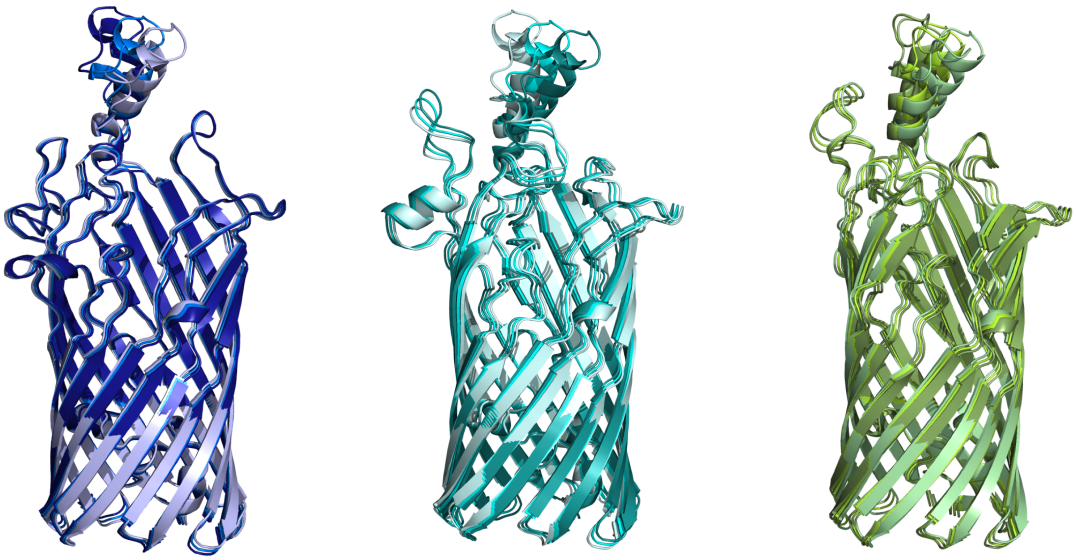


Figure S3.10: RMSD of the backbone of $FadL_{NTHi375}$ (blue) **with hCEACAM1** (green) on the left, **$FadL_{RdKW20}$**(cyan) **and hCEACAM1** (green) on the middle **and $FadL_{P608}$** (dark green) **and hCEACAM1** (green) on the right.

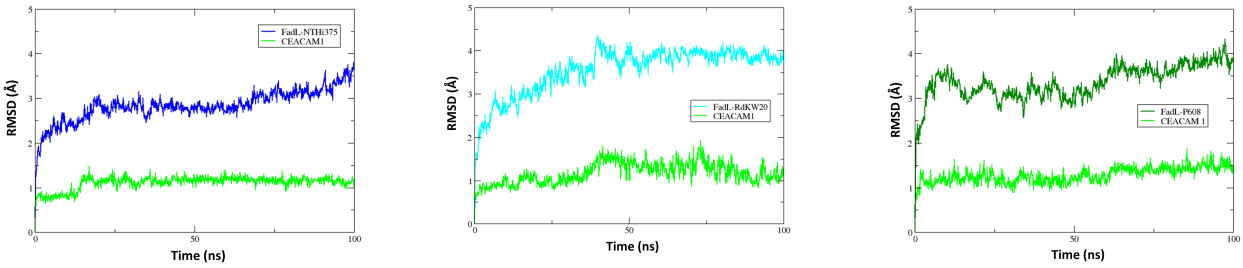


Figure S3.11: Binding poses of AA in $FadL_{P608}$ (lime green). Details of the interactions are represented in sticks.

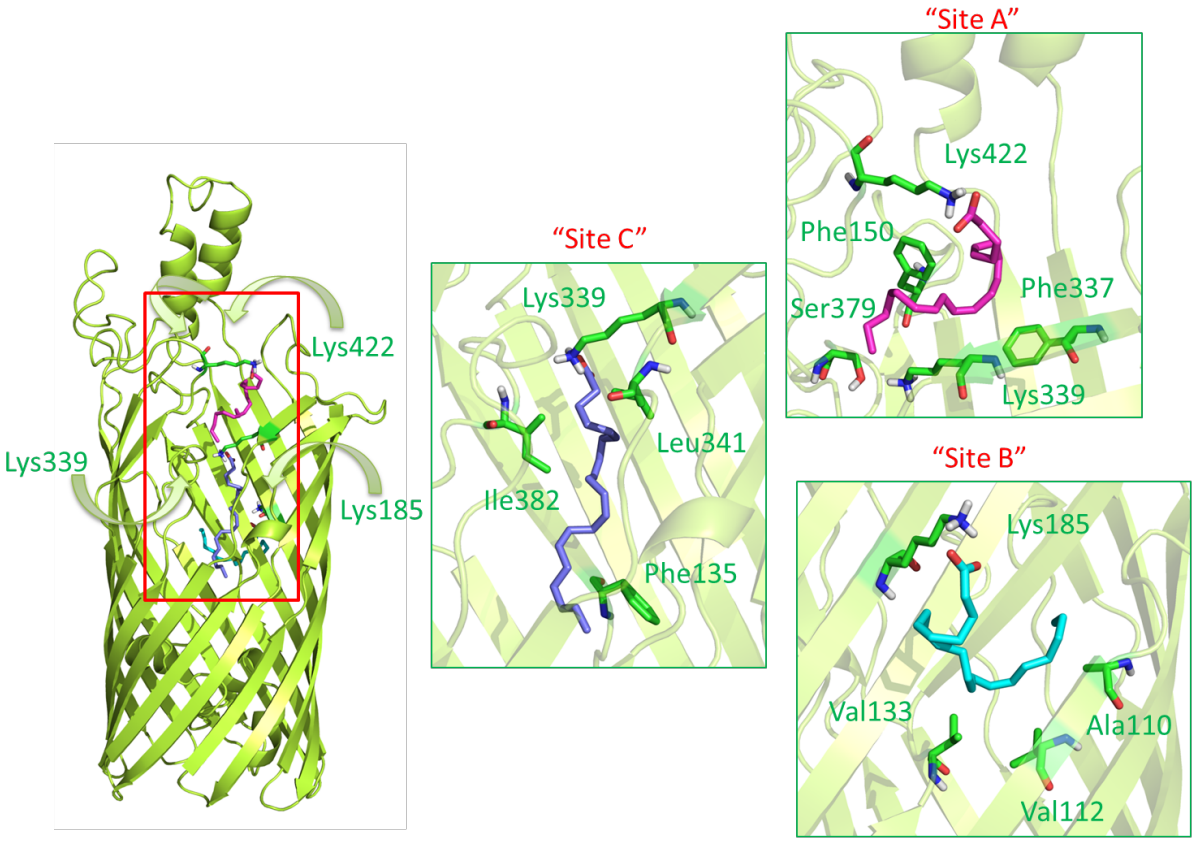


Figure S3.12: Binding poses of LDA in $FadL_{NTHi375}$ (blue) – on the left, **$FadL_{RdKW20}$** (cyan) – on the middle **and $FadL_{P608}$** (lime green) – on the right. Analogously, docking calculations of LDA in the three computational models ($FadL_{NTHi375}$, $FadL_{RdKW20}$ and $FadL_{P608}$) led to similar predicted binding poses at sites A and B for $FadL_{NTHi375}$ model. In the case of $FadL_{RdKW20}$ model, LDA is predicted to bind inside sites B and C. For $FadL_{P608}$ model, LDA is predicted to bind to the three sites A, B, and C.

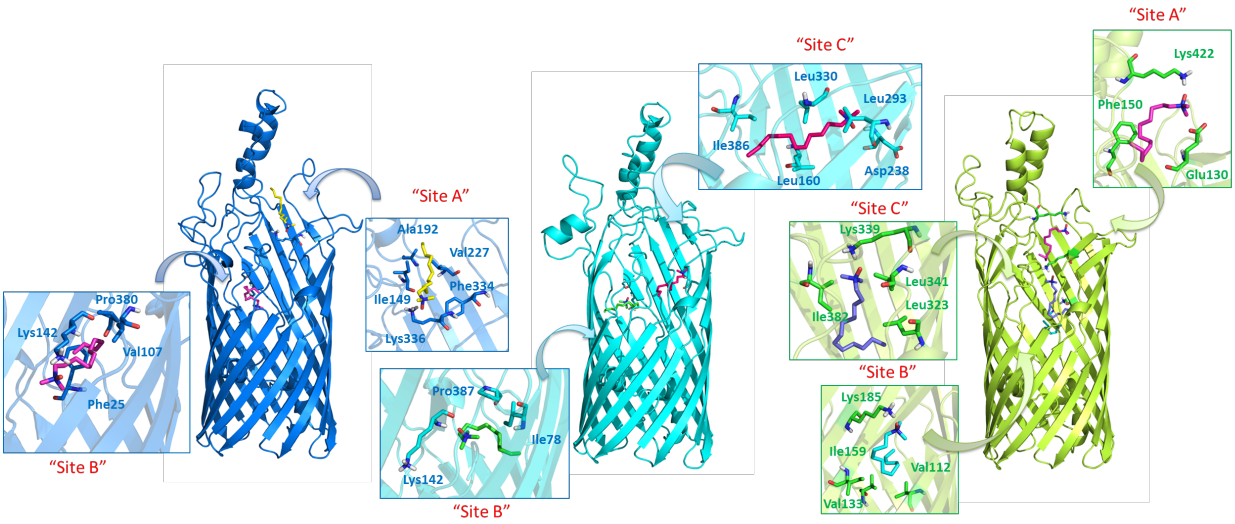


Figure S3.13: Binding poses of oleic acid (OA) in $FadL_{NTHi375}$ (blue) – left, **$FadL_{RdKW20}$** (cyan) – middle **and $FadL_{P608}$** (lime green) – right. Analogously, docking calculations of OA in the three computational models ($FadL_{NTHi375}$, $FadL_{RdKW20}$ and $FadL_{P608}$) led to similar predicted binding poses at sites A and B for $FadL_{NTHi375}$ model. In the case of $FadL_{RdKW20}$ model, OA is predicted to bind inside sites A and B. For $FadL_{P608}$ model, OA is predicted to bind to the three sites A, B, and C.

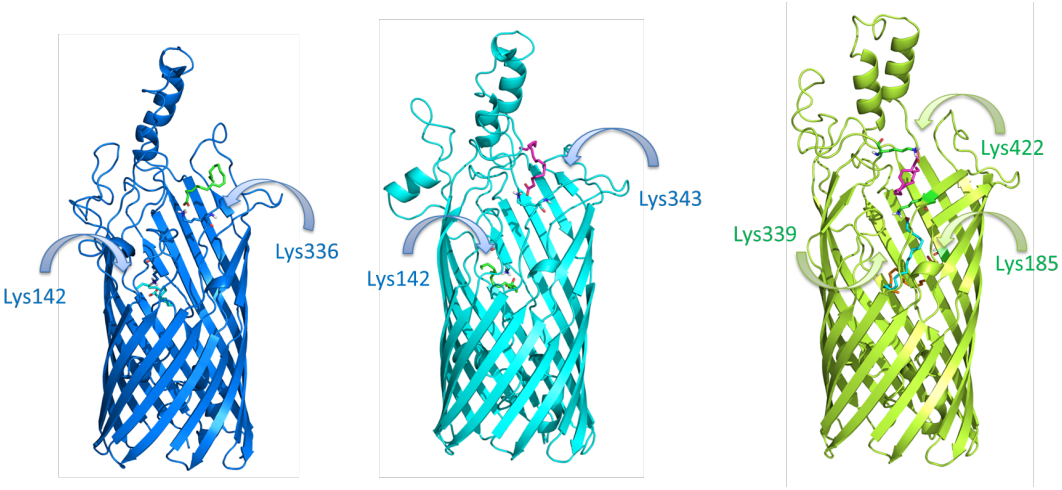
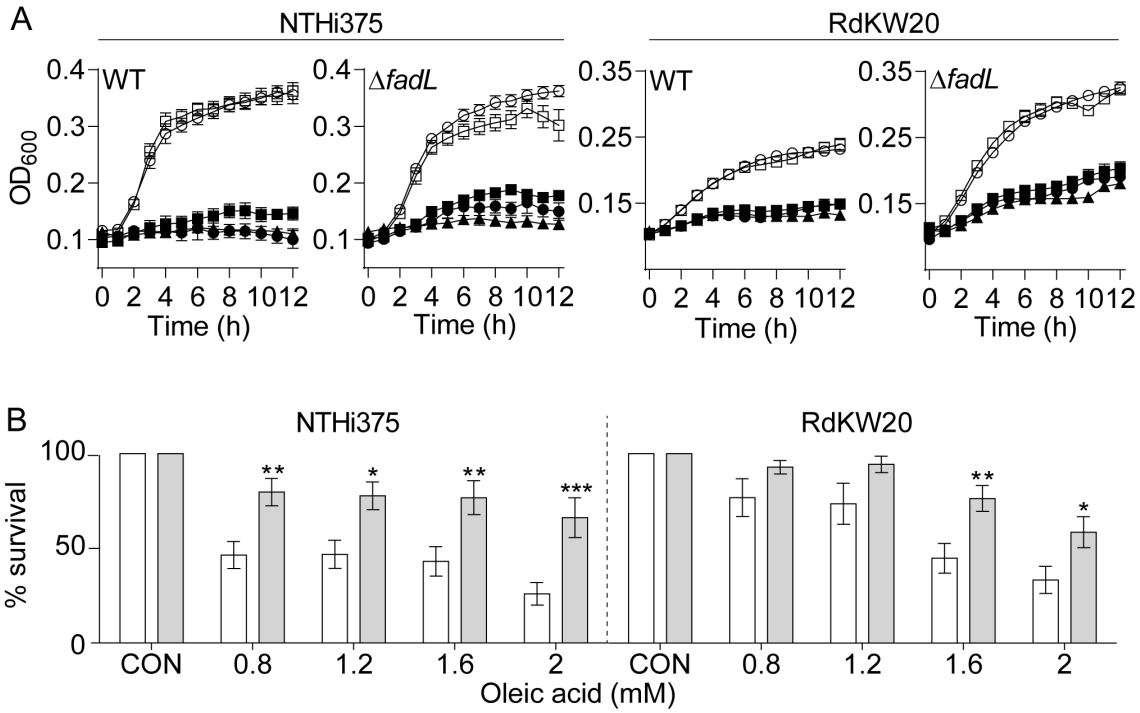


Figure S3.14: (**a**) **NTHi growth in Free Fatty Acid Minimum Media (MM-FFA)**, shown as a means of OD_{600} at the indicated time points; MM-FFA, black circle; MM-FFA + glucose 20 mM, white circle; MM-FFA + ethanol vehicle solution, black square; MM-FFA + ethanol + glucose 20 mM, white square; MM-FFA + AA 25 μM, black triangle. Glucose allowed exponential growth. Ethanol and AA did not render bacterial growth, similar to only MM-FFA conditions. (**b**) OA has a bactericidal effect on NTHi. NTHi375 and RdKW20 WT and *fadL* mutant strains grown on chocolate agar were used to generate normalized bacterial suspensions in MM-FFA, for further incubation with OA. Results are expressed as percentage of bacterial survival, referred to that in the presence of vehicle solution. A fatty acid dose dependent bactericidal effect was observed for both WT strains. Both *fadL* mutants were more resistant to OA than their respective parental strains (NTHi375: OA 0.8 mM, $p<0.005$; OA 1.2 mM, $p<0.05$; OA 1.6 mM, $p<0.005$; 2 mM, $p<0.0005$; RdKW20, OA 1.6 mM, $p<0.005$; 2 mM, $p<0.05$) *$p<0.05$, **$p<0.005$. ***$p<0.0005$.



3.6.4. Supplementary Tables

Table S3.1. Information on patients included in this study.

Patient	Year of birth	AECOPD/year	COPD stage, comorbidities, additional clinical information
1	1951	5	COPD GOLD IV (July 2012)[a], ex-smoker (40 pack year), severe emphysema, bronchiectasis, dyspnea mMRC-III[b], PBD[c] (+), DLCO[d] 11 ml/min/mg (42.2%), OCD[e] (+) **COPD specific treatment**: Relvar 92/222 µg, Spiriva 2.5 µg, Salbutamol, Combiprasal, Azithromycin 250 mg 3 days/week
2	1935	3	COPD GOLD III (August 2012), ex-smoker, emphysema, bronchiectasis, dyspnea mMRC-II, PBD (+), OCD (-) **COPD specific treatment**: Symbicort Forte, Spiriva, Azithromycin
3	1912	ND	COPD, ex-smoker, prostatism, urinary calculi, bronchiectasis, recurrent pneumonia, dyspnea **COPD specific treatment**: Atrovent, Ventolin
4	1945	ND	COPD GOLD II (November 2012), ex-smoker (50 pack year), emphysema, bronchiectasis, lung cancer with pulmonary lobectomy, pneumonia associated with neoplasia, dyspnea mMRC-I, PBD (-), DLCO 14 ml/min/mg (62%), OCD (-) **COPD specific treatment**: Budesonide Atrovent
5	1933	5	COPD GOLD II (October 2012), ex-smoker (40 pack year), emphysema, bronchiectasis, dyspnea mMRC-II, PBD (-), DLCO 11.3 ml/min/mg (51%), OCD (-) **COPD specific treatment**: Foster, Seebri, Salbutamol, Azithromycin
6	1947	7	COPD GOLD IV (October 2013), ex-smoker (30 pack year), emphysema, bronchiectasis, dyspnea mMRC-III, PBD (+), DLCO 14 ml/min/mg (59%), OCD (+) **COPD specific treatment**: Budesonide, Salbutamol, Colistin, Azithromycin
7	1963	3	COPD GOLD II (November 2011), smoker (>20 pack year), emphysema, bronchiectasis, HIV stadium 3, dyspnea mMRC-II, OCD (-) **COPD specific treatment**: Salbutamol, Atrovent
8	1939	3	COPD GOLD II (May 2009), ex-smoker, emphysema, bronchiectasis, dyspnea mMRC-II, PBD (-), DLCO (23%), OCD (+) **COPD specific treatment**: Seretide, Atrovent, Predmisone, Azithromycin
9	1921	2	COPD GOLD I (2007), ex-smoker, lung fibrosis (EPID), emphysema, bronchiectasis, chronic anemia **Treatment specific EPID:** Salbutanol, Atrovent, Salmetero/Fluticasone
10	1946	ND	COPD GOLD III (March 2008), ex-smoker, emphysema, bronchiectasis, recurrent bronchial infections, lung micobacteriosis, thyroid cancer, PBD (-), OCD (+) **COPD specific treatment**: Plusvent, Spiriva, Atrovent, Combiprasal, Colistin
11	1946	>3	COPD GOLD IV (November 2012), ex-smoker (100 pack year), cancer colon, bronchiolectasis, dyspnea mMRC-III, PBD (-), DLCO 9 ml/min/mg (37%), OCD (+) **COPD specific treatment**: Teofilin, Salbutanol, Ipratropi, Salmeterol, Azithromycin
12	1937	3	COPD GOLD III (March 2011), ex-smoker (67 pack

			year), bronchiectasis, dyspnea mMRC-II, PBD (-), DLCO 4.3 ml/min/mg (19%), OCD (+) **COPD specific treatment**: Daxar, Spiriva, Rilast, Ventolin
13	1926	7	COPD GOLD III (2007), ex-smoker (20 pack year), emphysema, dyspnea mMRC-III, DLCO (9.1%), OCD (-) **COPD specific treatment**: Symbicort, Spiriva, Salbutanol, Azithromycin

[a]Date of sperimetry test used to assign COPD GOLD stage
[b]dyspnea mMRC, modified Medical Research Council scale
[c]PBD, post-bronchodilator tests
[d]DLCO, carbon monoxide diffusion in the lung
[e]OCD, Chronic Coughing

Table S3.2. Information on the NTHi COPD isolates included in the longitudinal strain collection under study. Patient treatment at each NTHi isolation time-point, and isolation information.

Patient	NTHi strain	Isolation date dd/mm/yyyy	Isolation medical center	Treatment
COPD 1	P588-8079	13/02/2012	Hospital admission	Moxifloxacin + Ceftriaxone in ER Levofloxacin (10 days) Corticosteroids
	P589-8275	22/05/2012	Primary Health Care	No AECOPD – No treatment given
	P590-8360	10/07/2012	Primary Health Care	Cefditoren (cephalosporin) Corticosteroids
	P591-8654	12/02/2013	NO DATA	No AECOPD – No treatment given
	P592-9166	18/10/2013	Primary Health Care	Amoxicillin / Clavulanic acid (7 days) Corticosteroids
COPD 2	P593-7974	02/01/2012	Hospital admission	Levofloxacin (4 days) Ciprofloxacin (10 days) Corticosteroids
	P594-8239	02/05/2012	Primary Health Care	Cefuroxime 500 mg/12 h
	P595-8370	19/07/2012	NO DATA	NO DATA
	P596-8591	11/01/2013	NO DATA	NO DATA
	P597-8879	17/05/2013	Primary Health Care	Amoxicillin / Clavulanic acid (7 days) Corticosteroids
COPD 3	P674-1509	01/02/2000	Hospital admission	Ceftriaxone (10 days)
	P675-1755	13/12/2000	Emergency Room	Ceftriaxone (7 days) Amoxicillin / Clavulanic acid (5 days)
	P676-2514	16/10/2002	Hospital admission	Ceftriaxone (7 days) Amoxicillin / Clavulanic acid (3 days)
	P677-2580	29/12/2002	Hospital admission	Ceftriaxone (6 days) Amoxicillin / Clavulanic acid (9 days)
	P678-2720	23/03/2003	Hospital admission	Amoxicillin / Clavulanic acid (3 days) Ceftriaxone (3 days) Amoxicillin / Clavulanic acid (17 days)
	P679-2791	09/05/2003	NO DATA	NO DATA
	P851-2713	23/03/2003	NO DATA	NO DATA
	P853-2792	09/05/2003	NO DATA	NO DATA
COPD 4	P598-8560	22/12/2012	Hospital admission	No treatment
	P599-8624	28/01/2013	Hospital admission	Amoxicillin / Clavulanic acid (10 days) Corticosteroids
	P600-8643	06/02/2013	Hospital admission	Amoxicillin / Clavulanic acid (9 days)
	P601-8794	10/04/2013	Hospital admission	Amoxicillin / Clavulanic acid (7 days)
	P602-8883	16/05/2013	Hospital admission	Ciprofloxacin 750/12h (initial) Amoxicillin / Clavulanic acid (4 weeks) Corticosteroids
COPD 5	P603-4482	22/02/2010	Hospital admission	Ceftriaxone + Levofloxacin
	P604-7629	11/04/2011	Primary Health Care	Amoxicillin / Clavulanic acid (10 days) Corticosteroids
	P605-7719	08/06/2011	NO DATA	NO DATA
	P606-8650	07/02/2013	Primary Health Care	Trimethoprim / Sulfamethoxazole (15 days)

				Corticosteroids
	P607-8844	03/05/2013	NO DATA	No AECOPD – No treatment given
	P608-8895	24/05/2013	Primary Health Care	Corticosteroids
	P609-9724	08/04/2014	Primary Health Care	No AECOPD – Start chronic weekly Azithromycin treatment
COPD 6	P610-7579	21/03/2011	Hospital admission	Amoxicillin / Clavulanic acid Levofloxacin (10 days) Corticosteroids
	P611-7680	20/05/2011	Hospital admission	Cefepime (10 days) Corticosteroids
	P612-8066	07/02/2012	NO DATA	No AECOPD – No treatment given
	P613-8456	28/09/2012	NO DATA	No AECOPD – No treatment given
	P614-8522	26/11/2012	Hospital admission	Amoxicillin / Clavulanic acid (15 days) Corticosteroids
	P615-8618	25/01/2013	NO DATA	Amoxicillin / Clavulanic acid (10 days) Corticosteroids
	P616-8647	11/02/2013	NO DATA	No AECOPD – No treatment given
	P617-9224	13/11/2013	Primary Health Care	Amoxicillin / Clavulanic acid (15 days) Corticosteroids
	P618-9380	13/01/2014	Primary Health Care	Amoxicillin / Clavulanic acid (21 days) Corticosteroids
	P619-9590	31/03/2014	NO DATA	NO DATA
COPD 7	P620-6998	08/09/2010	Hospital admission	Ceftriaxone (4 days) Cefuroxime (3 days)
	P621-7028	23/09/2010	Emergency Room	Amoxicillin / Clavulanic acid (10 days) Corticosteroids
	P622-7806	25/08/2011	Hospital admission	Amoxicillin / Clavulanic acid (7 days) Corticosteroids
	P623-7854	13/10/2011	Hospital admission	Amoxicillin / Clavulanic acid (10 days) Corticosteroids
	P624-7870	25/10/2011	Hospital admission	Ceftriaxone (10 days) Corticosteroids
	P625-8065	06/02/2012	Hospital admission	Amoxicillin / Clavulanic acid (10 days) Corticosteroids
	P626-9221	15/11/2013	NO DATA	Ceftriaxone + Clarithromycin
COPD 8	P627-3766	19/01/2009	NO DATA	NO DATA
	P628-3847	20/02/2009	NO DATA	NO DATA
	P629-7106	24/11/2010	Hospital admission	Amoxicillin / Clavulanic acid (10 days) Corticosteroids
	P630-7232	21/01/2011	Emergency Room	Levofloxacin (7 days)
	P631-8237	01/05/2012	Hospital admission	Levofloxacin (3 days) Cefditoren (7 days) Corticosteroids
	P632-8384	27/07/2012	Hospital admission	Ceftriaxone + Levofloxacin (2 days) Levofloxacin (5 days) Start chronic weekly Azithromycin treatment (August 2012)
	P633-8486	26/10/2012	Hospital admission	Amoxicillin / Clavulanic acid (7 days) Corticosteroids
COPD 9	P634-7311	22/02/2011	Hospital admission	Amoxicillin / Clavulanic acid (10 days) Corticosteroids

	P635-7537	08/03/2011	Hospital admission	Levofloxacin (15 days) Corticosteroids
	P636-8296	05/06/2012	Hospital admission	Amoxicillin / Clavulanic acid (7 days)
	P637-8346	30/06/2012	NO DATA	No AECOPD Levofloxacin (7 days)
COPD 10	P638-3144	19/01/2005	NO DATA	NO DATA
	P639-3649	03/03/2008	NO DATA	NO DATA
	P640-3960	31/03/2009	NO DATA	NO DATA
	P641-4342	28/12/2009	NO DATA	NO DATA
	P642-4396	19/01/2010	NO DATA	NO DATA
	P643-7915	28/11/2010	NO DATA	NO DATA
	P644-8083	13/02/2012	NO DATA	NO DATA
	P645-8193	27/03/2012	NO DATA	NO DATA
	P646-8484	25/10/2012	NO DATA	NO DATA
	P647-8506	13/11/2012	NO DATA	NO DATA
	P648-8526	29/11/2012	NO DATA	NO DATA
	P649-8561	24/12/2012	NO DATA	NO DATA
	P650-8603	18/01/2013	NO DATA	NO DATA
	P651-8849	06/05/2013	NO DATA	NO DATA
	P652-8881	18/05/2013	NO DATA	NO DATA
	P653-8956	20/06/2013	NO DATA	NO DATA
	P654-8983	05/07/2013	NO DATA	NO DATA
	P655-9489	05/02/2014	NO DATA	NO DATA
COPD 11	P656-8020	20/01/2012	Emergency Room	Amoxicillin / Clavulanic acid (10 days) Corticosteroids
	P657-8759	26/03/2013	Emergency Room	Amoxicillin / Clavulanic acid (7 days) Corticosteroids
	P658-8889	22/05/2013	Hospital admission	Amoxicillin / Clavulanic acid (7 days) Corticosteroids
	P659-8953	19/06/2013	Hospital admission	Amoxicillin / Clavulanic acid (10 days) Corticosteroids
	P660-9367	08/01/2014	Hospital admission	Levofloxacin (14 days) Corticosteroids
COPD 12	P661-4460	10/02/2010	Emergency Room	Amoxicillin / Clavulanic acid (7 days) Corticosteroids
	P662-7189	30/12/2010	Emergency Room	Amoxicillin / Clavulanic acid (15 days) Corticosteroids
	P663-7216	17/01/2011	Hospital admission	Amoxicillin / Clavulanic acid (10 days) Corticosteroids
	P664-7614	04/04/2011	Hospital admission	Amoxicillin / Clavulanic acid (15 days) Corticosteroids
	P665-7858	15/10/2011	Hospital admission	Amoxicillin / Clavulanic acid (10 days) Corticosteroids
	P666-8053	01/02/2012	NO DATA	Cefuroxime (15 days) Corticosteroids
COPD 13	P667-4462	13/02/2010	Emergency Room	Amoxicillin / Clavulanic acid
	P668-6062	24/05/2010	Hospital admission	Amoxicillin / Clavulanic acid (10 days) Corticosteroids

	P669-6977	16/08/2010	Emergency Room	Amoxicillin / Clavulanic acid (7 days) Corticosteroids
	P670-7113	28/11/2010	Hospital admission	Amoxicillin / Clavulanic acid (10 days) Corticosteroids
	P671-7552	10/03/2011	Emergency Room	Ciprofloxacin (10 days)
	P672-7661	07/05/2011	Emergency Room	Levofloxacin
	P673-8639	05/02/2013	COPD Consulting Room	Levofloxacin (10 days) Corticosteroids

Table S3.3. Clonal type (CT) organization of the longitudinal collection of NTHi COPD isolates under study.

Clonal Type (CT)	NTHi strain	Patient	PFGE profile	WGS type	Biosample Accession number	MLST	FadL allele type number
3	P667-4462	13	AY	Illumina	SAMN07421942	3	25
	P668-6062	13	AY	Illumina	SAMN07421943	3	25
	P669-6977	13	AY	PacBio	SAMN07421944	3	25
6	P626-9221	7	AW	Illumina	SAMN07421903	ND	T12
7	P670-7113	13	AI	Illumina	SAMN07421945	ND	26
	P672-7661	13	AI	PacBio	SAMN07421947	ND	26
8	P619-9590	6	AX	Illumina	SAMN07421896	57	34
9	P589-8275	1	C	Illumina	SAMN07421866	18	2
	P590-8360	1	C	Illumina	SAMN07421867	18	T1
14	P594-8239	2	F	Illumina	SAMN07421871	ND	T3
	P595-8370	2	F	PacBio	SAMN07421872	ND	T4
	P596-8591	2	F	Illumina	SAMN07421873	ND	T4
	P650-8603	10	AD	PacBio	SAMN07421927	ND	T4
16	P676-2514	3	P	PacBio	SAMN07421951	1069	29
	P679-2791	3	A	PacBio	SAMN07421955	1069	29
	P853-2792	3	P	Illumina	SAMN07421956	1069	29
17	P627-3766	8	W	Illumina	SAMN07421904	215	11
	P628-3847	8	W	Illumina	SAMN07421905	215	11
18	P641-4342	10	BD	PacBio	SAMN07421918	145	17
	P642-4396	10	BD	PacBio	SAMN07421919	145	17
27	P588-8079	1	B	Illumina	SAMN07421865	11	1
28	P671-7552	13	AJ	Illumina	SAMN07421946	ND	27
35	P610-7579	6	P	Illumina	SAMN07421887	139	9

38	P643-7915	10	AC	Illumina	SAMN07421920	921	18
	P644-8083	10	AC	Illumina	SAMN07421921	921	18
	P645-8193	10	AC	Illumina	SAMN07421922	921	18
40	P607-8844	5	O	Illumina	SAMN07421884	272	T10
	P608-8895	5	O	Illumina	SAMN07421885	272	8
44	P617-9224	6	P	PacBio	SAMN07421894	12	T11
	P634-7311	9	P	Illumina	SAMN07421911	12	13
	P635-7537	9	P	Illumina	SAMN07421912	12	13
	P636-8296	9	P	PacBio	SAMN07421913	12	14
	P637-8346	9	P	Illumina	SAMN07421914	12	14
45	P674-1509	3	A	Illumina	SAMN07421949	519	28
	P675-1755	3	A	Illumina	SAMN07421950	519	28
	P677-2580	3	A	Illumina	SAMN07421952	519	28
	P678-2720	3	A	Illumina	SAMN07421953	519	28
47	P661-4460	12	AZ	Illumina	SAMN07421936	147	23
48	P600-8643	4	AP	Illumina	SAMN07421877	1025	5
	P601-8794	4	AP	Illumina	SAMN07421878	1025	5
	P602-8883	4	AP	PacBio	SAMN07421879	1025	T8
	P612-8066	6	AP	Illumina	SAMN07421889	1025	5
	P613-8456	6	AP	Illumina	SAMN07421890	ND	5
	P614-8522	6	AP	Illumina	SAMN07421891	ND	T7
	P615-8618	6	AP	PacBio	SAMN07421892	1025	5
	P616-8647	6	AP	Illumina	SAMN07421893	1025	5
	P618-9380	6	AP	Illumina	SAMN07421895	ND	5
	P620-6998	7	AP	Illumina	SAMN07421897	1025	5
	P621-7028	7	AP	PacBio	SAMN07421898	1025	5
	P622-7806	7	AP	Illumina	SAMN07421899	1025	T8

	P623-7854	7	AP	Illumina	SAMN07421900	1025	T8
	P624-7870	7	AP	Illumina	SAMN07421901	1025	T8
	P629-7106	8	AP	Illumina	SAMN07421906	1025	5
	P632-8384	8	AP	Illumina	SAMN07421909	1025	T8
	P633-8486	8	AP	Illumina	SAMN07421910	ND	5
52	P609-9724	5	AV	Illumina	SAMN07421886	201	30
54	P646-8484	10	AD	Illumina	SAMN07421923	986	19
	P647-8506	10	AD	Illumina	SAMN07421924	986	19
	P648-8526	10	AD	Illumina	SAMN07421925	986	19
	P649-8561	10	AD	Illumina	SAMN07421926	986	19
59	P597-8879	2	G	Illumina	SAMN07421874	1030	T5
72	P604-7629	5	B	Illumina	SAMN07421881	644	6
	P605-7719	5	B	Illumina	SAMN07421882	644	6
73	P599-8624	4	AP	Illumina	SAMN07421876	485	4
	P651-8849	10	AE	Illumina	SAMN07421928	485	4
	P652-8881	10	AE	PacBio	SAMN07421929	485	4
	P653-8956	10	AE	Illumina	SAMN07421930	485	4
	P654-8983	10	AE	Illumina	SAMN07421931	485	4
76	P664-7614	12	AH	Illumina	SAMN07421939	474	T14
	P665-7858	12	AH	PacBio	SAMN07421940	474	T15
	P666-8053	12	AH	Illumina	SAMN07421941	474	24
83	P640-3960	10	BC	Illumina	SAMN07421917	165	16
87	P638-3144	10	BA	Illumina	SAMN07421915	1230	15
91	P639-3649	10	BB	Illumina	SAMN07421916	238	T9
95	P625-8065	7	U	Illumina	SAMN07421902	253	10
	P673-8639	13	AK	Illumina	SAMN07421948	253	10
100	P657-8759	11	AM	PacBio	SAMN07421933	670	20

	P658-8889	11	AM	Illumina	SAMN07421934	670	21
	P660-9367	11	AM	Illumina	SAMN07421935	670	22
105	P656-8020	11	AL	Illumina	SAMN07421932	ND	T13
106	P662-7189	12	AG	PacBio	SAMN07421937	46	T6
	P663-7216	12	AG	Illumina	SAMN07421938	46	T6
107	P591-8654	1	D	Illumina	SAMN07421868	34	32
119	P603-4482	5	L	Illumina	SAMN07421880	ND	T8
124	P630-7232	8	Y	Illumina	SAMN07421907	203	12
135	P593-7974	2	E	Illumina	SAMN07421870	932	33
137	P598-8560	4	AO	Illumina	SAMN07421875	1066	3
	P631-8237	8	AO	Illumina	SAMN07421908	1066	3
138	P611-7680	6	Q	Illumina	SAMN07421888	ND	31
140	P851-2713	3	I	Illumina	SAMN07421954	ND	T16
145	P606-8650	5	N	Illumina	SAMN07421883	409	7
146	P592-9166	1	AQ	Illumina	SAMN07421869	996	T2

ND: Not determined

Table S3.4. FadL allelic variants encoding full-length proteins.

fadL alleles encoding full length proteins	NTHi strain	Patient	Clonal Type (CT)
1	P588-8079	1	CT27
2	P589-8275	1	CT9
3	P598-8560	4	CT137
	P631-8237	8	CT137
4	P599-8624	4	CT73
	P651-8849	10	CT73
	P652-8881	10	CT73
	P653-8956	10	CT73
	P654-8983	10	CT73
5	P600-8643	4	CT48
	P601-8794	4	CT48
	P612-8066	6	CT48
	P613-8456	6	CT48
	P615-8618	6	CT48
	P616-8647	6	CT48
	P618-9380	6	CT48
	P620-6998	7	CT48
	P621-7028	7	CT48
	P629-7106	8	CT48
	P633-8486	8	CT48
6	P604-7629	5	CT72
	P605-7719	5	CT72
7	P606-8650	5	CT145
8	P608-8895	5	CT40
9	P610-7579	6	CT35
10	P625-8065	7	CT95
	P673-8639	13	CT95
11	P627-3766	8	CT17
	P628-3847	8	CT17
12	P630-7232	8	CT124
13	P634-7311	9	CT44
	P635-7537	9	CT44
14	P636-8296	9	CT44
	P637-8346	9	CT44
15	P638-3144	10	CT87
16	P640-3960	10	CT83
17	P641-4342	10	CT18
	P642-4396	10	CT18
18	P643-7915	10	CT38
	P644-8083	10	CT38
	P645-8193	10	CT38
19	P646-8484	10	CT54
	P647-8506	10	CT54
	P648-8526	10	CT54
	P649-8561	10	CT54
20	P657-8759	11	CT100
21	P658-8889	11	CT100
22	P660-9367	11	CT100
23	P661-4460	12	CT47
24	P666-8053	12	CT76
25	P667-4462	13	CT3
	P668-6062	13	CT3
	P669-6977	13	CT3
26	P670-7113	13	CT7
	P672-7661	13	CT7
27	P671-7552	13	CT28
28	P674-1509	3	CT45
	P675-1755	3	CT45
	P677-2580	3	CT45
	P678-2720	3	CT45

29	P676-2514 P679-2791 P853-2792	3 3 3	CT16 CT16 CT16
30	P609-9724	5	CT52
31	P611-7680	6	CT138
32	P591-8654	1	CT107
33	P593-7974	2	CT135
34	P619-9590	6	CT8

Table S3.5. FadL allelic variants rendering truncated proteins.

FadL truncated (T) variant type number (Tn)	Strain	Alteration in *fadL* gene	Predicted point mutation (aa)	Predicted FadL ancestor allele[1]	Truncated FadL length (aa 1 to n)	Patient	Clonal type (CT)
T1	P590-8360	C686A	S-STOP (229)	2	1-228	1	CT9
T2	P592-9166	ND[2]	ND	ND	1-47	1	CT146
T3	P594-8239	ND	ND	ND	1-215	2	CT14
T4	P595-8370	ND	ND	ND	1-395	2	CT14
	P596-8591	ND	ND	ND	1-395	2	CT14
	P650-8603	ND	ND	ND	1-395	10	CT14
T5	P597-8879	ND	ND	ND	1-214	2	CT59
T6	P662-7189	ND	ND	ND	1-41	12	CT106
	P663-7216	ND	ND	ND	1-41	12	CT106
T7	P614-8522	T485A C486A	F-STOP (162)	5	1-161	6	CT48
T8	P602-8883	G insertion in nt 207	V-STOP (75)	5	1-74	4	CT48
	P622-7806	ΔG207	V-STOP (78)	5	1-77	7	CT48
	P623-7854	ΔG207	V-STOP (78)	5	1-77	7	CT48
	P624-7870	ΔG207	V-STOP (78)	5	1-77	7	CT48
	P632-8384	ΔG207	V-STOP (78)	5	1-77	8	CT48
	P603-4482	ND	ND	ND	1-77	5	CT119
T9	P639-3649	ND	ND	ND	1-36	10	CT91
T10	P607-8844	A659G	W-STOP (220)	8	1-219	5	CT40
T11	P617-9224	ΔCTCA (nt 1322-1325)	S-STOP (441)	13	1-440	6	CT44
T12	P626-9221	ND	ND	ND	1-167	7	CT6
T13	P656-8020	ND	ND	ND	1-241	11	CT105
T14	P664-7614	ΔT520	L-STOP (174)	24	1-173	12	CT76
T15	P665-7858	ΔATCG (nt 503-506)	Y-STOP (168)	24	1-167	12	CT76
T16	P851-2713	ND	ND	ND	1-268	3	CT140

[1]*fadL* gene variant number predicted to be an ancestor allele for specific truncated variants (see **Table S3.4**).
[2]ND = not determined

Table S3.6. Primers used in this study.

Primer	Sequence (5´-3´)	Reference
fadL-qPCR-F1/1605	CTACTTCTGGGCTTGGTCGT	This study
fadL-qPCR-F2/1606	CTACTTCTGGGCTTGGCCGT	This study
fadL-qPCR-F3/1658	CTACTTCAGGGCTTGGTCGT	This study
fadL-qPCR-R1/1607	GCAACTAACCCAGCTTTGATGA	This study
fadL-qPCR-R2/1608	GCAACAAACCCGGCTTTAATGA	This study
fadL-F1/1302	GTTTTCAACTGCCACGATTTCGCT	This study
fadL-R1/1303	CCCACGGCACGAACAGCTTTAGGA	This study
fadL-F2/1358	TGTCGTGGCAACTAACCCAGCTTTGATGAGTTTATTT AAAACGGCACAGTATTCCGGGGATCCGTCGACC	This study
fadL-R2/1359	TCAAGATTATAACTTGCCCCTAATGCAACACGAGAG TTATTACTGTATTGTGTAGGCTGGAGCTGCTTCG	This study
gyrA-qPCR-F/1078	ATATGTTGGTTGATGGGCAAGG	This study
gyrA-qPCR-R/1079	GGCGAGAAATTGACGGTTTCT	This study
Hmw2A-Seq-F1/1655	TGGACATAGCCGACACCTGTA	This study
Hmw1A-Seq-F1/1656	ATAGGTGTTGCCCAAAAATAA	This study
Hmw12A-Seq-F2/1665	ATTGACACGGGGTTGTGACCA	This study
Hmw12A-Seq-R/1657	ACGCACTTTCATGCGAGCAGG	This study
Hmw12A-Seq-PCR-R/1666	TGAGCTTTAACCCACGCAGTA	This study
Hmw2A-Seq-R2/1667	CTTTCTTCATCAGATAAATCT	This study
Hmw1A-Seq-R2/1668	TTGTTGCAGTTAAAGTCGCAG	This study
Hmw1A-Seq-R3$_{P653}$/1669	TCCTCGATTCATCGGTATCAT	This study
Hmw1A-Seq-F3$_{P651,P653}$/1680	GCATTGCTGCACCCGAAAACG	This study
Hmw1A-Seq-F3$_{P654}$/1681	ATAAAGCTCTCGCTGAAATCG	This study
Hmw1A-Seq-F4/1686	GGCTCAACAATTAATGTATCA	This study
Hmw1A-Seq-F5/1687	TTACCATTAGCAATGGTAACG	This study
Hmw1A-Seq-F6/1692	ATTTCCGTTTTAATAATGTAT	This study
Hmw1A-Seq-F7/1693	GAAATATTACTGTTAAAGGAG	This study
Hmw1A-Seq-R4/1694	CCACGCCACCTGTGGTCGCAT	This study
Hmw1A-Seq-R5/1695	GTGCCTGAATCAGAATTCTCC	This study
Hmw1A-Seq-R6/1696	TACCGCTAATATCGCCTGATT	This study
Hmw1A-qPCR-F/1649	TGTATCACTAAATGGAACTGGTAGAGG	This study
Hmw1A-qPCR-R/1650	TCTTATTTTCATCGCCAGGG	This study
Hmw2A-qPCR-F/1651	ATCCCATGTTCGCAAGGATA	This study
Hmw2A-qPCR-R/1652	AATTTTGCCCACCAAGAGTG	This study
gyrA-qPCR-F/1078	ATATGTTGGTTGATGGGCAAGG	This study
gyrA-qPCR-R/1079	GGCGAGAAATTGACGGTTTCT	This study

Table S3.7. Genome sequencing, assembly, and annotation metrics.

NTHi strain	Pacbio: mean subread length	Pacbio: subread N50	Pacbio: number of Reads	Illumina: raw read pairs	Number of contigs	Total length of assembly	N50 contig length	CDS	Biosample
P588				1218022	41	1863517	125945	1788	SAMN07421865
P589				727393	31	1840041	184812	1747	SAMN07421866
P590				1193982	28	1791978	309556	1696	SAMN07421867
P591				1324384	33	1793310	178430	1696	SAMN07421868
P592				1386966	38	1853450	996752	1754	SAMN07421869
P593				1465603	48	1919335	412503	1863	SAMN07421870
P594				1647336	38	1807486	216968	1714	SAMN07421871
P595	4873	6377	54146		1	1833864	1833864	1722	SAMN07421872
P596				2051996	38	1807638	216989	1714	SAMN07421873
P597				1962314	35	1795660	125279	1679	SAMN07421874
P598				2216617	40	1830334	123682	1748	SAMN07421875
P599				1574004	37	1781828	245977	1670	SAMN07421876
P600				1427329	45	1813961	158790	1718	SAMN07421877
P601				1506704	38	1812867	158794	1714	SAMN07421878
P602	4869	6369	46224		1	1843253	1843253	1725	SAMN07421879
P603				1421772	30	1846254	248091	1747	SAMN07421880
P604				1433923	45	1867586	139469	1783	SAMN07421881
P605				1257030	53	1867715	110659	1783	SAMN07421882
P606				1578414	35	1780644	193231	1684	SAMN07421883
P607				2511214	53	1857239	346957	1767	SAMN07421884
P608				1481632	35	1854390	347056	1766	SAMN07421885
P609				1346563	24	1777290	382859	1665	SAMN07421886
P610				1381243	76	1900746	126146	1828	SAMN07421887
P611				1461090	34	1802064	418518	1713	SAMN07421888
P612				1467367	53	1814489	150594	1709	SAMN07421889
P613				1201981	47	1813916	158797	1713	SAMN07421890
P614				1393628	47	1814104	158793	1711	SAMN07421891
P615	4996	6395	57402		1	1840062	1840062	1724	SAMN07421892
P616				1252291	42	1813400	158794	1715	SAMN07421893
P617	5421	6986	65426		1	1848210	1848210	1746	SAMN07421894

P618				1676463	53	1815258	158790	1711	SAMN07421895
P619				1517152	49	1925390	150222	1882	SAMN07421896
P620				2134269	55	1812678	112299	1712	SAMN07421897
P621	5331	6792	69369		1	1838740	1838740	1723	SAMN07421898
P622				1427373	41	1810835	158790	1713	SAMN07421899
P623				1953947	50	1811952	158790	1712	SAMN07421900
P624				1561436	54	1812889	158790	1716	SAMN07421901
P625				1235475	48	1777993	121364	1653	SAMN07421902
P626				1382764	27	1805573	195538	1721	SAMN07421903
P627				1368387	39	1815554	150168	1728	SAMN07421904
P628				2506537	41	1816183	140323	1727	SAMN07421905
P629				1419087	53	1813291	158790	1711	SAMN07421906
P630				1394577	26	1789738	281266	1680	SAMN07421907
P631				1663371	40	1829990	184958	1749	SAMN07421908
P632				1316290	43	1811248	158798	1717	SAMN07421909
P633				1529110	36	1813708	158801	1712	SAMN07421910
P634				1110116	29	1807565	170669	1721	SAMN07421911
P635				1520633	33	1807415	180677	1717	SAMN07421912
P636	5531	7135	70370		1	1840498	1840498	1727	SAMN07421913
P637				1364402	30	1807553	180630	1715	SAMN07421914
P638				2349140	63	1924447	111202	1872	SAMN07421915
P639				339210	104	1808740	35790	1694	SAMN07421916
P640				1759602	49	1847908	193948	1743	SAMN07421917
P641	5832	7155	42610		1	1849483	1849483	1733	SAMN07421918
P642	5515	7105	55407		1	1897311	1897311	1802	SAMN07421919
P643				1587695	36	1870747	153969	1791	SAMN07421920
P644				1417754	42	1872324	190835	1792	SAMN07421921
P645				1553752	21	1869813	313128	1795	SAMN07421922
P646				1228871	25	1860429	404825	1760	SAMN07421923
P647				1491895	29	1861098	277234	1760	SAMN07421924
P648				1425725	26	1859595	404646	1761	SAMN07421925
P649				1470078	30	1859985	199929	1761	SAMN07421926
P650	5079	6530	56750		1	1833710	1833710	1723	SAMN07421927
P651				1387941	35	1781706	215695	1671	SAMN07421928
P652	5211	6670	63194		1	1811303	1811303	1674	SAMN07421929

P653				1337945	38	1781640	245983	1670	SAMN07421930
P654				1355870	36	1781811	215742	1669	SAMN07421931
P656				1276649	31	1847403	234655	1756	SAMN07421932
P657	5271	6768	63854		1	1849298	1849298	1727	SAMN07421933
P658				1028590	34	1814041	227538	1710	SAMN07421934
P660				1400389	38	1813850	143982	1708	SAMN07421935
P661				1151033	41	1860639	150189	1794	SAMN07421936
P662	5141	6504	54421		1	1904311	1904311	1824	SAMN07421937
P663				1231765	42	1870794	129917	1819	SAMN07421938
P664				1266640	37	1836665	362407	1766	SAMN07421939
P665	5051	6672	44217		1	1908143	1908143	1840	SAMN07421940
P666				1229411	44	1883972	229632	1832	SAMN07421941
P667				1349753	59	1963017	140701	1923	SAMN07421942
P668				1218064	51	1962890	127571	1930	SAMN07421943
P669	4978	6518	55447		1	2013003	2013003	1970	SAMN07421944
P670				1292627	44	1800239	143286	1679	SAMN07421945
P671				1381456	32	1879346	270458	1795	SAMN07421946
P672	5322	6932	47614		1	1833305	1833305	1693	SAMN07421947
P673				1812941	56	1780206	121521	1654	SAMN07421948
P674				1157631	27	1759001	315303	1638	SAMN07421949
P675				1113164	31	1761515	275605	1642	SAMN07421950
P676	4948	5874	33290		1	1858630	1858630	1751	SAMN07421951
P677				1088488	26	1761345	315179	1639	SAMN07421952
P678				1255680	27	1761907	275624	1639	SAMN07421953
P679	5287	6879	62530		1	1858634	1858634	1748	SAMN07421955
P851				1391028	38	1994953	199052	1964	SAMN07421954
P853				1478012	30	1824549	155076	1740	SAMN07421956

Table S3.8. DNA modification/motif report forom PacBio assemblies.

Strain	Motif String	Center Pos	Modification Type	fraction	N Detected	N Genome	Group Tag	Partner MotifS tring	Mean Score	Mean IpdRatio	Mean Coverage	Objective Score
P665	GACCNNNNRTC	1	m6A	1	54	54	GACCNNNNRTC		82,98	5,466296	50,38889	4311,8486
	CAKAC	3	m6A	0,99967	3042	3043	CAKAC		81,13	6,032867	49,024	246726
	GATC	1	m6A	0,99939	9874	9880	GATC	GATC	84,77	6,345542	48,4603	836569,5
	CGWAAT	4	m6A	0,99773	2648	2654	CGWAAT		79,09	6,036342	47,13821	209024,78
	GAYDNNNNRTC	1	m6A	0,99725	1452	1456	GAYDNNNNRTC		78,90	5,511953	48,72589	110466,58
P662			Modifi				GGNCTNNNCGNCT					
	GGNCTNNNCGNCTA	9	base	1	6	6	A		45	2,671667	55,66667	270
	GATC	1	m6A	0,99960	10046	10050	GATC	GATC	89,50	6,199823	52,28190	898801,94
	GAGNNNNRTC	1	m6A	0,99723	722	724	GAGNNNNRTC		85,92	7,172629	51,78116	61883,742
	GAYDNNNCTC	1	m6A	0,99463	556	559	GAYDNNNCTC		84,99	5,579242	52,86871	47026,695
	GACCNNNCTC	1	m6A	0,98	49	50	GACCNNNCTC		80,38	5,370867	51,89796	3868,0266
	CGACA	4	m6A	0,83044	1533	1846	CGACA		52,52	3,944658	53,40965	68124,73
P676	GATC	1	m6A	0,98428	9770	9926	GATC	GATC	61,19	6,381234	32,62364	589394,8
	TCANNNNNNTRCC	2	m6A	0,9825	393	400	TCANNNNNNTRCC/ GGYANNNNNNTGA	GGYANNNNN NTGA	57,54	6,44392	32,13741	21404,88
	GGYANNNNNNTGA	3	m6A	0,9725	389	400	TCANNNNNNTRCC/ GGYANNNNNNTGA	TCANNNNNN TRCC	57,78	6,298302	32,60154	21921,873
	RTAYNNNNNRTTT	2	m6A	0,96381	1172	1216	RTAYNNNNNRTTT/A AAYNNNNNRTAY	AAAYNNNNN RTAY	57,83	6,37447	32,85154	65565,74
	AAAYNNNNNRTAY	2	m6A	0,94736	1152	1216	RTAYNNNNNRTTT/A AAYNNNNNRTAY	RTAYNNNNN RTTT	54,5	5,801197	31,72049	59802,04
	ATGNNNNNNCCT	0	m6A	0,96370	478	496	ATGNNNNNNCCT/A GGNNNNNNCAT	AGGNNNNNN CAT	57,93	6,716006	32,28243	26785,879
	AGGNNNNNNCAT	0	m6A	0,94153	467	496	ATGNNNNNNCCT/A GGNNNNNNCAT	ATGNNNNNN CCT	54,95	6,604178	31,81371	24307,605
	GCGCGCNYDNNND	1	m4C	0,34482	30	87	GCGCGCNYDNNND		40,5	3,517667	36,33333	466,03452
	CGCGCNCNNNVNNN DNNV	0	m4C	0,17441	15	86	CGCGCNCNNNVNN NDNNV		37,93	3,004	35,73333	88,01312
P636	GATC	1	m6A	1	9786	9786	GATC	GATC	123,7	5,921341	77,14224	1211416
	TCANNNNNNTRCC	2	m6A	1	402	402	TCANNNNNNTRCC/ GGYANNNNNNTGA	GGYANNNNN NTGA	118,8	6,06699	76,5995	46188,965
	GGYANNNNNNTGA	3	m6A	1	402	402	TCANNNNNNTRCC/ GGYANNNNNNTGA	TCANNNNNN TRCC	116,6	5,990945	76,39801	46912
	CCGAA	4	m6A	0,99909	2213	2215	CCGAA		97,75	5,011841	75,17216	216148,19
	RTAYNNNNNNRTTT	2	m6A	0,99811	1060	1062	RTAYNNNNNNRTTT	AAAYNNNNN	117,5	5,818105	77,08962	124426,73

							/AAAYNNNNNNRTAY	NRTAY				
							RTAYNNNNNNRTTT/AAAYNNNNNNRTAY	RTAYNNNNN NRTTT	106,9	5,233716	72,8763	112994,95
	ATGNNNNNNCCT	0	m6A	0,99190	490	494	ATGNNNNNNCCT/AGGNNNNNNCAT	AGGNNNNNN CAT	120,4	6,054267	76,74898	58332,47
	AGGNNNNNNCAT	0	m6A	0,98987	489	494	ATGNNNNNNCCT/AGGNNNNNNCAT	ATGNNNNNN CCT	111,0	5,835478	75,56032	53647,72
	GCGCGCNYD	1	m4C	0,69090	76	110	GCGCGCNYD		51,19	2,576316	76,8421	2789,588
	BNNDNNNNNSGCGCYA	11	Modif base	0,22222	38	171	BNNDNNNNNSGCGCYA		43,21	2,203158	83,5	424,11328
P657							RTTAYNNNNNRTTA/TAAYNNNNNRTAAY	TAAYNNNNN RTAAY	100,8	5,765165	64,29334	30249
	RTTAYNNNNNRTTA	3	m6A	1	300	300	RTTAYNNNNNRTTA/TAAYNNNNNRTAAY	RTTAYNNNN NRTTA	99,86	5,384866	63,03	29959
	TAAYNNNNNRTAAY	2	m6A	1	300	300	GATC	GATC	105,3	5,911549	64,37299	1048482,7
	GATC	1	m6A	0,99959	9952	9956	CAKAC		101,1	5,65208	65,20251	314170,6
	CAKAC	3	m6A	0,99839	3111	3116	AAGCTT	AAGCTT	95,51	5,825534	64,97612	67837,52
	AAGCTT	0	m6A	0,99719	712	714	GTTAAC	GTTAAC	94,29	5,252295	61,52459	5668,4355
	GTTAAC	4	m6A	0,98387	61	62	HNVAAGCTCT		44,18	2,759091	65,12727	1009,288
	HNVAAGCTCT	3	m6A	0,37671	55	146						
P602	GATC	1	m6A	0,99828	9867	9884	GATC	GATC	70,26	6,078947	40,84342	692238,7
	CAKAC	3	m6A	0,99610	3069	3081	CAKAC		68,59	5,905868	41,07103	209769,95
	AACNNNNNGTT	1	m6A	0,99139	1152	1162	AACNNNNNGTT	AACNNNNNG TT	64,98	5,494919	39,79427	74282,914
	CRAANNNNNNNTCAA	3	m6A	0,97969	386	394	CRAANNNNNNNTCAA		65,24	6,014405	40,67358	24723,314
	TTGANNNNNNNKTYG	3	m6A	0,85240	514	603	TTGANNNNNNNKTYG		65,27	5,857394	41,44163	24277,299
P621	CAKAC	3	m6A	1	3079	3079	CAKAC		110,7	5,610148	72,69601	341049
	AACNNNNNGTT	1	m6A	1	1162	1162	AACNNNNNGTT	AACNNNNNG TT	102	5,055878	69,79948	118585
	GATC	1	m6A	0,99969	9873	9876	GATC	GATC	116,3	5,941190	71,85931	1147930,1
	CRAANNNNNNNTCAA	3	m6A	0,99748	396	397	CRAANNNNNNNTCAA		99,30	5,395857	70,2702	39236,836
	TTGANNNNNNNKTYG	3	m6A	0,85808	520	606	TTGANNNNNNNKTYG		100	5,340289	72,34807	37872,73
P650							GTANNNNNNTAAA/TTTANNNNNNTAC	TTTANNNNN NTAC	89,64	6,435799	55,95211	31823
	GTANNNNNNTAAA	2	m6A	1	355	355	GTANNNNNNTAAA/	GTANNNNNN	89,11	6,181155	56,28732	31635
	TTTANNNNNNTAC	3	m6A	1	355	355						

							TTTANNNNNNTAC	TAAA				
	DNNNNNCGCGCGTG	6	m4C	1	10	10	DNNNNNCGCGCGTG		39,2	2,501	56,4	392
	GATC	1	m6A	0,99979	9870	9872	GATC	GATC	94,76	6,152125	56,16251	935138,44
	RTAYNNNNNNRTAAY	2	m6A	0,994169 1	341	343	RTAYNNNNNNRTAAY/RTTAYNNNNNNRTAY	RTTAYNNNNNNRTAY	89,50	5,843783	56,39003	30359,791
	RTTAYNNNNNNRTAY	3	m6A	0,99416	341	343	RTAYNNNNNNRTAAY/RTTAYNNNNNNRTAY	RTAYNNNNNNRTAAY	88,25	5,700910	56,42815	29937,021
	CTGGAGNNYNNNNDM	4	m6A	0,83720	36	43	CTGGAGNNYNNNNDM		49,86	4,043056	59,22222	1529,7311
	CTKMAG	4	m6A	0,78631	2609	3318	CTKMAG	CTKMAG	65,09	5,320430	56,49981	136801,45
	RNCTGGAGNNNNDNNND	6	m6A	0,5	46	92	RNCTGGAGNNNNDNNND		48,39	3,815870 3	59,21739	524,0717
	HNNNNCTCCAGNT	9	m6A	0,45238	38	84	HNNNNCTCCAGNT		47,78	3,850000	61,5	889,3439
P595							GTANNNNNNTAAA/TTTANNNNNNTAC	TTTANNNNNNTAC				
	GTANNNNNNTAAA	2	m6A	1	355	355		NTAC	75,80	6,8102		
	TTTANNNNNNTAC	3	m6A	0,99436	353	355	GTANNNNNNTAAA/TTTANNNNNNTAC	GTANNNNNNTAAA	75,19	6,566741		
	GATC	1	m6A	0,99959	9868	9872	GATC	GATC	78,23	6,464427		
	RTTAYNNNNNNRTAY	3	m6A	0,99416	341	343	RTTAYNNNNNNRTAY/RTAYNNNNNNRTAAY	RTAYNNNNNNRTAAY	74,48	6,129033	44,97746	26910
	RTAYNNNNNNRTAAY	2	m6A	0,99416	341	343	RTTAYNNNNNNRTAY/RTAYNNNNNNRTAAY	RTTAYNNNNNNRTAY	75,20	6,159120	45,18130	26408,377
	CTKMAG	4	m6A	0,74532	2473	3318	CTKMAG	CTKMAG	57,78	5,917789	44,59485	771706,5
	ANCTGGAGNNNNNNH	6	m6A	0,40963	34	83	ANCTGGAGNNNNNNH		44,94	3,999118	44,35191	25265,672
	HNNNTCTCCAG	9	m6A	0,38666	29	75	HNNNTCTCCAG		46,17	4,58931	44,41642	25510,38
	HBCSCGCGNNNNNND	2	m4C	0,24550	41	167	HBCSCGCGNNNNNND		40,29	2,896341	45,66599	109693,69
P672	GACCNNNNRTC	1	m6A	1	54	54	GACCNNNNRTC		82,98	5,466297	50,38889	4311,8486
	CAKAC	3	m6A	0,99967	3042	3043	CAKAC		81,13	6,032867	49,024	246726
	GATC	1	m6A	0,99939	9874	9880	GATC	GATC	84,77	6,345542	48,4603	836569,5
	CGWAAT	4	m6A	0,99773	2648	2654	CGWAAT		79,09	6,036342	47,13822	209024,78
	GAYDNNNNRTC	1	m6A	0,99725	1452	1456	GAYDNNNNRTC		78,90	5,511953	48,72590	110466,59

Table S3.9. List of of NTHi database and sequencing metrics used in this study.

Strain	pnas_clade	fadL_shorter_than_1350	Lower respiratory tract isolation	Number_of_contigs	Total_length_of_assembly	N50_contig_length
7P49H1	-	full length	yes	19	1827667	368081
6P18H1	-	short	yes	28	1912236	152038
584	-	short	yes	17	1794570	246701
1104	-	short	yes	16	1803206	247076
60294N1	-	full length	yes	19	1817364	262415
536_HINF	-	full length	yes	43	1940759	76214
40_HINF	-	short	yes	45	1800996	113320
552_HINF	-	full length	yes	45	1929827	76214
Hi1513	IV	full length	yes	33	1950434	178305
Hi1549	III	full length	yes	19	1782819	139126
Hi1553	V	short	yes	18	1778061	245650
Hi1556	I	short	yes	14	1825943	313270
Hi1557	IV	full length	yes	23	1865262	120687
Hi1558	I	full length	yes	16	1876051	260440
Hi1559	VI	full length	yes	8	1824293	989603
Hi1560	VI	short	yes	11	1807443	371631
Hi1566	V	short	yes	19	1829612	187385
Hi1568	VI	short	yes	16	1863033	190445
Hi1606	I	full length	yes	36	1956703	266883
Hi1607	VI	full length	yes	13	1838144	405737
Hi1621	I	full length	yes	17	1900066	544974
Hi1622	III	short	yes	20	1782901	363715
Hi1623	VI	full length	yes	24	1810594	133415
Hi1630	VI	full length	yes	14	1768101	991116
Hi398	II	full length	yes	22	1853767	353765
Hi805	II	full length	yes	27	1771899	144099
Hi968	V	full length	yes	12	1784375	482743
86-028NP	V	full length	no	1	1914490	1914490
PittEE	II	full length	no	1	1813033	1813033

PittGG	VI	full length	no	1	1887192	1887192
Rd	VI	full length	no	1	1830138	1830138
R2866	VI	full length	no	1	1932306	1932306
R2846	-	full length	no	1	1819370	1819370
22.1-21	-	full length	no	18	1888582	440092
3655	-	full length	no	23	1878368	184545
PittAA	-	full length	no	40	1876892	123700
PittII	VI	full length	no	25	1954291	135736
22.4-21	-	full length	no	44	1851097	129943
NT127	-	full length	no	41	1865669	77084
RdAW	VI	full length	no	32	1802170	131858
F3031	VI	full length	no	1	1985832	1985832
F3047	VI	full length	no	1	2007018	2007018
10810	-	full length	no	1	1981535	1981535
KR494	-	full length	no	1	1856176	1856176
411	-	full length	no	23	1787944	224231
CGSHiCZ412602	-	full length	no	1	1811802	1811802
Hi375	V	full length	no	1	1850897	1850897
RMHi93	-	full length	no	35	1935599	102797
MiHi270	-	full length	no	26	1835605	123586
Hi381	-	full length	no	24	1838395	139962
Hi361	-	full length	no	21	1843569	140842
Hi322	-	full length	no	33	1910278	104546
Hi359	-	short	no	30	1857899	114141
Hi345	-	full length	no	22	1811128	138608
Hi378	-	full length	no	26	1922014	164747
Hi394	-	full length	no	31	1896215	134543
Hi403	-	full length	no	31	1870733	120787
1209	-	full length	no	2	1925993	1799554
477	-	full length	no	1	1846259	1846259
C486	-	full length	no	1	1846503	1846503
723	-	full length	no	1	1887620	1887620
2019	-	full length	yes	1	1969659	1969659
156_HINF	-	full length	no	41	1990134	93303
159_HINF	-	full length	no	38	1946952	102526

781_HINF	-	full length	no	45	1986583	101265
839_HINF	-	full length	no	47	1937310	112549
167_HINF	-	full length	no	27	2105931	139903
HI1988	-	full length	no	18	1844600	193481
HI2004	-	full length	no	21	1773636	217329
HI2007	-	full length	no	37	1852997	131207
HI2116	-	full length	no	34	1831027	117064
HI1974	-	full length	no	41	1774730	137916
HI1722	-	full length	no	41	1849674	99272
HI1980	-	short	no	20	1761053	256756
HI1426	-	full length	no	18	1847224	300082
HI1417	-	full length	no	15	1829133	352109
HI1394	-	full length	no	21	1780723	143259
HI1388	-	full length	no	16	1812459	173959
HI1374	-	full length	no	30	1857438	106282
HI1373	-	full length	no	24	1839670	134804
HI1408	-	full length	no	27	1901953	116781
HI2428	-	full length	no	45	1798808	175901
HI2192	-	full length	no	22	1771410	231574
NCTC8143	-	full length	no	1	1890645	1890645
2842STDY5882078	-	full length	no	16	1835945	223914
2842STDY5882087	-	full length	no	20	1874760	243316
2842STDY5882032	-	full length	no	27	1932701	182847
2842STDY5882021	-	full length	no	13	1814878	405861
2842STDY5882040	-	full length	no	18	1839740	309325
2842STDY5882085	-	full length	no	28	1869361	182808
2842STDY5882019	-	full length	no	14	1835921	309420
2842STDY5882056	-	full length	no	26	1869958	182859
2842STDY5882058	-	full length	no	20	1830159	186117
2842STDY5882036	-	full length	no	10	1816181	528310
2842STDY5882049	-	full length	no	24	1869586	182838
2842STDY5882055	-	full length	no	22	1869605	306305
2842STDY5882046	-	full length	no	18	1868001	193028
2842STDY5882079	-	full length	no	15	1835755	230509
2842STDY5882034	-	full length	no	14	1858382	400959

2842STDY5882088	-	full length	no	17	1874827	420718
2842STDY5882080	-	full length	no	17	1892135	536968
2842STDY5882050	-	full length	no	25	1869582	182804
2842STDY5882035	-	full length	no	12	1858605	420643
2842STDY5882037	-	full length	no	12	1815468	422606
2842STDY5882018	-	full length	no	13	1835717	461837
2842STDY5882045	-	full length	no	17	1867511	193028
2842STDY5882051	-	full length	no	31	1966994	155173
2842STDY5882060	-	full length	no	15	1786362	463740
2842STDY5882086	-	full length	no	21	1868956	274285
2842STDY5882041	-	full length	no	12	1839558	446266
2842STDY5882048	-	full length	no	19	1828470	289594
2842STDY5882052	-	full length	no	36	1967016	155175
2842STDY5882062	-	full length	no	18	1860659	233089
2842STDY5882081	-	full length	no	19	1892644	536975
2842STDY5882059	-	full length	no	17	1786682	463756
2842STDY5882076	-	full length	no	11	1808707	1093881
2842STDY5882057	-	full length	no	20	1830018	289557
2842STDY5882065	-	full length	no	19	1804411	261383
2842STDY5882061	-	full length	no	14	1860100	279274
2842STDY5882064	-	full length	no	22	1805346	247032
2842STDY5882047	-	full length	no	17	1828077	332702
2842STDY5882075	-	full length	no	12	1808791	349569
2842STDY5882033	-	full length	no	26	1932832	182864
2842STDY5882022	-	full length	no	12	1815267	472448
DC7331102	VI	full length	no	12	1829869	1040802
Hi1008	VI	full length	no	26	1844757	380646
Hi1124	V	full length	no	27	1801322	128245
Hi1158	V	full length	no	30	1859828	261769
Hi1180	IV	full length	no	40	1948722	138922
Hi11	II	full length	no	18	1843648	429493
Hi1200	II	full length	no	30	1892173	250436
Hi1207	VI	full length	no	48	1870141	124078
Hi1231	III	full length	no	35	1846754	140417
Hi1233	VI	full length	no	38	1869940	123893

Hi1247	V	full length	no	20	1831925	410486
Hi1268	II	full length	no	24	1775701	377269
Hi1363	V	full length	no	17	1828129	220296
Hi1619	VI	full length	no	13	1930618	402048
Hi162	IV	full length	no	38	1916245	114017
Hi167	IV	full length	no	28	1914763	114021
Hi16	I	full length	no	13	1849145	988686
Hi176	III	full length	no	28	1779620	368989
Hi177	V	full length	no	17	1836430	294434
Hi17	V	full length	no	18	1813928	206369
Hi199	IV	full length	no	31	1913740	159238
Hi206	II	full length	no	21	1778049	138072
Hi24	II	full length	no	17	1773895	354233
Hi264	VI	full length	no	28	1870282	130390
Hi285	II	full length	no	22	1773058	363910
Hi432	V	full length	no	27	1853329	123745
Hi443	IV	full length	no	21	1832500	158805
Hi492	V	full length	no	25	1936449	193252
Hi525	IV	full length	no	24	1835945	139965
Hi609	I	full length	no	16	1902725	268377
Hi639	V	short	no	24	1846936	176889
Hi658	III	full length	no	24	1838136	191592
Hi667	IV	full length	no	50	2001952	124524
Hi709	VI	full length	no	10	1752603	986211
Hi740	I	full length	no	23	1904848	188188
Hi787	VI	full length	no	11	1831515	1022095
Hi794	I	full length	no	10	1807795	1139564
Hi88	V	full length	no	22	1848531	134839
Hi973	II	full length	no	18	1778751	377410
Hi981	V	short	no	20	1850572	164761
HiR3021	VI	full length	no	15	1906101	206323
RM600672	VI	full length	no	19	1801299	385129
RM601173	V	full length	no	11	1771509	476441
RM601974	VI	full length	no	10	1803765	1029746
RM603375	VI	short	no	13	1840054	281797

RM605177	VI	full length	no	20	1826632	189198
RM701878	VI	full length	no	19	1850592	270336
RM702879	VI	full length	no	28	1761368	204197
RM702980	I	full length	no	25	1900516	224631
RM706883	VI	full length	no	9	1837902	1005071
RM712284	VI	full length	no	23	1818502	281049
RM730885	V	full length	no	17	1840999	397085
RM730986	VI	full length	no	13	1826886	336084
RM734787	VI	full length	no	15	1775917	379952
RM744888	III	full length	no	21	1827993	183769
RM745989	VI	full length	no	21	1840562	338293
RM746590	VI	full length	no	26	1887114	224833
RM747791	V	full length	no	23	1875577	266497
RM749092	VI	full length	no	25	1937916	183088
RM761793	VI	full length	no	19	1771419	351369
RM763794	I	full length	no	21	1942710	190598
RM787695	VI	full length	no	14	1760085	387583

Table S3.10. Coding regions SNVs in strains from CTs 9 (P590 and P589) and 76 (P665 and P666).

		CT9									
nt REF	**nt ALT**	**P590**	**P589**	**Mutation type**	**Impact**	**Gene name**	**aa REF**	**aa ALT**	**aa position**	**Protein length**	**Function**
A	C	0	1	missense	MODERATE	L_00407	Ser	Arg	22	470	Putative ATPase,hypothetical protein
A	C	0	1	missense	MODERATE		Asn	Thr	282	470	
G	T	0	1	missense	MODERATE	L_00640	Pro	Thr	513	844	Translation initiation factor IF-2
A	C	0	1	stop lost & splice region	HIGH	L_00728	Ter	Ser	229	228	Outer membrane protein P1 precursor,long-chain fatty acid outer membrane transporter
T	C	0	1	missense	MODERATE	L_01000	Asn	Asp	59	137	Phage GP46 family protein,hypothetical protein
A	G	0	1	missense	MODERATE	L_01011	Ser	Pro	91	101	Putative uncharacterised protein,putative protein
C	A	0	1	missense	MODERATE	L_01014	Glu	Asp	131	161	Phage virion morphogenesis protein
T	C	0	1	missense	MODERATE	L_01016	Asn	Asp	521	546	Putative uncharacterized protein,Hypothetical protein,hypothetical protein
TG	CA	0	1	missense	MODERATE		Lys	Glu	514	546	

		CT76									
nt REF	**nt ALT**	**P665**	**P666**	**Mutation type**	**Impact**	**Gene name**	**aa REF**	**aa ALT**	**aa position**	**Protein length**	**Function**
AGGA	AA	0	1	frameshift	HIGH	L_00010	Gly	fs	140	258	Hypothetical protein
A	C	0	1	missense	MODERATE	L_00404	Lys	Thr	413	1046	Outer membrane receptor protein, putative Fe transport,hemoglobin/hemoglobin-haptoglobin binding protein B
CT	CTCGAT	0	1	frameshift & stop lost & splice region	HIGH	L_00626	Ter	fs	168	167	Outer membrane protein P1 precursor,long-chain fatty acid outer membrane transporter
T	C	0	1	non coding transcript	MODIFIER	L_00890					rRNA
A	C	0	1	missense	MODERATE	L_00973	Tyr	Asp	345	366	Outer membrane protein P2 precursor
C	T	0	1	missense	MODERATE		Gly	Ser	93	366	
A	C	0	1	missense	MODERATE	L_01164	Gln	Pro	288	704	Guanosine-3',5'-bis 3'-pyrophosphohydrolase
C	T	0	1	missense	MODERATE		Pro	Leu	402	704	
T	A	0	1	non coding transcript	MODIFIER	L_01171					rRNA
A	T	0	1	non coding transcript	MODIFIER						
T	G	0	1	missense	MODERATE	L_01757	Ser	Ala	322	366	Histidinol-phosphate aminotransferase 2

3.6.5. Supplementary Files

File S3.1. Multiple alignment of representative protein sequence for 36 full-length FadL protein variants, and 16 *fadL* alleles encoding truncated variants. Signal peptide is underlined. Predicted surface exposed loops are colored in pink (loop 1), yellow (loop 2), grey (loop 3), blue (loop 4), green (loop 5), orange (loop 6), red (loop 7).

Full-length FadL variants:

```
P600-8643       MKKFNQSILATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNAAVIATNPALMT
P674-1509       MKKFNQSILATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P606-8650       MKKFNQSILATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P630-7232       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P660-9367       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P657-8759       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P658-8889       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P646-8484       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P671-7552       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P643-7915       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P676-2514       MKKFNQSLLATAMLLAAGGANAAAFQLAEISTSGLGRAYAGEAAIADNASVVATNPALMS
P608-8895       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P666-8053       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P670-7113       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P611-7680       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P627-3766       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P661-4460       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P609-9724       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P634-7311       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P638-3144       MKKFYQSLLATEMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P619-9590       MKKFNQSILATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
HiRdKW20        MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P640-3960       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P588-8079       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P589-8275       MKKFNQSILATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P667-4462       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P599-8624       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P625-8065       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P598-8560       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P610-7579       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
NTHi375         MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P591-8654       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P593-7974       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P604-7629       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P641-4342       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
P636-8296       MKKFNQSLLATAMLLAAGGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATNPALMS
                **** **:*** **********:**********:*:*******:

P600-8643       VFKRAQFSAGGVYINSKVDMRGDVATNINN-----LTGKGKKGSASKNDVVPEAFIPNLY
P674-1509       LFKTNQFSVGGVYVDSRINMNGDVKTEAPA-----LQAINQNGSASERKVVPGAFVPNLY
P606-8650       LFKTAQFSTGGVYVDSRINMNGDVASSVTIG--G-KMKATREGSASQRNVIPGAFVPNLY
P630-7232       LFKTAQFSTGGVYVDSRINMNGDVDVRASIT--GVSVKTTKNGSASQRNVVPGAFVPNLY
P660-9367       LFKTAQFSTGGVYIDSRINMNGDVDVSASIT--GASMKTTRNGSASQRNVIPGAFVPNLY
P657-8759       LFKTAQFSTGGVYIDSRINMNGDVDVSASIT--GASMKTTRNGSASQRNVIPGAFVPNLY
P658-8889       LFKTAQFSTGGVYIDSRINMNGDVDVSASIT--GASMKTTRNGSASQRNVIPGAFVPNLY
P646-8484       LFKTAQFSTGGVYVDSRINMSGDVTSSAIV---SKRMNATKYGSASQRNVVPGAFVPNLY
P671-7552       LFKTAQFSTGGVYIDSRINMNGDVASSAIV---SNQMKATHNGSASARNVVPGAFVPNLY
P643-7915       LFKTAQFSTGGVYVDSRINMNGDVASSVLI---NNTTKATQYGSASARNVVPGAFVPNLY
P676-2514       LFKTAQFSTGGVYVDSRINMNGDVASSVTVS--NNMVKATQDGSASAHNVVPGAFVPNLY
P608-8895       LFKTAQFSTGGVYVDSRINMNGDVASSIRTS--SK--QVTKGGSASERNVVPGAFVPNLY
P666-8053       LFKTAQFSTGGVYVDSRINMNGNVASSVTTN--ST-MKTTMDGSASERNVVPGAFVPNLY
P670-7113       LFKTAQFSTGGVYIDSRINMNGDVAAHI----ATTGMNSAKYGSASKRNVVPGAFVPNLY
P611-7680       LFKTAQFSTGGVYVDSRINMNGDVNSYATITSASSGVRTIKHGSASQRNVVPGAFVPNLY
P627-3766       LFKTAQFSTGGVYIDSRINMNGDVTSYAKIITNNIGMKAIKDGSASQRNVVPGAFVPNLY
P661-4460       LFKTAQFSTGGVYIDSRINMNGDVTSYAQIITNKIGMKAIKDGSASQRNVVPGAFVPNLY
P609-9724       LFKTAQFSTGGVYIDSRINMNGDVASSAII---SSAINSMKDGSASQRNVVPGAFVPNLY
P634-7311       LFKTNQFSVGGVYVDSRINMNGDVDSYAII---SDSIKVTNDGSASARNVVPGAFVPNLY
P638-3144       LFKTAQFSTGGVYVDSRINMNGDVDSYAII---SNSIKVTNDGSASARNVVPGAFVPNLY
P619-9590       LFKTAQFSTGGVYIDSRINMNGDVTSSAT--TSSSGMRATRYGSASQRNVIPGAFVPNLY
HiRdKW20        LFKTAQFSTGGVYVDSRINMNGDVTSHATIITSSSGIKAIEGGSASARNVVPGAFVPNLY
P640-3960       LFKTAQFSTGGVYVDSRINMNGDVTASIT----SASMIATKNGSASERNVVPGAFVPNLY
P588-8079       LFKTAQFSTGGVYVDSRINMNGDVESSTK----NLQMLAAKYGSASQRNVVPGAFVPNLY
P589-8275       LFKTAQFSTGGVYVDSRINMNGDVESSIK----NFQMLAAKYGSASARNVVPGAFVPNLY
P667-4462       LFKTAQFSTGGVYIDSRINMSGDVTASIK----NSSRSTTKDDAASARNVVPGAFVPNLY
P599-8624       LFKTNQFSVGGVYVDSRINMNGDVASSIT----GASMKTTKYGSASERNVVPGAFVPNLY
P625-8065       LFKTAQFSTGGVYVDSRINMNGDVESSIK----NTKMLTAKHGSASQHNVVPGAFVPNLY
P598-8560       LFKTAQFSTGGVYVDSRINMNGDVAASVK----NTNMSTTKHGSASARNVVPGAFVPNLY
P610-7579       LFKTAQFSTGGVYVDSRINMNGDVASSVT----GTTMITTKNGSASARNVVPGAFVPNLY
```

```
NTHi375         LFKTAQFSTGGVYVDSRINMNGDVASSVT----GSAMITTKNGSASERNVVPGAFVPNLY
P591-8654       LFKTAQFSTGGVYVDSRINMNGDVDSYLR----SGTTQFTKYGSASQRNVVPGAFVPNLY
P593-7974       LFKTAQFSTGGVYIDSRINMNGDVDAYLQ----SGVVKFTKYGSASQRNVVPGAFVPNLY
P604-7629       LFKTAQFSTGGVYIDSRINMNGDVDAAVK----SLAMSFTKSGSASQRNVVPGAFVPNLY
P641-4342       LFKTAQFSTGGVYIDSRINMNGDVDAAVK----SLAMSFTKSGSASQRNVVPGAFVPNLY
P636-8296       LFKTAQFSTGGVYIDSRINMNGNVDAAVK----SLAMSFTKSGSASQRNVVPGAFVPNLY
                :**  ***.*****::*:::* *:*                  .:** ..*:* **:****

P600-8643       FVAPVNDKLALGAGMNVNFGLKSEYGRDYDAGLFGGETKLSTINLNFSGAYRVTQGLSVG
P674-1509       FVAPVNDKLALGAGMNVNFGLKSEYGDSYDAGVFGGKTDLTAINLNLSGAYRVTEGLSVG
P606-8650       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLTAINLNLSGAYRVTEGLSLG
P630-7232       FVAPVNDKFALGAGMNVSFGLKSEYDDSYDAGVFGGKTDLTAINLNLSGAYRVTQGLSLG
P660-9367       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSAG
P657-8759       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSAG
P658-8889       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSAG
P646-8484       FVTPVNDKFALGAGMNVNFGLKSEYDDSYDAGIFGGKTDLTAINLNLSGAYRVTEGLSLG
P671-7552       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGIFGGKTDLTAINLNLSGAYRVTEGLSLG
P643-7915       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSLG
P676-2514       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGIFGGKTDLSAINLNLSGAYRVTEGLSLG
P608-8895       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGIFGGKTDLSAINLNLSGAYRVTEGLSLG
P666-8053       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSLG
P670-7113       FVAPVNDKFALGAGMNVNFGLKSEYGDSYDAGIFGGKTDLSAINLNLSGAYRVTEGLSLG
P611-7680       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSLG
P627-3766       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSLG
P661-4460       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSLG
P609-9724       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGIFGGKTDLSAINLNLSGAYRVTEGLSLG
P634-7311       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLTAINLNLSGAYRVTEGLSLG
P638-3144       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLTAINLNLSGAYRVTEGLSLG
P619-9590       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGIFGGKTDLSAINLNLSGAYRVTEGLSLG
HiRdKW20        FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGIFGGKTDLSAINLNLSGAYRVTEGLSLG
P640-3960       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLTAINLNLSGAYRVTEGLSLG
P588-8079       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSLG
P589-8275       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSLG
P667-4462       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSLG
P599-8624       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLTAINLNLSGAYRVTEGLSLG
P625-8065       FVAPVNDKFAVGAGMNVNFGLKSEYDDSYDAGVFGGKTDLTAINLNLSGAYRVTEGLSLG
P598-8560       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSLG
P610-7579       FVAPVNDKLAVGAGMNVNFGLKSEYGDSYDAGIFGGKTDLTAINLNLSGAYRVTEGLSLG
NTHi375         FVAPVNDKLAVGAGMNVNFGLKSEYGDSYDAGIFGGKTDLTAINLNLSGAYRVTEGLSLG
P591-8654       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSLG
P593-7974       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGIFGGKTDLTAINLNLSGAYRVTEGLSLG
P604-7629       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSLG
P641-4342       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSLG
P636-8296       FVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTEGLSLG
                **:*****:*:******.*******. .****:***:*.*::****:*******:*** *

P600-8643       LGLNAVHAKAKLDRTAGILTKTIDGVRSNLNTLASSVPELKLVGDYLTSTDRSVVQLQDK
P674-1509       LGVNAVHAKAKVERNAGIVADSIMVAKAG-NALTALPPKQKALGQYLTSKDKSVVSLQDK
P606-8650       LGVNAVYAKAQVERNAGIIANTVNDDQVK-PALLTQPELLRDLSKNLPSKDKSVVSLQDR
P630-7232       LGVNAVYAKAQVKRNAGIIANTVNDNQVK-LPLSTQPETLRDLHKHLPSKDKSVVSLQDR
P660-9367       LGVNAVYAKAQVERNAGIIANTVNDNQVK-GALSTQSETLRDLPKYLPSKDKSVVSLQDR
P657-8759       LGVNAVYAKAQVERNAGIIANTVNDNQVK-GALSTQSETLRDLPKYLPSKDKSVVSLQDR
P658-8889       LGVNAVYAKAQVERNAGIIANTVNDNQVK-GALSTQSETLRGLPKYLPSKDKSVVSLQDR
P646-8484       LGVNAVYANAQVERNAGIIKDTVNDNQIT-GALLTQQEPLKSLNKYLSSKDKSVVSLQDR
P671-7552       LGVNAVYAKAQVERNAGTIADSIQDQTVK-NAFSVLSEELKHLPQYLSSKDKSVVSLQDR
P643-7915       LGVNAVYAKAQVERNAGIIADSVKDGQVT-QALSVLSPPFKELNKHLASKDKSVVSLQDR
P676-2514       LGVNAVYAKAQVERNAGTIAESVNDTQVS-QAFSVLSEPYKSFPKYLPSKDKSVVSLQDR
P608-8895       LGVNAVYAKAQVERNAGTIVDTVNDTQVK-PAFSVLGELYKKFPWYLTSKDKSVVSLQDR
P666-8053       LGVNAVYAKAQVERNAGVIVDIVNDTQIK-GALAALSKPYKDFPNYLTSKDKSVVSLQDR
P670-7113       LGVNAVYANAQVERNAGIIANSVNDTQVQ-RALSVLAPPLKDLDKNLPSKDKSVVSLQDR
P611-7680       LGVNAVYANAQVERNAGIIVETID-DQVN-TALSVLQEPFRDLKKHLPSKDKSVVSLQDR
P627-3766       LGVNAVYAKAQVERNAGIIVETAKDNQVT-GALSTLGDPYKNLPQYLPSKDTSVVSLQDS
P661-4460       LGVNAVYAKAQVERNAGLIADSVKDDQIT-SALSTQQEPFRDLKKYLPSKDKSVVSLQDR
P609-9724       LGVNAVYAKAQVERNAGIIVDSIQDRQVK-TALTVLGEPLRDLNQHLPSKDKSVVSLQDR
P634-7311       LGVNAVYAKAQVERNAGLIADTVKDNQVK-NTLTVQQEPLKFIDKYLPSKDTSVVSLQDR
P638-3144       LGVNAVYAKAQVERNAGLIADTVKDSQVK-TTLIVQQEPLKSIDKYLPSKDTSVVSLQDR
P619-9590       LGVNAVYAKAQVERNAGIIVDSLKDNQVQ-TALSVQKEPLKYLHKYLPSKDTSVVSLQDR
HiRdKW20        LGVNAVYAKAQVERNAGIIADSVKDNQVK-TALTVQQEPLKFLDKYLPSKDTSVVSLQDR
P640-3960       LGVNAVHAKAQVERNAGLIAESAKIAQSQ-NAFSVGTNEQKAAGKYLTSKDKSVVSLQDR
P588-8079       LGVNAVYAKAQVERNAGIIKESVELA--K-TAFTVAQDDEKAIPTYLTSKNTSVVSLQDR
P589-8275       LGVNAVYAKAQVERNAGIIKESVELA--K-TAFTVAQDDEKAIPTYLTSKNTSVVSLQDR
P667-4462       LGVNAVYANAQVERNAGIIADTAKMSQG---AFNVGSEAEKLIPGYLTSKDKSVVSLQDR
P599-8624       LGVNAVYAKAQVERNAGIIADSVKDDQAK-ILFTVGSPEDKAIPTYLTSKDKSVVSLQDR
P625-8065       LGVNAVYAKAQVERNAGIIADTAKISQ---DAFTVGSLEDKEIPKYLTSKDKSVVSLQDR
P598-8560       LGVNAVYAKAQVERNAGIITESVKI--AQ-GALNTVVP-GTKIPGYLTSKDKSVVSLQDR
P610-7579       LGVNAVYAKAQVERNAGIITDSVKI--AQ-DALKVVKP-KTVIPDYLTSKDKSVVSLQDR
NTHi375         LGVNAVYAKAQVERNAGIITDSVKI--AQ-NALKIVKP-KTVIPDYLTSKDKSVVSLQDR
P591-8654       LGVNAVYANAQVERNAGIIADSVQDGKVK-NALKTVAP-GTTIPDYLTSKDKSVVSLQDR
P593-7974       LGVNAVYAKAQVERNAGIITDSIQDQQIK-RALTTVDP-QTKIHEYLTSKDKSVVSLQDR
```

```
P604-7629       LGVNAVYAKAQVERNAGIIVDSVKDQQVQ-LALKQVNS-QTKIPNYLTSKDKSVVSLQDR
P641-4342       LGVNAVYAKAQVERNAGIIVDSVKDQQVQ-RALKQVNS-QTKIPNYLTSKDKSVVSLQDR
P636-8296       LGVNAVYANAQVERNAGIIVDSVQDKQIQ-QALKAVDS-QTKIHEYLTSKDKSVVSLQDR
                **:***:*:*::.*.** : .          :             * *.: ***.***

P600-8643       SAWGFGWNAGLMYQFNENNRMGLAYHSAVDIDFTDYTATGLSAGK--SSKESSLTLRLPD
P674-1509       AAWGFGWNAGVMYQFNEGNRIGLAYHSKVDIDFTDRTATSLKAGVIKAGEKGNLTLKLPD
P606-8650       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLEANVIKAGKKGDLTLTLPD
P630-7232       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFADRTATSLEAEAIKAGKKGDLTLTLPD
P660-9367       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLEAEAIKTAKKGDLTLTLPD
P657-8759       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLEAEAIKAGKKGDLTLTLPD
P658-8889       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLEAEAIKAGKKGDLTLTLPD
P646-8484       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSL-----GVGKKGDLTLTLPD
P671-7552       AAWGFGWNAGVMYQFNEGNRIGLAYHSKVDIDFTDRTATSL-----GVGKKGDLTLTLPD
P643-7915       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLEAGAIGAGKKGNLTLTLPD
P676-2514       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLEAGVIGAGKTGNLTLTLPD
P608-8895       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFVDRTATSLEAGVIGAGKKGDLTLTLPD
P666-8053       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSVEANVIKAGKKGDLTLTLPD
P670-7113       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSVEANVIKEGKKGNLTLTLPD
P611-7680       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLEANVIKAGKKGDLTFTLPD
P627-3766       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLEANVIKAGKKGDLTLTLPD
P661-4460       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFADRTATSLEANVIKEGKKGNLTFTLPD
P609-9724       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSVEANVIKEGKKGNLTFTLPD
P634-7311       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLEANVIKASKTGDLTLTLPD
P638-3144       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLEAEAIKAGKKGDLTLTLPD
P619-9590       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFADRTATSVEADVIKAGKKGDLTLTLPD
HiRdKW20        AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSVEANVIKAGKKGDLTLTLPD
P640-3960       AAWGFGWNAGVMYQFNEGNRIGLAYHSKVDIDFTDRTATSLEAGLIEAGKKGDLTLTLPD
P588-8079       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFVDRTATSLEAGVIGAGKKGDLTLTLPD
P589-8275       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFVDRTATSLEAGVIGAGKKGDLTLTLPD
P667-4462       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLGAGVIKAGKTGNLTLTLPD
P599-8624       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLEAGAIGAGKKGDLTLTLPD
P625-8065       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLEAGAIGAGKKGDLTLTLPD
P598-8560       AAWGFGWNAGVMYQFNEGNRIGLAYHSKVDIDFTDRTATSLGAKDIEAGKTDNLTLTLPD
P610-7579       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLESEVIEAGKKGDLTLTLPD
NTHi375         AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLESEVIEAGKKADLTLTLPD
P591-8654       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFSDRTATSLGAKDIEAGKTGDLTLALPD
P593-7974       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLGKKDIVAGKTGNLTFTLPD
P604-7629       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTASSLGKNDIVAGKTGNLTFTLPD
P641-4342       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTASSLGKNDIVAGKTGNLTFTLPD
P636-8296       AAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLGKKDIVAGKTGNLTFTLPD
                :*********:****** **:****** ***** * **:.:       .:  .**: ***

P600-8643       YLEFSGLHQVTNRFAIHYSYKYTHWSHLNKLYARSNDGKTAFEKELQYSNSSRVAFGGTY
P674-1509       YLELSGFHQLTNKFAVHYSYKYTHWSRLTRLHASFEDGKEAFDKELQYSNNSRVALGASY
P606-8650       YLELSGFHQLTDKFAVHYSYKYTHWSRLTKLYASFENGKKAFDKELQYSNNSRVALGASY
P630-7232       YLELSGFHQLTDKLAVHYSYKYTHWSRLTRLYASSENGKKAFDKELQYSNNSRVALGASY
P660-9367       YLELSGFHQLTDKLAVHYSYKYTHWSRLTRLYASSENGKKAFDKELQYSNNSRVALGASY
P657-8759       YLELSGFHQLTDKLAVHYSYKYTHWSRLTRLYASSENGKKAFDKELQYSNNSRVALGASY
P658-8889       YLELSGFHQLTDKLAVHYSYKYTHWSRLTRLYASSENGKKAFDKELQYSNNSRVALGASY
P646-8484       YLELSGFHQLTDELAVHYSYKYTHWSRLTRLYASSENGKKAFDKELQYSNNSRVALGASY
P671-7552       YLELSGFHQLTDKLAVHYSYKYTHWSRLTRLYASSENGKKAFDKELQYSNNSRVALGASY
P643-7915       YLELSGFHQLTDKLAVHYSYKYTHWSRLTRLYASSENGKKAFDKELQYSNNSRIALGASY
P676-2514       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLHASFEDGKKAFDKELQYSNNSRIALGASY
P608-8895       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLNASFEDGKKAFDKELQYSNNSRVALGASY
P666-8053       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLHASFEDGKKAFDKELQYSNNSRVALGASY
P670-7113       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLNANFEDGKKAFDKELQYSNNSRVALGASY
P611-7680       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLHASFEDGKKAFEKELQYSNNSRVALGASY
P627-3766       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLNASFENGKKAFDKELQYSNNSRVALGASY
P661-4460       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLHASFENGKKAFDKELQYSNNSRVALGASY
P609-9724       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLHASFEDGKKAFDKELQYSNNSRVALGASY
P634-7311       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLNASFEDGKKAFDKELQYSNNSRVALGASY
P638-3144       YLELSGFHQLTDKLAVHYSYKYTHWSRLTRLYASSENGKKAFDKELQYSNNSRVALGASY
P619-9590       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLNASFEDGKKAFDKELQYSNNSRIALGASY
HiRdKW20        YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLNASFEDGKKAFDKELQYSNNSRVALGASY
P640-3960       YLELSGFHQLTDKLAVHYSYKYTHWSRLTRLYASSENGKKAFDKELQYSNNSRVALGASY
P588-8079       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLNASFEDGKKAFDKELQYSNNSRVALGASY
P589-8275       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLNASFEDGKKAFDKELQYSNNSRVALGASY
P667-4462       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLNASFEDGKKAFDKELQYSNNSRVALGASY
P599-8624       YLELSGFHQLTDKFAVHYSYKYTHWSRLTRLYASSENGKKAFDKELQYSNNSRIALGASY
P625-8065       YLELSGFHQLTDKFAVHYSYKYTHWSRLTRLYASSENGKKAFDKELQYSNNSRIALGASY
P598-8560       YLELSGFHQLTDKFAVHYSYKYTHWSRLTKLHASFENGKKAFDKELQYSNNSRVALGASY
P610-7579       YLELSGFHQLTDKFAVHYSYKYTHWSRLTKLHASFEDGKKAFDKELQYSNNSRVALGASY
NTHi375         YLELSGFHQLTDKFVVHYSYKYTHWSRLTKLHASFEDGKKAFDKELQYSNNSRVALGASY
P591-8654       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLHASYENGKKAFEKELQYSNNSRVALGASY
P593-7974       YLELSGFHQLTDKFAVHYSYKYTHWSRLTKLHASYENGEKAFDKELQYSNNSRVALGASY
P604-7629       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLHASYENGEKAFDKELQYSNNSRVALGASY
P641-4342       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLHASYENGEKAFDKELQYSNNSRVALGASY
P636-8296       YLELSGFHQLTDKLAVHYSYKYTHWSRLTKLNASYENGEKAFDKELQYSNNSRVALGASY
```

```
                ***:**:**:*:.:.:**********:*.:* *  ::*: **:*******.**:*:*.:*

P600-8643       DIDDKLTLRAGVAYDQAASRHYRSASIPDTDRTWYSLGATYSFTPNFSVDFGYAYLKGKK
P674-1509       NLDEKLTLRAGIAYDQAASRHHRSAAIPDTNRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P606-8650       NLDEKLTLRAGIAYDQAASRHHRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P630-7232       NLDEKLTLRAGIAYDQAASRHQRSAAIPDTNRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P660-9367       NLDEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P657-8759       NLDEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P658-8889       NLDEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P646-8484       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPHLSVDLGYAYLKGKK
P671-7552       NLYEKLTLRAGIAYDQAASRHHRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P643-7915       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P676-2514       DLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P608-8895       DLYEKLTLRAGIAYDQAASRHHRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P666-8053       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P670-7113       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P611-7680       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P627-3766       NLYEKLTLRAGIAYDQAASRHHRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P661-4460       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P609-9724       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P634-7311       NLDEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P638-3144       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P619-9590       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
HiRdKW20        NLDEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P640-3960       DLYEKLTLRAGIAYDQAASRHHRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P588-8079       DLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P589-8275       DLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P667-4462       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P599-8624       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P625-8065       NLYEKLTLRAGIAYDQAASRHHRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P598-8560       NLDEKLTLRAGIAYDQAASRHHRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P610-7579       NLDEKLTLRAGIAYDQAASRHHRSAAIPDTNRTWYSLGTTYKFTPNLSVDLGYAYLKGKK
NTHi375         NLDEKLTLRAGIAYDQAASRHHRSAAIPDTNRTWYSLGTTYKFTPNLSVDLGYAYLKGKK
P591-8654       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P593-7974       NLDEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGGTYKFTPNLSVDLGYAYLKGKK
P604-7629       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P641-4342       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
P636-8296       NLYEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYLKGKK
                :: :*******:********* ***:****:******* **.***::***:**********

P600-8643       VNFKEEAKLAKD----QLTTTANYTSKAHANLYGLNLNYSF*
P674-1509       VHFKEVNAIGDE-RTLTINTTANYTSQAHANLYGLNLNYSF*
P606-8650       VHFKEVKAIDGKGATLTIDTTANYTSQAHANLYGLNLNYSF*
P630-7232       VHFKEVKAITK--SVLTLDTTANYTSQAHANLYGLNLNYSF*
P660-9367       VHFKEVKAIAK--APLTLDTTANYTSQAHANLYGLNLNYSF*
P657-8759       VHFKEVKAIAK--APLTLDTTANYTSQAHANLYGLNLNYSF*
P658-8889       VHFKEVKAIAK--APLTLDTTANYTSQAHANLYGLNLNYSF*
P646-8484       VHFKEVNTIGDE-RSLTFNTTANYTSQAHANLYGLNLNYSF*
P671-7552       VHFKEVNAIGDE-RTLTVNTTANYTSQAHANLYGLNLNYSF*
P643-7915       VHFKEEKTIAEK-SGLPLTSTANYTSQAHANLYGLNLNYSF*
P676-2514       VHFKEVKTIGKE-STLPVTATANYTSQAHANLYGLNLNYSF*
P608-8895       VHFKEVKEIGDK-SGLPLTTTANYTSQAHANLYGLNLNYSF*
P666-8053       VHFKEVKTIGKE-SGLPLTTTANYTSQAHANLYGLNLNYSF*
P670-7113       VHFKEVKAIGQA-P-STLDTTANYTSQAHANLYGLNLNYSF*
P611-7680       VHFKEVKTIGEP-RALTFNTTANYTSQAHANLYGLNLNYSF*
P627-3766       VHFKEVKTIGEQ-RALTFNTTANYTSQAHANLYGLNLNYSF*
P661-4460       VHFKEVKAIGDE-RTLTFNTTANYTSQAHANLYGLNLNYSF*
P609-9724       VHFKEVKTIGQK-NVLELNTTANYTSQAHANLYGLNLNYSF*
P634-7311       VHFKEVKTIGDK-RSLALSTTANYTSQAHANLYGLNLNYSF*
P638-3144       VHFKEVKTIGDA-RSLALNTTANYTSQAHANLYGLNLNYSF*
P619-9590       VHFKEVKTIGDE-RSLALNTTANYTSQAHANLYGLNLNYSF*
HiRdKW20        VHFKEVKTIGDE-RSLTLNTTANYTSQAHANLYGLNLNYSF*
P640-3960       VHFKEAQKVAGG----FITTTANYTSQAHANLYGLNLNYSF*
P588-8079       VHFKEVKTTGKT---VILNATADYTSQAHANLYGLNLNYSF*
P589-8275       VHFKEVKTTGKT---VILNATADYTSQAHANLYGLNLNYSF*
P667-4462       VHFKEVATLGKT---VIVNTTADYTSQAHANLYGLNLNYSF*
P599-8624       VHFKEVATIGKN---VTLNTTANYTSQAHANLYGLNLNYSF*
P625-8065       VHFKEVATIGKN---VILNTTANYTSQAHANLYGLNLNYSF*
P598-8560       VHFKEVQQAVGG---H-IITTANYTSQAHANLYGLNLNYSF*
P610-7579       VHFKEVQKVADG---Q-IITTANYTSQAHANLYGLNLNYSF*
NTHi375         VHFKEVQKVADG---Q-ITTTANYTSQAHANLYGLNLNYSF*
P591-8654       VHFKEVQQAVGG---Q-IITTANYTSQAHANLYGLNLNYSF*
P593-7974       VHFKEVQKAAGG---H-ITTTANYTSQAHANLYGLNLNYSF*
P604-7629       VHFKEAQEVAGG---Q-IITTANYTSQAHANLYGLNLNYSF*
P641-4342       VHFKEAQEVAGG---Q-IITTANYTSQAHANLYGLNLNYSF*
P636-8296       VHFKETQEAASG---Q-IITTANYTSQAHANLYGLNLNYSF*
                *:***           . :**:***:***************
```

Truncated FadL variants:

```
P662      MKKFNQSLFSNCNVVGCRWCKCGGFPIGGSFYFWAWS-CLCR*-----------------
P639      MKKFNQSILATAMLLAA-----GGANAAAFLYFWAWS-CLCR*-----------------
P602      MKKFNQSILATAMLLAA-----GGANAAAFQLAEVSTSGLGRAYAGEAAIADNAAVIATN
P614      MKKFNQSILATAMLLAA-----GGANAAAFQLAEVSTSGLGRAYAGEAAIADNAAVIATN
P590      MKKFNQSILATAMLLAA-----GGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATN
P851      MKKFNQSLLATAMLLAA-----GGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATN
P664      MKKFNQSLLATAMLLAA-----GGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATN
P665      MKKFNQSLLATAMLLAA-----GGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATN
P656      MKKFNQSLLATAMLLAA-----GGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATN
P626      MKKFNQSLLATAMLLAA-----GGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATN
P617      MKKFNQSLLATAMLLAA-----GGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATN
P607      MKKFNQSLLATAMLLAA-----GGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATN
P597      MKKFNQSLLATAMLLAA-----GGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATN
P592      MKKFNQSLLATAMLLAA-----GGANAAAFQLAEVSTSGLGRAYAG*-------------
P594      MKKFNQSLLATAMLLAA-----GGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATN
P595      MKKFNQSLLATAMLLAA-----GGANAAAFQLAEVSTSGLGRAYAGEAAIADNASVVATN
            *******:::.. ::..     **   ..     . :  * *

P662      ------------------------------------------------------------
P639      ------------------------------------------------------------
P602      PALMTVFKRAQFSAGG*--------------------------------------------
P614      PALMTVFKRAQFSAGGVYINSKVDMRGDVATNINNLTG-----KGKKGSASKNDVVPEAF
P590      PALMSLFKTAQFSTGGVYVDSRINMNGDVESSIKNF----QMLAAKYGSASARNVVPGAF
P851      PALMSLFKTAQFSTGGVYVDSRINMNGDVDASIK-----ATMNMTKYGSASQRNVVPGAF
P664      PALMSLFKTAQFSTGGVYVDSRINMNGNVASSVTT---NSTMKTTMDGSASERNVVPGAF
P665      PALMSLFKTAQFSTGGVYVDSRINMNGNVASSVTT---NSTMKTTMDGSASERNVVPGAF
P656      PALMSLFKTAQFSTGGVYVDSRINMNGDVDSYAIISN---SIKVTNDGSASARNVVPGAF
P626      PALMSLFKTAQFSTGGVYVDSRINMNGNVDASIT----GTGMAATKYGSVSERNVVPGAF
P617      PALMSLFKTNQFSVGGVYVDSRINMNGDVDSYAIISD---SIKVTNDGSASARNVVPGAF
P607      PALMSLFKTAQFSTGGVYVDSRINMNGDVASSIRTS----SKQVTKGGSASERNVVPGAF
P597      PALMSLFKTAQFSTGGVYVDSRINMNGDVTSSAATNTQRSKMIATKYGSASERNVVPGAF
P592      ------------------------------------------------------------
P594      PALMSLFKTAQFSTGGVYVDSRINMNGDVTSYATILTSSTGMKAIKEGSASARNVVPGAF
P595      PALMSLFKTAQFSTGGVYVDSRINMNGDVTSYATILTSSTGMKAIKEGSASARNVVPGAF


P662      ------------------------------------------------------------
P639      ------------------------------------------------------------
P602      ------------------------------------------------------------
P614      IPNLYFVAPVNDKLALGAGMNVNFGLKSEYGRDYDAGLFGGETKLSTINLN*--------
P590      VPNLYFVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTE
P851      VPNLYFVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGIFGGKTDLSAINLNLSGAYRVTE
P664      VPNLYFVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGAYRVTE
P665      VPNLYFVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLSAINLNLSGA*----
P656      VPNLYFVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLTAINLNLSGAYRVTE
P626      VPNLYFVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGIFGGKTDLSAINLNLSGAY*---
P617      VPNLYFVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGVFGGKTDLTAINLNLSGAYRVTE
P607      VPNLYFVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGIFGGKTDLSAINLNLSGAYRVTE
P597      VPNLYFVAPVNDKFALGAGMNVNFGLKSEYDDSYDAGIFGGKTDLSAINLNLSGAYRVTE
P592      ------------------------------------------------------------
P594      VPNLYFVAPVNDKFALGAGMNVNFGLKSEYDDSYNAGIFGGKTDLSAINLNLSGAYRVTQ
P595      VPNLYFVAPVNDKFALGAGMNVNFGLKSEYDDSYNAGIFGGKTDLSAINLNLSGAYRVTQ


P662      ------------------------------------------------------------
P639      ------------------------------------------------------------
P602      ------------------------------------------------------------
P614      ------------------------------------------------------------
P590      GLSLGLGVNAVYAKAQVERNAGIIKESVELA--KTAFTVAQDDEKAIPTYLTSKNTSVV*
P851      GLSLGLGVNAVYANAQVERNAGIIVDTVKDKQAQTALRAVDSK-TKIPDILTSKDKSVVS
P664      G*----------------------------------------------------------
P665      ------------------------------------------------------------
P656      GLSLGLGVNAVYAKAQVERNAGLIADTVKDSQVKTTLIVQQEPLKSIDKYLPSKDTSVVS
P626      ------------------------------------------------------------
P617      GLSLGLGVNAVYAKAQVERNAGLIADTVKDNQVKNTLTVQQEPLKFIDKYLPSKDTSVVS
P607      GLSLGLGVNAVYAKAQVERNAGTIVDTVNDTQVKPAFSVLGELYKKFP*-----------
P597      GLSLGLGVNAVYANAQVERNAGIIKDTVNDNQIKSALLT*--------------------
P592      ------------------------------------------------------------
P594      GLSLGLGVNAVYAKAQVERNAGIIADSVQDGQITQALTVL*--------------------
P595      GLSLGLGVNAVYAKAQVERNAGIIADSVQDGQITQALTVLQEPYKHLNQYLSSKDKSVVS


P662      ------------------------------------------------------------
P639      ------------------------------------------------------------
P602      ------------------------------------------------------------
P614      ------------------------------------------------------------
P590      ------------------------------------------------------------
```

```
P851    LQDRAAWGFGWNAGVMYQFNEANRIGLAYHSKV*-------------------------
P664    ------------------------------------------------------------
P665    ------------------------------------------------------------
P656    LQDRAAW*---------------------------------------------------
P626    ------------------------------------------------------------
P617    LQDRAAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLEANVIKASKTGDLTL
P607    ------------------------------------------------------------
P597    ------------------------------------------------------------
P592    ------------------------------------------------------------
P594    ------------------------------------------------------------
P595    LQDRAAWGFGWNAGVMYQFNEANRIGLAYHSKVDIDFTDRTATSLEAGAIKAGKKGDLTL

P662    ------------------------------------------------------------
P639    ------------------------------------------------------------
P602    ------------------------------------------------------------
P614    ------------------------------------------------------------
P590    ------------------------------------------------------------
P851    ------------------------------------------------------------
P664    ------------------------------------------------------------
P665    ------------------------------------------------------------
P656    ------------------------------------------------------------
P626    ------------------------------------------------------------
P617    TLPDYLELSGFHQLTDKLAVHYSYKYTHWSRLTKLNASFEDGKKAFDKELQYSNNSRVAL
P607    ------------------------------------------------------------
P597    ------------------------------------------------------------
P592    ------------------------------------------------------------
P594    ------------------------------------------------------------
P595    TLPDYLELSGFHQLTDKLAVHYSYKYTHWSRLTRLYASSENGKKAFDKELQYSNNSRVAL

P662    ------------------------------------------------------------
P639    ------------------------------------------------------------
P602    ------------------------------------------------------------
P614    ------------------------------------------------------------
P590    ------------------------------------------------------------
P851    ------------------------------------------------------------
P664    ------------------------------------------------------------
P665    ------------------------------------------------------------
P656    ------------------------------------------------------------
P626    ------------------------------------------------------------
P617    GASYNLDEKLTLRAGIAYDQAASRHQRSAAIPDTDRTWYSLGATYKFTPNLSVDLGYAYL
P607    ------------------------------------------------------------
P597    ------------------------------------------------------------
P592    ------------------------------------------------------------
P594    ------------------------------------------------------------
P595    GASYNLDEKLTLRAGIAYDQAASRHQRSAAIPDTDRT*----------------------

P662    ----------------------------
P639    ----------------------------
P602    ----------------------------
P614    ----------------------------
P590    ----------------------------
P851    ----------------------------
P664    ----------------------------
P665    ----------------------------
P656    ----------------------------
P626    ----------------------------
P617    KGKKVHFKEVKTIGDKRSLALSTTANYT*
P607    ----------------------------
P597    ----------------------------
P592    ----------------------------
P594    ----------------------------
P595    ----------------------------
```

File S3.2. Alignment of FolH (upper) and FolP (lower) variants found in the COPD NTHi strain collection under study. Variants are indicated with numbers (FolH_1, etc.). S=SXT sensitive; R=SXT resistant. Amino acids found to be present in all SXT^S strains are indicated in **bold blue**. Amino acid transitions likely to be associated to acquisition of SXT resistance are indicated in **bold black**.

Alignment of FolH variants:

```
S-P603_FolH_11      MTFSLIVATTLNSVIGKDNQMPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
R-P671_FolH_22      MTFSLIVATTLNSVIGKDNQMPWHLPADLARFRQNTTGKPVIMGRKTFESIGRPLPKRTN
S-P640_FolH_19      MTFSLIVAITLNNVIGKDNQMPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
R-P599_FolH_8       MTFSLIVATTLNSVIGKDNQMPWHLPADLARFRQNTTGKPVIMGRKTFESIGRPLPKRTN
S-P636_FolH_13      MTFSLIVATTLNSVIGKDNQMPWHLPADLARFRQNTTGKPVIMGRKTFESIGRPLPKRTN
R-P641_FolH_2       MTFSLIVATTLNSVIGKDNQMPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
R-P592_FolH_5       MTFSLIVATTLNSVIGKDNQMPWHLPADLAWFRQNTTGKPIIMGRKTFESIGRPLPKRTN
R-P619_FolH_15      MTFSLIVATTLNNVIGKDNQMPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
R-P600_FolH_10      MTFSLIVATTLNSVIGKDNQIPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
R-P609_FolH_9       MTFSLIVATTLNSVIGKDNQIPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
R-P634_FolH_12      MTFSLIVATTLNSVIGKDNQIPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
S-P588_FolH_1       MTFSLIVATTLNNVIGKDNQMPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
S-P638_FolH_16      MTFSLIVATTLNNVIGKDNQMPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
S-P656_FolH_20      MTFSLIVATTLNSVIGKDNQMPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
S-P662_FolH_7       MTFNLIVATTLNNVIGKDNQMPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
S-P626_FolH_18      MTFSLIVATTLNNVIGKDNQMPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
S-P667_FolH_14      MTFSLIVATTLNNVIGKDNQMPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
S-P594_FolH_6       MTFSLIVATTLNNVIGKDNQMPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
S-P589_FolH_3       MTFSLIVATTLNNVIGKDNQMPWHLPADLAWFRQNTTGKPVIMGRKTFESIGRPLPKRTN
                    ***.**** ***.*******:********* *********:*******************

S-P603_FolH_11      IILSRQPFKHEGVVWKDSLESAVDFVRDFDEIMLIGGGELFKQYLPKSDKLYLTQIQTEL
R-P671_FolH_22      IVLSRQPFKHEGVVWKDSLESAVNFVRDFDEIMLIGGGELFKQYFPQADKLYLTQIQTEL
S-P640_FolH_19      IVLSRQPFKHEGVVWKDSLESAVDFVRDFDEIMLIGGGELFKQYLPKADKLYLTQIQTEL
R-P599_FolH_8       IVLSRQPFKHEGVVWKDSLESAVNFVRDFDEIMLLGGGELFKQYLPQADKLYLTQIQTEL
S-P636_FolH_13      IVLSRQPFKHEGVVWKDSLESAVDFVRDFDEIMLIGGGELFKQYLPKADKLYLTQIQTEL
R-P641_FolH_2       IVLSRQPFEHEGVVWKDSFESAVDFVRDFDEIMLIGGGELFKQYLPQADKLYLTQIQAEL
R-P592_FolH_5       IVLSRQPFEHEGVVWKDSFESAVNFVRDFDEIMLIGGGELFKQYLPQADKLYFTQIQAEL
R-P619_FolH_15      IVLSRQPFEREGVVWKDSLESAVYFVRDFDEIMLLGGGELFKQYLPQADKLYLTQIQTEL
R-P600_FolH_10      IVLSRQLFEHEGVIWKDSFESAVNFVRDFDEIMLLGGGELFKQYLPKADKLYLTQIQTEL
R-P609_FolH_9       IVLSRQLFEHEGVIWKDSFESAVNFVRDFDEIMLLGGGELFKQYLPKADKLYLTQIQTEL
R-P634_FolH_12      IVLSRQLFEHEGVIWKDSFESAVNFVRDFDEIMLLGGGELFKQYLPKADKLYLTQIQTEL
S-P588_FolH_1       IVLSRQLFEHDGVVWKDSLESAVNFVRDFDEIMLIGGGELFKQYLPKADKLYLTQIQTEL
S-P638_FolH_16      IVLSRQLFEHDGVVWKGSLESAVNFVRDFDEIMLIGGGELFKQYLPKADKLYLTQIQTEL
S-P656_FolH_20      IVLSRQLFEHEGVIWKNSLESAVNFVRDFDEIMLIGGGELFKQYLPQADKLYLTQIQTEL
S-P662_FolH_7       IVLSRQLFEHEGVIWKDSFESAVNFVRDFDEIMLIGGGELFKQYLPKADKLYLTQIQTEL
S-P626_FolH_18      IVLSRQLFEYEGVIWKDSFESAVNFVRDFDEIMLIGGGELFKQYLPKADKLYLTQIQTEL
S-P667_FolH_14      IVLSRQLFEHEGVIWKDSFESAVNFVRDFDEIMLIGGGELFKQYLPKADKLYLTQIQTEL
S-P594_FolH_6       IVLSRQLFEHEGVIWKNSFESAVNFVRDFDEIMLIGGGELFKQYLPKADKLYLTQIQTEL
S-P589_FolH_3       IVLSRQLFEHDGVIWKDSFESAVNFVRDFDEIMLIGGGELFKQYLPKADKLYLTQIQTEL
                    *:**** *: :**:**.*:**** **********:*********:*::****:****:**

S-P603_FolH_11      DGDTFFPQLNWEEWKIEFDEYRKADEKNRYDCRFLILTRK*
R-P671_FolH_22      DGDTFFPQLNWEEWKIEFDEYRKADEQNRYDCRFLILIRK*
S-P640_FolH_19      DGDTFFPQLNWEEWKIEFDEYRKADEQNRYDCRFLILTRK*
R-P599_FolH_8       DGDTFFPQLNWEEWKIEFDEYRKADEQNRYDCRFLILTRK*
S-P636_FolH_13      DGDTFFPQLNWEEWKIEFDEYRKADEQNRYDCRFLILTRK*
R-P641_FolH_2       DGDIFFPQLNWDEWTIEFDEYRKADEQNRYDCRSLILTRK*
R-P592_FolH_5       DGDTFFPQLNWDEWKIEFDEYREADEQNRYDCRSLILTRK*
R-P619_FolH_15      DGDTFFPQLNWEEWEIEFDEYRKADEQNRYDCRVLILTRK*
R-P600_FolH_10      DGDTFFPQLNWDEWEIEFDEYRKADEQNHYDCRFLILTRK*
R-P609_FolH_9       DGDTFFPQLNWEEWEIEFDEYRKADEQNRYDCRFLILTRK*
R-P634_FolH_12      DGDTFFPQLNWEEWKIEFDEYRKADEQNRYDCRFLILTRK*
S-P588_FolH_1       DGDTFFPQLNWEEWEIEFDEYRKADEQNRYDCRFLILTRK*
S-P638_FolH_16      DGDTFFPQLNWEEWEIEFDEYRKADEQNRYDCRFLILTRK*
S-P656_FolH_20      DGDTFFPQLNWEEWKIEFDEYRKADEQNRYDCRFLILTRK*
S-P662_FolH_7       DGDTFFPQLNWEEWKIEFDEYRKADEQNRYDCRFLILTRK*
S-P626_FolH_18      DGDTFFPQLNWEEWEIEFDEYRKADEQNRYDCRFLILTRK*
S-P667_FolH_14      DGDTFFPQLNWEEWEIEFDEYRKADAQNRYDCRFLILTRK*
S-P594_FolH_6       DGDTFFPQLNWEEWEIEFDEYRKADEQNRYDCRFLILTRK*
S-P589_FolH_3       DGDTFFPQLNWEEWEIEFDEYRKADEQNRYDCRFLILTRK*
                    *** *******:** *******:** :*:**** *** ***
```

Alignment of FolP variants:

```
R-P609_FolP_11      MKLYANNKCLDLSVPNIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGAKIIDIGGE
R-P599_FolP_10      MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
R-P851_FolP_27      MKLYANNKCLDLSVPKIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
R-P671_FolP_25      MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
S-P640_FolP_20      MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGHFFSLDKALFQVEKMLEEGATIIDIGGE
S-P597_FolP_8       MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGAAIIDIGGE
S-P656_FolP_22      MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
S-P607_FolP_14      MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
R-P606_FolP_13      MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
S-P591_FolP_4       MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
S-P626_FolP_18      MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
S-P593_FolP_6       MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGAVIIDIGGE
S-P638_FolP_16      MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
S-P646_FolP_21      MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
S-P589_FolP_3       MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
S-P674_FolP_26      MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
S-P603_FolP_12      MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
S-P639_FolP_19      MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
S-P661_FolP_24      MKLYANNKCLDLSVPKIMGILNFTPDSFSDSGQFFSLDKALFQVEKMLEEGATIIDIGGE
R-P641_FolP_2       MKLYANNKCLDLSVPQIMGILNFTPDSFSDRGQFFSLDKALFQVEKMLEEGATIIDIGGE
R-P619_FolP_15      MKLYANNKCLDLSVPQIMGILNFTPDSFSDRGQFFSLDKALFQVEKMLEEGATIIDIGGE
R-P592_FolP_5       MKLYANNKCLDLSVPQIMGILNFTPDSFSDRGQFFSLDKALFQVEKMLEEGATIIDIGGE
                    ***************:************** *:******************* *******

R-P609_FolP_11      SARP----NDADEVSEQEELHRVVPVVEAVRSRFDCWISVDSSKAVVMREAASVGMDLIN
R-P599_FolP_10      STRPSFLYNDADEVSEQEELHRVVPVVEAVRSRFDCWISVDSSKAVVMREAASVGMDLIN
R-P851_FolP_27      STRP----NDADEVSEQEELHRVVPVVEAVRSRFDCWISVDSSKAVVMREAASVGMDLIN
R-P671_FolP_25      STRP----NDADEVSEQEELHRVVPVVEAVRSRFDCWISVDSSKAVVMREAASVGMDLIN
S-P640_FolP_20      STRP-----NADEVSEQEELHRVVPVVEAVRNRFDCWISVDSSKAIVMREAANVGMDLIN
S-P597_FolP_8       STRP-----NADEVSEQEELHRVVPVVEAVRNRFDCWISVDSSKAIVMREAANVGMDLIN
S-P656_FolP_22      STRP-----NADEVSEQEELHRVVPVVEAVRNRFDCWISVDSSKAVVMREAANVGMDLIN
S-P607_FolP_14      STRP-----NADEVSEQEELHRVVPVVEAVRNRFDCWISVDSSKAIVMREAANVGMDLIN
R-P606_FolP_13      STRP-----NADEVSEQEELHRVVPVVEAVRNRFDCWISVDSSKAIVMREAANVGMDLIN
S-P591_FolP_4       STRP-----NADEVSEQEELHRVVPVVEAVRNRFDCWISVDSSKAIVMREAANVGMDLIN
S-P626_FolP_18      STRP-----NADEVSEQEELHRVVPVVEAVRNRFDCWISVDSSKAIVMREAANVGMDLIN
S-P593_FolP_6       STRS-----NADEVSEQEELHRVVPVVEAVRSRFDCWISVDSSKAVVMRESASVGMDLIN
S-P638_FolP_16      STRP-----NADEVSEQEELHRVVPVVEAVRSRFDCWISVDSSKAIVIREAASVGMDLIN
S-P646_FolP_21      STRS-----NADEVSEQEELHRVVPVVEAVRSRFDCWISVDSSKAIVIREAASVGMDLIN
S-P589_FolP_3       STRP-----NADEVSEQEELHRVVPVVEAVRSRFDCWISVDSSKAIVMREAASVGMDLIN
S-P674_FolP_26      STRS-----NADEVSEQEELHRVVPVVEAVRSRFDCWISVDSSKAIVMREAASVGMDLIN
S-P603_FolP_12      STRP-----NADEVSEQEELHRVVPVVEAVRSRFDCWISVDSSKAIVMREAASVGMDLIN
S-P639_FolP_19      STRP-----NADEVSEQEELHRVVPVVEAMRSRFDCWISVDSSKAIVMREAASVGMDLIN
S-P661_FolP_24      STRQ-----NADEVSEQEELHRVVPVVEAVRSRFDCWISVDSSKAVVMREAANVGMDLIN
R-P641_FolP_2       STRE-----NADEVSEQEELHRVVPVVEAVRSRFDCWISVDSSKAIVMREAANVGMDLIN
R-P619_FolP_15      STRE-----NADEVSEQEELHRVVPVVEAVRSRFDCWISVDSSKAVVMREAASVGMDLIN
R-P592_FolP_5       STRE-----NADEVSEQEELHRVVPVVEAVRSRFDCWISVDSSKAVVMREAASVGMDLIN
                    *:*      :************** ****:*.*************:*: *:*.*******

R-P609_FolP_11      DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQVNPHYENVVQDVLAFLQKRTNECLSA
R-P599_FolP_10      DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQANPHYENVVQDVLAFLQKRTNECLSA
R-P851_FolP_27      DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQANPHYENVVQDVFAFLQKRTNECLSA
R-P671_FolP_25      DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQANPHYENVVQDVFAFLQKRTNECLSA
S-P640_FolP_20      DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQANPYYENVVQDVLAFLQKRTNECLSA
S-P597_FolP_8       DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQANPYYENVVQDVLAFLQKRTNECLSA
S-P656_FolP_22      DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQANPYYENVVQDVLAFLQKRTNECLSA
S-P607_FolP_14      DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQANPYYENVVQDVLAFLQKRTNECLSA
R-P606_FolP_13      DIRALQEPNALETTVKLALPVCIMHMQGQPRTMQANPYYENVVQDVLAFLQKRTNECLSA
S-P591_FolP_4       DIRALQEPNALEMAVKLALPVCIMHMQGQPRTMQANPYYENVVQDVLAFLQKRTNECLSA
S-P626_FolP_18      DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQANPYYENVVQDVLAFLQKRTNECLSA
S-P593_FolP_6       DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQVNPHYENVVQDVFAFLQKRTNECLSA
S-P638_FolP_16      DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQVNPYYENVVQDVLAFLQKRTNECLSA
S-P646_FolP_21      DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQVNPYYENVVQDVLAFLQKRTNECLSA
S-P589_FolP_3       DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQANPYYENVVQDILAFLQKRTNECLSA
S-P674_FolP_26      DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQANPYYENVVQDILAFLQKRTNECLSA
S-P603_FolP_12      DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQVNPYYENVVQDVLAFLQKRTNECLSA
S-P639_FolP_19      DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQVNPYYENVVQDVLAFLQKRTNECLSA
S-P661_FolP_24      DVRALQEPNALETAVKLALPVCIMHMQGQPRTMQENPHYENVVQDVFAFLQKRTNECLSA
R-P641_FolP_2       DIRALQEPNALETAVKLALPVCIMHMQGQPRTMQVNPYYENVVQDVLAFLQKRTNECLSA
R-P619_FolP_15      DVRALQEPNALETAVKLALPVCIMHMQGQPRTMQVNPYYENVVQDVLAFLQKRTNECLSA
R-P592_FolP_5       DVRALQEPNALETAVKLALPVCIMHMQGQPRTMQVNPHYENVVQDVLAFLQKRTNECLSA
                    *:*********: :******************** **:*******::************

R-P609_FolP_11      GVKKENLIWDMGFCFGKSVQHNYQLLQNLNEFCHSGYPVLAGLSRKSMIGAVLDKPVDQR
R-P599_FolP_10      GVKKENLIWDMGFCFGKSVQHNYQLLQNLNEFCQNGYPVLAGLSRKSMIGAVLNKAVDQR
R-P851_FolP_27      GVKKENLIWDMGFCFGKSVQHNYQLLQNLNEFCHSGYPVLAGLSRKSMIGAVLNKAVDQR
R-P671_FolP_25      GVKKENLIWDMGFCFGKSVQHNYQLLQNLNEFCHSGYPVLAGLSRKSMIGAVLNKAVDQR
```

```
S-P640_FolP_20      GIKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCKNGYPVLAGLSRKSMIGAVLDKPVDQR
S-P597_FolP_8       GIKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCKNGYPVLAGLSRKSMIGAVLDKPVDQR
S-P656_FolP_22      GIKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCKNGYPVLAGLSRKSMIGAVLDKPVDQR
S-P607_FolP_14      GIKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCKNGYPVLAGLSRKSMIGAVLDKPVDQR
R-P606_FolP_13      GIKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCQNGYPVLAGLSRKSMIGAVLDKPVDQR
S-P591_FolP_4       GIKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCQNGYPVLAGLSRKSMIGAVLDKPVDQR
S-P626_FolP_18      GIKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCQNGYPVLAGLSRKSMIGAVLDKPVDQR
S-P593_FolP_6       GIKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCQNGYPVLAGLSRKSMIGAVLDKPVDQR
S-P638_FolP_16      GVKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCQNGYPVLAGLSRKSMIGAVLDKPVDQR
S-P646_FolP_21      GVKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCQNGYPVLAGLSRKSMIGAVLDKPVDQR
S-P589_FolP_3       GVKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCQNGYPVLAGLSRKSMIGAVLDKPVDQR
S-P674_FolP_26      GVKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCQNGYPVLAGLSRKSMIGAVLDKPVDQR
S-P603_FolP_12      GVKKENLIWDMGFGFGKSVQHNYQLLQNLNKFCQIGYPVLAGLSRKSMIGAVLDKPVDQR
S-P639_FolP_19      GVKKENLIWDMGFGFGKSVQHNYQLLQNLNKFCQIGYPVLAGLSRKSMIGAVLDKPVDQR
S-P661_FolP_24      GVKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCQNGYPVLAGLSRKSMIGAVLDKPVDQR
R-P641_FolP_2       GVKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCQNGYPVLAGLSRKSMIGAVLDKPVDQR
R-P619_FolP_15      GIKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCQNVYPVLAGLSRKSMIGAVLDKPVDQR
R-P592_FolP_5       GIKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCQNVYPVLAGLSRKSMIGAVLDKPVDQR
                    *:*********** ****************:**:  *****************:* ****

R-P609_FolP_11      VIGSAAGALIAAQKGAKILRVHDVAATSDMLKVWQATENA*
R-P599_FolP_10      VIGSAAGALIAAQKGAKILRVHDVAATSDMLKVWQATENA*
?-P851_FolP_27      VIGSAAGALIAAQKGAKILRVHDVAATSDMLKVWQATENA*
R-P671_FolP_25      VIGSAAGALIAAQKGAKILRVHDVAATSDMLKVWQATENA*
S-P640_FolP_20      IIGSAAGALIAVQKGAKILRVHDIAATSDMLKVWQATANA*
S-P597_FolP_8       IIGSAAGTLIAAQKGAKILRVHDVAATSDMLKVWQATENA*
S-P656_FolP_22      IIGSAAGALIAAQKGAKILRVHDVAATSDMLKVWQATENA*
S-P607_FolP_14      IIGSAAGALIAAQKGAKILRVHDVAATSDMLKVWQATENA*
S-P606_FolP_13      IIGSAAGALIAAQKGAKILRVHDVAATSDMLKVWQATENA*
S-P591_FolP_4       IIGSAAGALIAAQKGAKILRVHDVAATSDMLKVWQATENA*
S-P626_FolP_18      IIGSAAGALIAAQKGAKILRVHDVAATSDMLKVWQATENA*
S-P593_FolP_6       VIGSVAGALIAVQKGAKILRVHDVAATSDMLKVWQATENA*
S-P638_FolP_16      IIGSAVGALIAAQKGAKILRVHDVAATSDMLKIWQATENA*
S-P646_FolP_21      IIGSAVGALIAAQKGAKILRVHDVAATSDMLKIWQATENA*
S-P589_FolP_3       IIGSAAGALIATQKGAKILRVHDVAATSDMLKIWQATENA*
S-P674_FolP_26      IIGSAAGALIATQKGAKILRVHDVAATSDMLKIWQATENA*
S-P603_FolP_12      IIGSAAGALIATQKGAKILRVHDVAATSDILKIWQATENA*
S-P639_FolP_19      IIGSAAGTLIAVQKGAKILRVHDVAATSDMLKIWQATENA*
S-P661_FolP_24      VIGSAAGALIAAQKGAKILRVHDVAATSDMLKVWQATENA*
R-P641_FolP_2       VIGSAAGALIAAQKGAKILRVHDVASTSDMLKVWQATENA*
R-P619_FolP_15      IIGSAVGALIAAQKGAKILRVHDVAATSDMLKIWQATENA*
R-P592_FolP_5       VIGSAAGALIAAQKGAKILRVHDVAATSDMLKVWQATENA*
                    :***..*:***.***********:*:***:**:**** ***
```

File S3.3. DNA sequence of the hmw1A together with 800 bp upstream of the predicted ATG for this gene in strains P651-8849, P652-8881, P653-8956 and P654-8983 and its flanking regions. *hmw1A* start codon is highlighted in green, *hmw1A* stop codon is highlighted in red, *hmw1B* start codon is highlighted in pink. Primers used to perform Sanger sequencing on the *hmw1A* gene from these four strains are highlighted in turquoise.

AACTCAATTTCACAAATCGGTTGCTCAGATTCGCCTGCAATAATTTTGCCTTGATCGAACGCCACTTCAATTTTGGAT
TGTTGAAATTCAACTAACCAAAAAGTGCGGTTAAAATCTGTGGAGAAAATAGGTTGTAGTGAAGAACGAGGTAATTGT
TCAAAAGGATAAAGCTCTCTTAATTGGGCATTGGTTGGCGTTTCTTTTTCGGTTAATAGTAAATTATATTCTGGACGA
CTATGCAATCCACCAACAACTTTACCGTTGGTTTTAAGCGTTAATGTAAGTTCTTGCTCTTCTTGGCGAATACGTAAT
CCCATTTTTTGTTTAGCAAGAAAATGATCGGGATAATCATAATAGGTGTTGCCCAAAAATAAATTTTGATGTTCTAAA
ATCGTAAATTTTGCAAGATATTGTGGCAATTCAATACCTATTTGTGGCGAAATCGCCAATTTTAATTCAATTTCTTGT
AGCATAATATTTCCCACTCAAATCAACTGGTTAAATATACAAGATAATAAAAATAAATCAAGATTTTTGTGATGACAA
ACAACAATTACAACAGCTTCTTTGCAATCTATATGCAAATATTTAAAAAAATAGTATAAATCCGCCATATAAAATGGT
ATAATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTTTCATCTT
TCATCTTTCATCTTTCACATGAAATGATGAACCGAGGGGAGGGAGGGCAAGAATGAAGAGGGAGCTGAACGAACGCA
AATGATAAAGTAATTTAATTGTTCAACTAACCTTAGGAGAAAATATGAACAAGATATATCGTCTCAAATTCAGCAAAC
GCCTGAATGCTTTGGTTGCTGTGTCTGAATTGACACGGGGTTGTGACCATTCCACAGAAAAAGGCAGCGAAAAACCTG
CTCGCATGAAAGTGCGTCACTTGGCGTTAAAGCCACTTTCCGCTATATTACTATCTTTAGGTGTAACATCTATTCCAC
AATCTGTTTTAGCAAGCGGCTTACAAGGAATGGATGTAGTACACGGCACAGCCACTATGCAAGTAGATGGTAATAAAA
CCATTATCCGCAACAGTGTTGACGCTATCATTAATTGGAAGCAATTTAACATCGACCAAAATGAAATGGTGCAGTTTT
TGCAAGAAAACAACAACTCCGCCGTATTTAACCGCGTTACATCTAACCAAATTTCCCAATTAAAAGGGATTTTAGATT
CTAACGGACAAGTCTTTTTAATCAACCCAAATGGTATCACAATAGGTAAAGACGCAATTATTAACACCAATGGCTTTA
CGGCTTCTACGCTAGATATTTCTAACGAAAACATCAAAGCGCGTAATTTTACCCTTGAGCAAACCAAGGATAAAGCTC
TCGCTGAAATCGTGAATCACGGTTTAATTACCGTTGGTAAAGACGGTAGCGTAAACCTTATTGGCGGCAAAGTAAAAA
ATGAGGGTGTAATTAGCGTAAATGGCGGTAGCATTTCTTTACTTGCAGGGCAAAAAATTACCATCAGCGATATAATAA
ACCCAACCATTACTTACAGCATTGCTGCACCCGAAAACGAAGCGGTCAATCTGGGCGATATTTTTGCCAAAGGCGGTA
ACATTAATGTCCGTGCTGCCACTATTCGAAACCAAGGTAAACTTTCCGCTGATTCTGTAAGCAAAGATAAAAGCGGTA
ACATTATTCTTTCCGCCAAAGAGGGTGAAGCGGAAATTGGCGGCGTGATTTCCGCTCAAAATCAACAAGCCAAAGGTG
GTAAGTTGATGATTACAGGTGATAAAGTCACATTAAAAACAGGTGCAGTTATCGACCTTTCAGGTAAAGAAGGAGGAG
AAACTTATCTTGGCGGTGACGAGCGCGGCGAAGGTAAAAACGGCATTCAATTAGCAAAGAAAACCTCTTTAGAAAAAG
GCTCAACAATTAATGTATCAGGCAAAGAAAAAGGCGGGCGCGCTATTGTATGGGGCGATATTGCGTTAATTAATGGTA
ACATTAATGCTCAAGGCAGTGATATTGCTGAAACTGGTGGCTTTGTGGAAACGTCGGGGCATTATTTATCTATTGGTA
ATGATGCGGCTGTTGAGGCGAAAGAATGGTTGTTGGATCCGGATAATGTTACCATTAGCAATGGTAACGATGATCAAT
CTCAATTAAAGGACGATAGAGGCGATTCACCTAATAAAATTCTAGCAGATAATAAACACGTTTAATAACAAAACATTA
TCTACAGCTTTAGCTAAAGGTATTGGAGTTAATATTTCTGCTAAAAAGAAAGTAAATGTTACTGCAGATATTAATGTT
CATAATGGAACATTAACGTTACACTCAGAACAAGGTGGAGTTGAAATTAACGGTGATATTACCTCAGAACAAAATGGT
AATTTAACCATTAAAGCAGGTAACTGGGTTGATGTTCACAAAAATATCACGATTGGTACGGGGTTTTTAAATATTACC
GCTGGGGGTTCTGTAGCTTTTGAAAAGGCAGGGGGTGATAAAGGCCGTGCTGCAAGTGATGCTAAAATTGTTGCTCAG
GGTGTAATTACTGCAGGCAGTGGTCAAGATTTCCGTTTTAATAATGTATCACTAAATGGAACTGGTAGAGGGTTGAAA
TTCATTACAGCTAAAGGAAATAAAGGTAATTTCTCAGCCAAATTTGATGGTGTGTTAAATATTTCAGGTAATATTTCA
ATTAATCACACTGCTAATAATCAATTATCTTATTTTCATCGCCAGGGTTATACATATTGGAATTTGACGCAGCTTAAT
GTTGATTCTGATTCTAGTTTTTCTTTAACATCAATAAAAGACGCGATTAAAGTTGGGGGATATGACAATGCTAAAGAC
AAAAAACACACAGGAGGGATTGGTTTTACACGAGATACCATTTTCAATGTAAAACAAGGTGCTAGAGTTGATATTTCC
TACACCCTACCGATATCTCCAGTTAAAAATTCAAGAATTGCTGCTGTCAATTTTGACGGAAATATTACTGTTAAAGGA
GGAGGAGTTGTTAATTTAAAATTTAATGCGCTTAGCAATAACTACAAAACTCCAGGGGTGAACATTTCCTCTAGATTT
ATTAATGTTACCGAAGGTTCGCAGCTAAACATTACTGGTAGTATGCCATCTACAACCTTATTTAATGTAGCAAATGAT
TTAATCATAAACGCAACCAATTCATTTGTATCCATTAAAGAAATTGAAGGTACAGATACTCATCTTGATACAGGTTTA
AAGGTAAACGGTAATGTTACCATTAAAGGAGGGAATGTCACCCTAGGTTCAAATAAAGCTAAAACCAAATTTGATAAA
AATGTAACTGTTGAAAAGGGTGCAAATCTTACTCTTGCGTCAGCAAACTTTGGAAATCATAAGGGTGCTTTAACTGTA
GCTGGAAATATCAATACTCAAGGTAAGCTTGTTGCAACTGGCGATACTATTGATGTATCTGGTGATTTTACTGTTGGA
AACGATGCTACTTTTAATGGTAATACTAATAACAACCTAAACATTACCGGCAATTTTACCAACAACGGCACCTCCATA
ATTGATGTAAAAAAAGGGGCGGCAAAACTAGGTAATATTACCAATGAAGGTAGTTTAAATATTACTACTCACGCTAAC
ACTAATCAAAAAACCATTATTACCGGAAATATAACCAATAAAAAAGGCGACTTAAACATCAGGGATAATAAAAATAAT
GCCGAAATCCAAATTGGCGGCAATATCTCGCAAAAAGAAGGCAATCTCACAATTTCTTCTGATAAAGTCAATATTACC
AAGCAGATAACAATCAAAGCAGGCGTTAATGGGGAGAATTCTGATTCAGGCACAGAAAACAATGCCAATCTAACCATT
AAAACCAAAACGCTAGAATTAACAAACGACCTAAACATTTCAGGCTTTCATAAAGCAGAAATTACAGCTAAAGATAAC
AGTGATTTAATTATTGGCAAGGCTAGCAGTGATAGCGGTAATGCTGGTGCTCAAAAAGTAATCTTTGACAAGGTTAAA
GATTCAAAAATCTCAGCTGGCAATCACAATGTAACACTAAATAGCGAAGTGGAAACGTCTAATGGTAATAGTAATGCT
GCTGGCGATAGTAATGGTAACAACGCTGGTTTAACTATTTCCGCAAAAGATGTGGCAGTAAATAATAACATCACTTCT
CACAAGACAATAAATATCTCTGCCACAACAGGAAATGTAACAACCAAAGAAGGCACAACCATTAATGCGACCACAGGT
GGCGTGGAAGTAACCGCTAAAACAGGTGACATTAAAGGTGGAATTGAGTCCAAGTCTGGCGGAGTAACACTTACTGCA
ACCGGAGATACTCTTGCTGTAGGTAATATTTCGGGTAACACTGTTAGTGTTACTGCAAATAGCGGTACATTAACCACC
AAAGCAGACTCTACAATCAAAGGAACCGGGAGCGTAACCACCTTAAGTCAATCAGGCGATATCGGCGGTACGATTTCT
GGTAAGACAGTAAGTGTTACAGCAACCACCGACAGTTTAACTGTTAAAGGTGGCGCAAAAATTAATGCGACAGAAGGA
ACTGCAACCTTAACTGCATCATCGGGCAAATTAACCACCGAGGCCAACTCTGCGATTAGCGGGGCTAACGGTGTAACT
GCCTCAAGTCAATCAGGCGATATTAGCGGTACGATTTCCGGTAAGACAGTAAGTGTTACAGCAACCACCGACAGTTTA
ACTGTTAAAGGTGGCGCAAAAATTAATGCGACAGAAGGAACTGCAACCTTAACTGCATCATCGGGCAAATTAACCACC

GAGGCCAACTCTGCGATTAGCGGGGCTAACGGTGTAACTGCCTCAAGTCAATCAGGCGATATTAGCGGTACGATTTCC
GGTAAGACAGTAAGTGTTACAGCAACCACCGACAGTTTAACTGTTAAAGGTGGCGCAAAAATTAATGCGACAGAAGGA
GCTGCGACTTTAACTGCAACAAAAGGCACTTTAACTACCGTGAAGGGTTCAAACATTGACGCAAACAAAGGCACCTTA
GTTATTAACGCAAAAGACGCCACACTAAATGGTGATGCATCAGGCGACCGTACAGAAGTGAATGCAGTCAACGCAAGC
GGCTCTGGTAACGTAACTGCGGCAACCTCAAGCAGCGTGAATATCACCGGGGATTTAAACACAATAAATGGGTTAAAT
ATCATTTCGAAAAATGGTAAAAACACCGTAGTGTTAAAAGGTGCTGAAATTGATGTGAAATATATTCAACCAGGTGTA
GCAAGTGCGAATGAGGTTATTGAAGCGAAGCGTGCCCTTGAAAAAGTAAAAGATTTATCTGATGAAGAAAGAGAAACA
TTAGCTAAACTTGGTGTAAGTGCTGTACGTTTTATTGAACCAAATAATACCATTACGGTTAACACACAAAATGAGTTT
ACAACCAGACCATCAAGTCAAGTGACAATTTCTGAAGGTAAGGCGTGTTTCTCAAGTGGTAATGGCGCAGCAGTATGT
ACCAATGTTGCTGACGATGGACAGCAGTAGTCAGTAATTGACAAGGTAGATTTCATCCTGCAATGAAGTCATTTTATT
TTCGTATTATTTACTGCGTGGGTTAAAGCTCAGTGTGGGCTTTTTACCCACCTTGTAAAAAATTACGGAAAATGCAAT
AAAATATTTTTAACAGGTTATTATTATGAAAAACATAAAAAGCAAATTAAAGCTCAGTGCAATATCAGTATTGCTTGG
TTTAGCGTCTTTATCAGCGCATACGGAAGAAGCATTTTTAGTAAAAGGTTTTCAGCTATCTGGCGCACTTGAAACTTT
AAGTGAAAATGCCCAATTATCTGTAGCAAAATCTCTATCTAAATATCAAGGTACGCAAACTTTAACAAGTCTAAAAAC
AGCACAACTTGAATTACAAGCTGTGCTAGATAAGATTGAGCCAAATAAATTTGATGTGGTATTACCACAACAAACCAT
TACAGATGGTAATGTTATGTTTGAACTGGTCTCAAAATCAGTCGCAGAAAGCCAAGTTTTTTATAAGGCAAGTAAGGG
TTATAGTGAAGAAAATATCGCTCGCAGCCTGCCATCTTTAAAACAAGGAAAAATGTATGAAGATGGTCGTCAGTGGTT
CGATTTGCGTGAATTCAATATGGCAAAAGAAAATCCACTTAAAGTCACTCGCGTGCATTACGAGTTAAACCCTAAAAA
CAAAACCTCTGATTTGGTAGTTGCAGGTTTTTCGCCTTTTGGCAAAACGCGCAGCTTTGTTTCC

Sanger sequencing runs and primers used to complete *hmw1A* gene sequence are listed below for each of the following NTHi isolates:

P651-8849: PCR product (1656+1657) with primer 1656; PCR product (1656+1666) with primer 1665; PCR product (1656+1666) with primer 1666; PCR product (1656+1666) with primer 1668; PCR product (1656+1666) with primer 1680; PCR product (1687+1696) with primer 1692; PCR product (1687+1696) with primer 1693; PCR product (1687+1696) with primer 1694; PCR product (1687+1696) with primer 1695.

P653-8956: PCR product (1656+1657) with primer 1656; PCR product (1656+1666) with primer 1665; PCR product (1656+1666) with primer 1666; PCR product (1656+1666) with primer 1680; PCR product (1687+1696) with primer 1687; PCR product (1687+1696) with primer 1692; PCR product (1687+1696) with primer 1693; PCR product (1687+1696) with primer 1694; PCR product (1693+1668) with primer 1668.

P654-8983: PCR product (1656+1657) with primer 1656; PCR product (1656+1666) with primer 1665; PCR product (1656+1666) with primer 1666; PCR product (1656+1666) with primer 1680; PCR product (1656+1666) with primer 1686; PCR product (1687+1696) with primer 1687; PCR product (1687+1696) with primer 1692; PCR product (1687+1696) with primer 1693; PCR product (1687+1696) with primer 1694; PCR product (1687+1696) with primer 1695; PCR product (1687+1695) with primer 1692. PCR product (1656+1695) with primer 1681.

3.7. References

Ahearn, C. P., Gallo, M. C. & Murphy, T. F. (2017). Insights on persistent airway infection by non-typeable *Haemophilus influenzae* in chronic obstructive pulmonary disease. *Pathog Dis*, 75 (4). doi: 10.1093/femspd/ftx042.

Andersson, S. G. & Kurland, C. G. (1998). Reductive evolution of resident genomes. *Trends Microbiol*, 6 (7): 263-8.

Ankrum, A. & Hall, B. G. (2017). Population dynamics of *Staphylococcus aureus* in cystic fibrosis patients to determine transmission events by use of whole-genome sequencing. *J Clin Microbiol*, 55 (7): 2143-2152. doi: 10.1128/JCM.00164-17.

Anzueto, A. (2010). Impact of exacerbations on COPD. *Eur Respir Rev*, 19 (116): 113-8. doi: 10.1183/09059180.00002610.

Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 22 (2): 195-201. doi: 10.1093/bioinformatics/bti770.

Bandi, V., Apicella, M. A., Mason, E., Murphy, T. F., Siddiqi, A., Atmar, R. L. & Greenberg, S. B. (2001). Nontypeable *Haemophilus influenzae* in the lower respiratory tract of patients with chronic bronchitis. *Am J Respir Crit Care Med*, 164 (11): 2114-9. doi: 10.1164/ajrccm.164.11.2104093.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*, 19 (5): 455-77. doi: 10.1089/cmb.2012.0021.

Bari, M. R., Hiron, M. M., Zaman, S. M., Rahman, M. M. & Ganguly, K. C. (2010). Microbes responsible for acute exacerbation of COPD. *Mymensingh Med J*, 19 (4): 576-85.

Beasley, V., Joshi, P. V., Singanayagam, A., Molyneaux, P. L., Johnston, S. L. & Mallia, P. (2012). Lung microbiology and exacerbations in COPD. *Int J Chron Obstruct Pulmon Dis*, 7: 555-69. doi: 10.2147/COPD.S28286.

Berrens, Z. J., Marrs, C. F., Pettigrew, M. M., Sandstedt, S. A., Patel, M. & Gilsdorf, J. R. (2007). Genetic diversity of paired middle-ear and pharyngeal nontypeable *Haemophilus influenzae* isolates from children with acute otitis media. *J Clin Microbiol*, 45 (11): 3764-7. doi: 10.1128/JCM.00964-07.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., et al. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*, 42 (Web Server issue): W252-8. doi: 10.1093/nar/gku340.

Bliven, K. A. & Maurelli, A. T. (2016). Evolution of bacterial pathogens within the human host. *Microbiol Spectr*, 4 (1). doi: 10.1128/microbiolspec.VMBF-0017-2015.

Bolger, A. M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30 (15): 2114-20. doi: 10.1093/bioinformatics/btu170.

Brogden, K. A. (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat Rev Microbiol*, 3 (3): 238-50. doi: 10.1038/nrmicro1098.

Buscher, A. Z., Burmeister, K., Barenkamp, S. J. & St Geme, J. W., 3rd. (2004). Evolutionary and functional relationships among the nontypeable *Haemophilus influenzae* HMW family of adhesins. *J Bacteriol*, 186 (13): 4209-17. doi: 10.1128/JB.186.13.4209-4217.2004.

Buscher, A. Z., Grass, S., Heuser, J., Roth, R. & St Geme, J. W., 3rd. (2006). Surface anchoring of a bacterial adhesin secreted by the two-partner secretion pathway. *Mol Microbiol*, 61 (2): 470-83. doi: 10.1111/j.1365-2958.2006.05236.x.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10: 421. doi: 10.1186/1471-2105-10-421.

Case, D. A. V. B., J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, XWu and P.A. Kollman. (2014). AMBER 14. *University of California, San Francisco*.

Chattopadhyay, M. K. (2014). Use of antibiotics as feed additives: a burning question. *Frontiers in Microbiology*, 5: 334. doi: 10.3389/fmicb.2014.00334.

Chin, C. L., Manzel, L. J., Lehman, E. E., Humlicek, A. L., Shi, L., Starner, T. D., Denning, G. M., Murphy, T. F., Sethi, S. & Look, D. C. (2005). *Haemophilus influenzae* from patients with chronic obstructive pulmonary disease exacerbation induce more inflammation than colonizers. *Am J Respir Crit Care Med*, 172 (1): 85-91. doi: 10.1164/rccm.200412-1687OC.

Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*, 10 (6): 563-9. doi: 10.1038/nmeth.2474.

Cholon, D. M., Cutter, D., Richardson, S. K., Sethi, S., Murphy, T. F., Look, D. C. & St Geme, J. W., 3rd. (2008). Serial isolates of persistent *Haemophilus influenzae* in patients with chronic obstructive pulmonary disease express diminishing quantities of the HMW1 and HMW2 adhesins. *Infect Immun*, 76 (10): 4463-8. doi: 10.1128/IAI.00499-08.

Chung, H., Lieberman, T. D., Vargas, S. O., Flett, K. B., McAdam, A. J., Priebe, G. P. & Kishony, R. (2017). Global and local selection acting on the pathogen *Stenotrophomonas maltophilia* in the human lung. *Nat Commun*, 8: 14078. doi: 10.1038/ncomms14078.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X. & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*, 6 (2): 80-92. doi: 10.4161/fly.19695.

Clementi, C. F. & Murphy, T. F. (2011). Non-typeable *Haemophilus influenzae* invasion and persistence in the human respiratory tract. *Front Cell Infect Microbiol*, 1: 1. doi: 10.3389/fcimb.2011.00001.

Clementi, C. F., Hakansson, A. P. & Murphy, T. F. (2014). Internalization and trafficking of nontypeable *Haemophilus influenzae* in human respiratory epithelial cells and roles of IgA1 proteases for optimal invasion and persistence. *Infect Immun*, 82 (1): 433-44. doi: 10.1128/IAI.00864-13.

Croucher, N. J., Harris, S. R., Barquist, L., Parkhill, J. & Bentley, S. D. (2012). A high-resolution view of genome-wide pneumococcal transformation. *PLoS Pathog*, 8 (6): e1002745. doi: 10.1371/journal.ppat.1002745.

Cullen, L. & McClean, S. (2015). Bacterial adaptation during chronic respiratory infections. *Pathogens*, 4 (1): 66-89. doi: 10.3390/pathogens4010066.

Darling, A. E., Mau, B. & Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*, 5 (6): e11147. doi: 10.1371/journal.pone.0011147.

Dawid, S., Barenkamp, S. J. & St. Geme, J. W. (1999). Variation in expression of the *Haemophilus influenzae* HMW adhesins: A prokaryotic system reminiscent of eukaryotes. *Proceedings of the National Academy of Sciences*, 96 (3): 1077-1082. doi: 10.1073/pnas.96.3.1077.

De Chiara, M., Hood, D., Muzzi, A., Pickard, D. J., Perkins, T., Pizza, M., Dougan, G., Rappuoli, R., Moxon, E. R., Soriani, M., et al. (2014). Genome sequencing of disease and carriage isolates of nontypeable *Haemophilus influenzae* identifies discrete population structure. *Proc Natl Acad Sci U S A*, 111 (14): 5439-44. doi: 10.1073/pnas.1403353111.

de Groot, R., Campos, J., Moseley, S. L. & Smith, A. L. (1988). Molecular cloning and mechanism of trimethoprim resistance in *Haemophilus influenzae*. *Antimicrobial Agents and Chemotherapy*, 32 (4): 477-484.

de Groot, R., Sluijter, M., de Bruyn, A., Campos, J., Goessens, W. H., Smith, A. L. & Hermans, P. W. (1996). Genetic characterization of trimethoprim resistance in *Haemophilus influenzae*. *Antimicrobial Agents and Chemotherapy*, 40 (9): 2131-2136.

Deadman, M. E., Hermant, P., Engskog, M., Makepeace, K., Moxon, E. R., Schweda, E. K. & Hood, D. W. (2009). Lex2B, a phase-variable glycosyltransferase, adds either a glucose or a galactose to *Haemophilus influenzae* lipopolysaccharide. *Infect Immun*, 77 (6): 2376-84. doi: 10.1128/IAI.01446-08.

Desai, H., Eschberger, K., Wrona, C., Grove, L., Agrawal, A., Grant, B., Yin, J., Parameswaran, G. I., Murphy, T. & Sethi, S. (2014). Bacterial colonization increases daily symptoms in patients with chronic obstructive pulmonary disease. *Ann Am Thorac Soc*, 11 (3): 303-9. doi: 10.1513/AnnalsATS.201310-350OC.

Didelot, X., Walker, A. S., Peto, T. E., Crook, D. W. & Wilson, D. J. (2016). Within-host evolution of bacterial pathogens. *Nat Rev Microbiol*, 14 (3): 150-62. doi: 10.1038/nrmicro.2015.13.

Droge, J., Gregor, I. & McHardy, A. C. (2015). Taxator-tk: precise taxonomic assignment of metagenomes by fast approximation of evolutionary neighborhoods. *Bioinformatics*, 31 (6): 817-24. doi: 10.1093/bioinformatics/btu745.

Enne, V. I., King, A., Livermore, D. M. & Hall, L. M. (2002). Sulfonamide resistance in *Haemophilus influenzae* mediated by acquisition of sul2 or a short insertion in chromosomal folP. *Antimicrob Agents Chemother*, 46 (6): 1934-9.

Erkan, L., Uzun, O., Findik, S., Katar, D., Sanic, A. & Atici, A. G. (2008). Role of bacteria in acute exacerbations of chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis*, 3 (3): 463-7.

Erwin, A. L., Bonthuis, P. J., Geelhood, J. L., Nelson, K. L., McCrea, K. W., Gilsdorf, J. R. & Smith, A. L. (2006). Heterogeneity in tandem octanucleotides within *Haemophilus influenzae* lipopolysaccharide biosynthetic gene losA affects serum resistance. *Infect Immun*, 74 (6): 3408-14. doi: 10.1128/IAI.01540-05.

Euba, B., Moleres, J., Segura, V., Viadas, C., Morey, P., Moranta, D., Leiva, J., de-Torres, J. P., Bengoechea, J. A. & Garmendia, J. (2015a). Genome expression profiling-based identification and administration efficacy of host-directed antimicrobial drugs against respiratory infection by nontypeable *Haemophilus influenzae*. *Antimicrob Agents Chemother*, 59 (12): 7581-92. doi: 10.1128/AAC.01278-15.

Euba, B., Moleres, J., Viadas, C., Barberan, M., Caballero, L., Grillo, M. J., Bengoechea, J. A., de-Torres, J. P., Linares, J., Leiva, J., et al. (2015b). Relationship between azithromycin susceptibility and administration efficacy for nontypeable *Haemophilus influenzae* respiratory infection. *Antimicrob Agents Chemother*, 59 (5): 2700-12. doi: 10.1128/AAC.04447-14.

Euba, B., Moleres, J., Viadas, C., Ruiz de los Mozos, I., Valle, J., Bengoechea, J. A. & Garmendia, J. (2015c). Relative contribution of P5 and Hap surface proteins to nontypable

Haemophilus influenzae interplay with the host Upper and lower airways. *PLoS One*, 10 (4): e0123154. doi: 10.1371/journal.pone.0123154.

Euba, B., Lopez-Lopez, N., Rodriguez-Arce, I., Fernandez-Calvet, A., Barberan, M., Caturla, N., Marti, S., Diez-Martinez, R. & Garmendia, J. (2017). Resveratrol therapeutics combines both antimicrobial and immunomodulatory properties against respiratory infection by nontypeable *Haemophilus influenzae*. *Sci Rep*, 7 (1): 12860. doi: 10.1038/s41598-017-13034-7.

Evans, M. E., Feola, D. J. & Rapp, R. P. (1999). Polymyxin B sulfate and colistin: old antibiotics for emerging multiresistant gram-negative bacteria. *Ann Pharmacother*, 33 (9): 960-7. doi: 10.1345/aph.18426.

Fagerberg, L., Hallstrom, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., et al. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics, 13 (2): 397-406. doi: 10.1074/mcp.M113.035600.

Farjo, R. S., Foxman, B., Patel, M. J., Zhang, L., Pettigrew, M. M., McCoy, S. I., Marrs, C. F. & Gilsdorf, J. R. (2004). Diversity and sharing of *Haemophilus influenzae* strains colonizing healthy children attending day-care centers. *Pediatr Infect Dis J*, 23 (1): 41-6. doi: 10.1097/01.inf.0000106981.89572.d1.

Fernaays, M. M., Lesse, A. J., Cai, X. & Murphy, T. F. (2006). Characterization of *igaB*, a second immunoglobulin A1 protease gene in nontypeable *Haemophilus influenzae*. *Infect Immun*, 74 (10): 5860-70. doi: 10.1128/IAI.00796-06.

Finney, L. J., Ritchie, A., Pollard, E., Johnston, S. L. & Mallia, P. (2014). Lower airway colonization and inflammatory response in COPD: a focus on *Haemophilus influenzae*. *Int J Chron Obstruct Pulmon Dis*, 9: 1119-32. doi: 10.2147/COPD.S54477.

Fleischmann, R., Adams, M., White, O., Clayton, R., Kirkness, E., Kerlavage, A., Bult, C., Tomb, J., Dougherty, B., Merrick, J., et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 269 (5223): 496-512. doi: 10.1126/science.7542800.

Folkesson, A., Jelsbak, L., Yang, L., Johansen, H. K., Ciofu, O., Hoiby, N. & Molin, S. (2012). Adaptation of *Pseudomonas aeruginosa* to the cystic fibrosis airway: an evolutionary perspective. *Nat Rev Microbiol*, 10 (12): 841-51. doi: 10.1038/nrmicro2907.

Fox, K. L., Srikhanta, Y. N. & Jennings, M. P. (2007). Phase variable type III restriction-modification systems of host-adapted bacterial pathogens. *Mol Microbiol*, 65 (6): 1375-9. doi: 10.1111/j.1365-2958.2007.05873.x.

Francisco, A. P., Bugalho, M., Ramirez, M. & Carrico, J. A. (2009). Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics*, 10: 152. doi: 10.1186/1471-2105-10-152.

Gao, R., Hu, Y., Li, Z., Sun, J., Wang, Q., Lin, J., Ye, H., Liu, F., Srinivas, S., Li, D., et al. (2016). Dissemination and Mechanism for the MCR-1 Colistin Resistance. *PLoS Pathog*, 12 (11): e1005957. doi: 10.1371/journal.ppat.1005957.

Garmendia, J., Viadas, C., Calatayud, L., Mell, J. C., Marti-Lliteras, P., Euba, B., Llobet, E., Gil, C., Bengoechea, J. A., Redfield, R. J., et al. (2014). Characterization of nontypable *Haemophilus influenzae* isolates recovered from adult patients with underlying chronic lung disease reveals genotypic and phenotypic traits associated with persistent infection. *PLoS One*, 9 (5): e97020. doi: 10.1371/journal.pone.0097020.

Gilsdorf, J. R., Marrs, C. F. & Foxman, B. (2004). *Haemophilus influenzae*: genetic variability and natural selection to identify virulence factors. *Infect Immun*, 72 (5): 2457-61.

Gray-Owen, S. D., Dehio, C., Haude, A., Grunert, F. & Meyer, T. F. (1997). CD66 carcinoembryonic antigens mediate interactions between *Opa*-expressing Neisseria

gonorrhoeae and human polymorphonuclear phagocytes. *EMBO J*, 16 (12): 3435-45. doi: 10.1093/emboj/16.12.3435.

Guex, N., Peitsch, M. C. & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis*, 30 Suppl 1: S162-73. doi: 10.1002/elps.200900140.

Guttman, D. S., Gropp, S. J., Morgan, R. L. & Wang, P. W. (2006). Diversifying selection drives the evolution of the type III secretion system pilus of *Pseudomonas syringae*. *Mol Biol Evol*, 23 (12): 2342-54. doi: 10.1093/molbev/msl103.

Hancock, R. E. (1999). Host defence (cationic) peptides: what is their future clinical potential? *Drugs*, 57 (4): 469-73.

Hearn, E. M., Patel, D. R., Lepore, B. W., Indic, M. & van den Berg, B. (2009). Transmembrane passage of hydrophobic compounds through a protein channel wall. *Nature*, 458 (7236): 367-70. doi: 10.1038/nature07678.

Herath, S. C. & Poole, P. (2013). Prophylactic antibiotic therapy for chronic obstructive pulmonary disease (COPD). *Cochrane Database Syst Rev* (11): CD009764. doi: 10.1002/14651858.CD009764.pub2.

Hergott, C. B., Roche, A. M., Naidu, N. A., Mesaros, C., Blair, I. A. & Weiser, J. N. (2015). Bacterial exploitation of phosphorylcholine mimicry suppresses inflammation to promote airway infection. *J Clin Invest*, 125 (10): 3878-90. doi: 10.1172/JCI81888.

Hermsen, E. D., Sullivan, C. J. & Rotschafer, J. C. (2003). Polymyxins: pharmacology, pharmacokinetics, pharmacodynamics, and clinical applications. *Infect Dis Clin North Am*, 17 (3): 545-62.

Herriott, R. M., Meyer, E. M. & Vogt, M. (1970). Defined nongrowth media for stage II development of competence in *Haemophilus influenzae*. *J Bacteriol*, 101 (2): 517-24.

Hiltke, T. J., Schiffmacher, A. T., Dagonese, A. J., Sethi, S. & Murphy, T. F. (2003). Horizontal transfer of the gene encoding outer membrane protein P2 of nontypeable *Haemophilus influenzae*, in a patient with chronic obstructive pulmonary disease. *J Infect Dis*, 188 (1): 114-7. doi: 10.1086/375724.

Hogg, J. S., Hu, F. Z., Janto, B., Boissy, R., Hayes, J., Keefe, R., Post, J. C. & Ehrlich, G. D. (2007). Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome Biol*, 8 (6): R103. doi: 10.1186/gb-2007-8-6-r103.

Howell, K. J., Weinert, L. A., Chaudhuri, R. R., Luan, S. L., Peters, S. E., Corander, J., Harris, D., Angen, O., Aragon, V., Bensaid, A., et al. (2014). The use of genome wide association methods to investigate pathogenicity, population structure and serovar in *Haemophilus parasuis*. *BMC Genomics*, 15: 1179. doi: 10.1186/1471-2164-15-1179.

Investigators, S. (2014). Darapladib for Preventing Ischemic Events in Stable Coronary Heart Disease. *New England Journal of Medicine*, 370 (18): 1702-1711. doi: 10.1056/NEJMoa1315878.

Jorth, P., Staudinger, B. J., Wu, X., Hisert, K. B., Hayden, H., Garudathri, J., Harding, C. L., Radey, M. C., Rezayat, A., Bautista, G., et al. (2015). Regional Isolation Drives Bacterial Diversification within Cystic Fibrosis Lungs. *Cell Host Microbe*, 18 (3): 307-19. doi: 10.1016/j.chom.2015.07.006.

Kallberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H. & Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nat Protoc*, 7 (8): 1511-22. doi: 10.1038/nprot.2012.085.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*, 10 (6): 845-58. doi: 10.1038/nprot.2015.053.

Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L. & Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res*, 37 (Database issue): D387-92. doi: 10.1093/nar/gkn750.

King, P. T. & Sharma, R. (2015). The lung immune response to nontypeable *Haemophilus influenzae* (Lung Immunity to NTHi). *J Immunol Res*, 2015: 706376. doi: 10.1155/2015/706376.

Knapp, H. R. & Melly, M. A. (1986). Bactericidal effects of polyunsaturated fatty acids. *J Infect Dis*, 154 (1): 84-94.

Kostyanev, T. S. & Sechanova, L. P. (2012). Virulence factors and mechanisms of antibiotic resistance of *Haemophilus influenzae*. *Folia Med (Plovdiv)*, 54 (1): 19-23.

Krasan, G. P., Cutter, D., Block, S. L. & St Geme, J. W., 3rd. (1999). Adhesin expression in matched nasopharyngeal and middle ear isolates of nontypeable *Haemophilus influenzae* from children with acute otitis media. *Infect Immun*, 67 (1): 449-54.

Krol, E. & Becker, A. (2014). Rhizobial homologs of the fatty acid transporter FadL facilitate perception of long-chain acyl-homoserine lactone signals. *Proc Natl Acad Sci U S A*, 111 (29): 10702-7. doi: 10.1073/pnas.1404929111.

Kruse, H. & Sorum, H. (1994). Transfer of multiple drug resistance plasmids between bacteria of diverse origins in natural microenvironments. *Appl Environ Microbiol*, 60 (11): 4015-21.

Laabei, M., Uhlemann, A. C., Lowy, F. D., Austin, E. D., Yokoyama, M., Ouadi, K., Feil, E., Thorpe, H. A., Williams, B., Perkins, M., et al. (2015). Evolutionary Trade-Offs Underlie the Multi-faceted Virulence of *Staphylococcus aureus*. *PLoS Biol*, 13 (9): e1002229. doi: 10.1371/journal.pbio.1002229.

Lacross, N. C., Marrs, C. F., Patel, M., Sandstedt, S. A. & Gilsdorf, J. R. (2008). High genetic diversity of nontypeable *Haemophilus influenzae* isolates from two children attending a day care center. *J Clin Microbiol*, 46 (11): 3817-21. doi: 10.1128/JCM.00940-08.

LaCross, N. C., Marrs, C. F. & Gilsdorf, J. R. (2013). Population structure in nontypeable *Haemophilus influenzae*. *Infect Genet Evol*, 14: 125-36. doi: 10.1016/j.meegid.2012.11.023.

Langereis, J. D. & de Jonge, M. I. (2015). Invasive disease caused by nontypeable *Haemophilus influenzae*. *Emerg Infect Dis*, 21 (10): 1711-8. doi: 10.3201/eid2110.150004.

Lee, A. H., Flibotte, S., Sinha, S., Paiero, A., Ehrlich, R. L., Balashov, S., Ehrlich, G. D., Zlosnik, J. E., Mell, J. C. & Nislow, C. (2017). Phenotypic diversity and genotypic flexibility of *Burkholderia cenocepacia* during long-term chronic infection of cystic fibrosis lungs. *Genome Res*, 27 (4): 650-662. doi: 10.1101/gr.213363.116.

Lepore, B. W., Indic, M., Pham, H., Hearn, E. M., Patel, D. R. & van den Berg, B. (2011). Ligand-gated diffusion across the bacterial outer membrane. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (25): 10121-10126. doi: 10.1073/pnas.1018532108.

Lerner, C. G., Kakavas, S. J., Wagner, C., Chang, R. T., Merta, P. J., Ruan, X., Metzger, R. E. & Beutel, B. A. (2005). Novel approach to mapping of resistance mutations in whole genomes by using restriction enzyme modulation of transformation efficiency. *Antimicrob Agents Chemother*, 49 (7): 2767-77. doi: 10.1128/aac.49.7.2767-2777.2005.

Lieberman, T. D., Michel, J. B., Aingaran, M., Potter-Bynoe, G., Roux, D., Davis, M. R., Jr., Skurnik, D., Leiby, N., LiPuma, J. J., Goldberg, J. B., et al. (2011). Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet*, 43 (12): 1275-80. doi: 10.1038/ng.997.

Lieberman, T. D., Flett, K. B., Yelin, I., Martin, T. R., McAdam, A. J., Priebe, G. P. & Kishony, R. (2014). Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nat Genet*, 46 (1): 82-7. doi: 10.1038/ng.2848.

Lieberman, T. D., Wilson, D., Misra, R., Xiong, L. L., Moodley, P., Cohen, T. & Kishony, R. (2016). Genomic diversity in autopsy samples reveals within-host dissemination of HIV-associated *Mycobacterium tuberculosis*. *Nat Med*, 22 (12): 1470-1474. doi: 10.1038/nm.4205.

Liu, B., Yuan, J., Yiu, S. M., Li, Z., Xie, Y., Chen, Y., Shi, Y., Zhang, H., Li, Y., Lam, T. W., et al. (2012). COPE: an accurate k-mer-based pair-end reads connection tool to facilitate genome assembly. *Bioinformatics*, 28 (22): 2870-4. doi: 10.1093/bioinformatics/bts563.

Look, D. C., Chin, C. L., Manzel, L. J., Lehman, E. E., Humlicek, A. L., Shi, L., Starner, T. D., Denning, G. M., Murphy, T. F. & Sethi, S. (2006). Modulation of airway inflammation by *Haemophilus influenzae* isolates associated with chronic obstructive pulmonary disease exacerbation. *Proc Am Thorac Soc*, 3 (6): 482-3. doi: 10.1513/pats.200603-060MS.

Lopez-Collazo, E., Jurado, T., de Dios Caballero, J., Perez-Vazquez, M., Vindel, A., Hernandez-Jimenez, E., Tamames, J., Cubillos-Zapata, C., Manrique, M., Tobes, R., et al. (2015). In vivo attenuation and genetic evolution of a ST247-SCCmecI MRSA clone after 13 years of pathogenic bronchopulmonary colonization in a patient with cystic fibrosis: implications of the innate immune response. *Mucosal Immunol*, 8 (2): 362-71. doi: 10.1038/mi.2014.73.

Lopez-Gomez, A., Cano, V., Moranta, D., Morey, P., Garcia del Portillo, F., Bengoechea, J. A. & Garmendia, J. (2012). Host cell kinases, α5 and β1 integrins, and Rac1 signalling on the microtubule cytoskeleton are important for non-typable *Haemophilus influenzae* invasion of respiratory epithelial cells. *Microbiology*, 158 (Pt 9): 2384-98. doi: 10.1099/mic.0.059972-0.

Malhotra, S., Deshmukh, S. S. & Dastidar, S. G. (2012). COX inhibitors for airway inflammation. *Expert Opin Ther Targets*, 16 (2): 195-207. doi: 10.1517/14728222.2012.661416.

Markussen, T., Marvig, R. L., Gomez-Lozano, M., Aanaes, K., Burleigh, A. E., Hoiby, N., Johansen, H. K., Molin, S. & Jelsbak, L. (2014). Environmental heterogeneity drives within-host diversification and evolution of *Pseudomonas aeruginosa*. *MBio*, 5 (5): e01592-14. doi: 10.1128/mBio.01592-14.

Marvig, R. L., Damkiaer, S., Khademi, S. M., Markussen, T. M., Molin, S. & Jelsbak, L. (2014). Within-host evolution of *Pseudomonas aeruginosa* reveals adaptation toward iron acquisition from hemoglobin. *MBio*, 5 (3): e00966-14. doi: 10.1128/mBio.00966-14.

Marvig, R. L., Sommer, L. M., Molin, S. & Johansen, H. K. (2015). Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. *Nat Genet*, 47 (1): 57-64. doi: 10.1038/ng.3148.

Mason, K. M., Munson, R. S., Jr. & Bakaletz, L. O. (2005). A mutation in the sap operon attenuates survival of nontypeable *Haemophilus influenzae* in a chinchilla model of otitis media. *Infect Immun*, 73 (1): 599-608. doi: 10.1128/IAI.73.1.599-608.2005.

Maughan, H. & Redfield, R. J. (2009). Extensive variation in natural competence in *Haemophilus influenzae*. *Evolution*, 63 (7): 1852-66. doi: 10.1111/j.1558-5646.2009.00658.x.

Maughan, H. S., S.; Wilson, L. and Redfield, R. (2018). Competence, DNA uptake and transformation in *Pasteurellaceae*. Pasteurellaceae: Biology, Genomics and Molecular Aspects: Caister Academic Press.

McAdam, P. R., Holmes, A., Templeton, K. E. & Fitzgerald, J. R. (2011). Adaptive evolution of *Staphylococcus aureus* during chronic endobronchial infection of a cystic fibrosis patient. *PLoS One*, 6 (9): e24301. doi: 10.1371/journal.pone.0024301.

McIntyre, T. M., Prescott, S. M. & Stafforini, D. M. (2009). The emerging roles of PAF acetylhydrolase. *J Lipid Res*, 50 Suppl: S255-9. doi: 10.1194/jlr.R800024-JLR200.

Mell, J. C., Shumilina, S., Hall, I. M. & Redfield, R. J. (2011). Transformation of natural genetic variation into *Haemophilus influenzae* genomes. *PLoS Pathog*, 7 (7): e1002151. doi: 10.1371/journal.ppat.1002151.

Mell, J. C. & Redfield, R. J. (2014). Natural competence and the evolution of DNA uptake specificity. *J Bacteriol*, 196 (8): 1471-83. doi: 10.1128/JB.01293-13.

Mell, J. C., Sinha, S., Balashov, S., Viadas, C., Grassa, C. J., Ehrlich, G. D., Nislow, C., Redfield, R. J. & Garmendia, J. (2014). Complete Genome Sequence of *Haemophilus influenzae* Strain 375 from the Middle Ear of a Pediatric Patient with Otitis Media. *Genome Announc*, 2 (6). doi: 10.1128/genomeA.01245-14.

Mell, J. C., Viadas, C., Moleres, J., Sinha, S., Fernandez-Calvet, A., Porsch, E. A., St Geme, J. W., 3rd, Nislow, C., Redfield, R. J. & Garmendia, J. (2016). Transformed recombinant enrichment profiling rapidly identifies HMW1 as an intracellular invasion locus in *Haemophilus influenzae*. *PLoS Pathog*, 12 (4): e1005576. doi: 10.1371/journal.ppat.1005576.

Mes, T. H. & van Putten, J. P. (2007). Positively selected codons in immune-exposed loops of the vaccine candidate OMP-P1 of *Haemophilus influenzae*. *J Mol Evol*, 64 (4): 411-22. doi: 10.1007/s00239-006-0021-2.

Meyer, M. C., Rastogi, P., Beckett, C. S. & McHowat, J. (2005). Phospholipase A2 inhibitors as potential anti-inflammatory agents. *Curr Pharm Des*, 11 (10): 1301-12.

Morey, P., Cano, V., Marti-Lliteras, P., Lopez-Gomez, A., Regueiro, V., Saus, C., Bengoechea, J. A. & Garmendia, J. (2011). Evidence for a non-replicative intracellular stage of nontypable *Haemophilus influenzae* in epithelial cells. *Microbiology*, 157 (Pt 1): 234-50. doi: 10.1099/mic.0.040451-0.

Morey, P., Viadas, C., Euba, B., Hood, D. W., Barberan, M., Gil, C., Grillo, M. J., Bengoechea, J. A. & Garmendia, J. (2013). Relative contributions of lipooligosaccharide inner and outer core modifications to nontypeable *Haemophilus influenzae* pathogenesis. *Infect Immun*, 81 (11): 4100-11. doi: 10.1128/IAI.00492-13.

Morris, G. M. G., D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R.K.; Olson, A. J. (1998). AutoDock treats the ligand as a flexible unit and the protein as a rigid unit. *J. Comput. Chem.*, 19 (1639).

Morton, D. J., Hempel, R. J., Whitby, P. W., Seale, T. W. & Stull, T. L. (2012). An invasive *Haemophilus haemolyticus* isolate. *J Clin Microbiol*, 50 (4): 1502-3. doi: 10.1128/JCM.06688-11.

Murphy, T. F., Sethi, S., Klingman, K. L., Brueggemann, A. B. & Doern, G. V. (1999). Simultaneous respiratory tract colonization by multiple strains of nontypeable *Haemophilus influenzae* in chronic obstructive pulmonary disease: implications for antibiotic therapy. *J Infect Dis*, 180 (2): 404-9. doi: 10.1086/314870.

Murphy, T. F., Brauer, A. L., Schiffmacher, A. T. & Sethi, S. (2004). Persistent colonization by *Haemophilus influenzae* in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*, 170 (3): 266-72. doi: 10.1164/rccm.200403-354OC.

Murray, P. R., Baron, E. J. & American Society for, M. (2003). *Manual of clinical microbiology*. Washington, D.C.: ASM Press.

Nakamura, S., Shchepetov, M., Dalia, A. B., Clark, S. E., Murphy, T. F., Sethi, S., Gilsdorf, J. R., Smith, A. L. & Weiser, J. N. (2011). Molecular basis of increased serum resistance among pulmonary isolates of non-typeable *Haemophilus influenzae*. *PLoS Pathog*, 7 (1): e1001247. doi: 10.1371/journal.ppat.1001247.

Nedbalcova, K. S., P; Jaglic, Z.; Ondriasova, R. and Kucerova, Z. (2006). *Haemophilus parasuis* and Glässer's disease in pigs: A review. *Veterinarni Medicina*, 51 (5): 168-179.

Nicolas, P. & Mor, A. (1995). Peptides as weapons against microorganisms in the chemical defense system of vertebrates. *Annu Rev Microbiol*, 49: 277-304. doi: 10.1146/annurev.mi.49.100195.001425.

Notredame, C., Higgins, D. G. & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*, 302 (1): 205-17. doi: 10.1006/jmbi.2000.4042.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., Fookes, M., Falush, D., Keane, J. A. & Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, 31 (22): 3691-3. doi: 10.1093/bioinformatics/btv421.

Petrache, I. & Petrusca, D. N. (2013). The involvement of sphingolipids in chronic obstructive pulmonary diseases. I: Gulbins, E. & Petrache, I. (red.) *Sphingolipids in Disease*, s. 247-264. Vienna: Springer Vienna.

Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T. & Weng, Z. (2014). ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics*, 30 (12): 1771-3. doi: 10.1093/bioinformatics/btu097.

Platt, D. J., Guthrie, A. J. & Langan, C. E. (1983). The isolation of thymidine-requiring *Haemophilus influenzae* from the sputum of chronic bronchitic patients receiving trimethoprim. *J Antimicrob Chemother*, 11 (3): 281-6.

Power, P. M., Sweetman, W. A., Gallacher, N. J., Woodhall, M. R., Kumar, G. A., Moxon, E. R. & Hood, D. W. (2009). Simple sequence repeats in *Haemophilus influenzae*. *Infect Genet Evol*, 9 (2): 216-28. doi: 10.1016/j.meegid.2008.11.006.

Price, E. P., Sarovich, D. S., Mayo, M., Tuanyok, A., Drees, K. P., Kaestli, M., Beckstrom-Sternberg, S. M., Babic-Sternberg, J. S., Kidd, T. J., Bell, S. C., et al. (2013). Within-host evolution of *Burkholderia pseudomallei* over a twelve-year chronic carriage infection. *MBio*, 4 (4). doi: 10.1128/mBio.00388-13.

Price, E. P., Sarovich, D. S., Nosworthy, E., Beissbarth, J., Marsh, R. L., Pickering, J., Kirkham, L. A., Keil, A. D., Chang, A. B. & Smith-Vaughan, H. C. (2015). *Haemophilus influenzae*: using comparative genomics to accurately identify a highly recombinogenic human pathogen. *BMC Genomics*, 16: 641. doi: 10.1186/s12864-015-1857-x.

Puig, C., Calatayud, L., Marti, S., Tubau, F., Garcia-Vidal, C., Carratala, J., Linares, J. & Ardanuy, C. (2013). Molecular epidemiology of nontypeable H*aemophilus influenzae* causing community-acquired pneumonia in adults. *PLoS One*, 8 (12): e82515. doi: 10.1371/journal.pone.0082515.

Quinlan, A. R. & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26 (6): 841-2. doi: 10.1093/bioinformatics/btq033.

Rodriguez, C. A., Avadhanula, V., Buscher, A., Smith, A. L., St Geme, J. W., 3rd & Adderson, E. E. (2003). Prevalence and distribution of adhesins in invasive non-type b encapsulated *Haemophilus influenzae*. *Infect Immun*, 71 (4): 1635-42.

Rodriguez-Arce, I., Marti, S., Euba, B., Fernandez-Calvet, A., Moleres, J., Lopez-Lopez, N., Barberan, M., Ramos-Vivas, J., Tubau, F., Losa, C., et al. (2017). Inactivation of the thymidylate synthase thyA in non-typeable *Haemophilus influenzae* modulates antibiotic resistance and has a strong impact on its interplay with the host airways. *Front Cell Infect Microbiol*, 7: 266. doi: 10.3389/fcimb.2017.00266.

Roy, A., Kucukural, A. & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*, 5 (4): 725-38. doi: 10.1038/nprot.2010.5.

Salcedo, S. P. & Cid, V. J. (2011). Nontypable Haemophilus influenzae: an intracellular phase within epithelial cells might contribute to persistence. *Microbiology*, 157 (Pt 1): 1-2. doi: 10.1099/mic.0.046722-0.

San Millan, A., Garcia-Cobos, S., Escudero, J. A., Hidalgo, L., Gutierrez, B., Carrilero, L., Campos, J. & Gonzalez-Zorn, B. (2010). *Haemophilus influenzae* clinical isolates with plasmid pB1000 bearing bla_{ROB-1}: fitness cost and interspecies dissemination. *Antimicrob Agents Chemother*, 54 (4): 1506-11. doi: 10.1128/AAC.01489-09.

Schrodinger. (2012a). Epik version 2.3,Schrodinger, LLC, New York, NY, 2012.

Schrodinger. (2012b). Impaversion 5.8, Schrodinger, LLC, New York, NY.

Schrodinger. (2012c). Prime version 3.1, Schrodinger, LLC, New York, NY.

Schrodinger. (2012d). Suite 2012 Protein preparation Wizard. *NY*.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30 (14): 2068-9. doi: 10.1093/bioinformatics/btu153.

Sethi, S., Evans, N., Grant, B. J. & Murphy, T. F. (2002). New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. *N Engl J Med*, 347 (7): 465-71. doi: 10.1056/NEJMoa012561.

Sethi, S. (2004). Bacteria in exacerbations of chronic obstructive pulmonary disease: phenomenon or epiphenomenon? *Proc Am Thorac Soc*, 1 (2): 109-14. doi: 10.1513/pats.2306029.

Sethi, S., Wrona, C., Grant, B. J. & Murphy, T. F. (2004). Strain-specific immune response to *Haemophilus influenzae* in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*, 169 (4): 448-53. doi: 10.1164/rccm.200308-1181OC.

Sethi, S. & Murphy, T. F. (2008). Infection in the pathogenesis and course of chronic obstructive pulmonary disease. *N Engl J Med*, 359 (22): 2355-65. doi: 10.1056/NEJMra0800353.

Siegel, S. J. & Weiser, J. N. (2015). Mechanisms of bacterial colonization of the respiratory tract. *Annual review of microbiology*, 69: 425-444. doi: 10.1146/annurev-micro-091014-104209.

Sinha, S., Mell, J. C. & Redfield, R. J. (2012). Seventeen Sxy-dependent cyclic AMP receptor protein site-regulated genes are needed for natural transformation in *Haemophilus influenzae*. *J Bacteriol*, 194 (19): 5245-54. doi: 10.1128/JB.00671-12.

Smith, E. E., Buckley, D. G., Wu, Z., Saenphimmachak, C., Hoffman, L. R., D'Argenio, D. A., Miller, S. I., Ramsey, B. W., Speert, D. P., Moskowitz, S. M., et al. (2006). Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc Natl Acad Sci U S A*, 103 (22): 8487-92. doi: 10.1073/pnas.0602138103.

Smith-Vaughan, H. C., Sriprakash, K. S., Leach, A. J., Mathews, J. D. & Kemp, D. J. (1998). Low genetic diversity of *Haemophilus influenzae* type b compared to nonencapsulated *H. influenzae* in a population in which *H. influenzae* is highly endemic. *Infection and Immunity*, 66 (7): 3403-3409.

St Geme, J. W., 3rd, Falkow, S. & Barenkamp, S. J. (1993). High-molecular-weight proteins of nontypeable *Haemophilus influenzae* mediate attachment to human epithelial cells. *Proc Natl Acad Sci U S A*, 90 (7): 2875-9.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30 (9): 1312-3. doi: 10.1093/bioinformatics/btu033.

Staples, M., Graham, R. M. A. & Jennison, A. V. (2017). Characterisation of invasive clinical *Haemophilus influenzae* isolates in Queensland, Australia using whole-genome sequencing. *Epidemiol Infect*, 145 (8): 1727-1736. doi: 10.1017/S0950268817000450.

Tchoupa, A. K., Lichtenegger, S., Reidl, J. & Hauck, C. R. (2015). Outer membrane protein P1 is the CEACAM-binding adhesin of *Haemophilus influenzae*. *Mol Microbiol*, 98 (3): 440-55. doi: 10.1111/mmi.13134.

Tracy, E., Ye, F., Baker, B. D. & Munson, R. S., Jr. (2008). Construction of non-polar mutants in *Haemophilus influenzae* using FLP recombinase technology. *BMC Mol Biol*, 9: 101. doi: 10.1186/1471-2199-9-101.

Treangen, T. J. & Salzberg, S. L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet*, 13 (1): 36-46. doi: 10.1038/nrg3117.

Vaara, M. (1992). Agents that increase the permeability of the outer membrane. *Microbiol Rev*, 56 (3): 395-411.

Villullas, S., Hill, D. J., Sessions, R. B., Rea, J. & Virji, M. (2007). Mutational analysis of human CEACAM1: the potential of receptor polymorphism in increasing host susceptibility to bacterial infection. *Cell Microbiol*, 9 (2): 329-46. doi: 10.1111/j.1462-5822.2006.00789.x.

Virji, M. (2000). The structural basis of CEACAM-receptor targeting by neisserial opa proteins: response. *Trends Microbiol*, 8 (6): 260-1.

Voges, M., Bachmann, V., Kammerer, R., Gophna, U. & Hauck, C. R. (2010). CEACAM1 recognition by bacterial pathogens is species-specific. *BMC Microbiol*, 10: 117. doi: 10.1186/1471-2180-10-117.

Whitby, P. W., Morton, D. J., Vanwagoner, T. M., Seale, T. W., Cole, B. K., Mussa, H. J., McGhee, P. A., Bauer, C. Y., Springer, J. M. & Stull, T. L. (2012). Haemophilus influenzae *OxyR*: characterization of its regulation, regulon and role in fitness. *PLoS One*, 7 (11): e50588. doi: 10.1371/journal.pone.0050588.

Winstanley, C., O'Brien, S. & Brockhurst, M. A. (2016). *Pseudomonas aeruginosa* evolutionary adaptation and diversification in cystic fibrosis chronic lung infections. *Trends Microbiol*, 24 (5): 327-37. doi: 10.1016/j.tim.2016.01.008.

Wu, Z., Periaswamy, B., Sahin, O., Yaeger, M., Plummer, P., Zhai, W., Shen, Z., Dai, L., Chen, S. L. & Zhang, Q. (2016). Point mutations in the major outer membrane protein drive hypervirulence of a rapidly expanding clone of *Campylobacter jejuni*. *Proc Natl Acad Sci U S A*, 113 (38): 10690-5. doi: 10.1073/pnas.1605869113.

Yang, J. & Zhang, Y. (2015). Protein structure and function prediction using I-TASSER. *Curr Protoc Bioinformatics*, 52: 5 8 1-15. doi: 10.1002/0471250953.bi0508s52.

Yang, L., Jelsbak, L., Marvig, R. L., Damkiaer, S., Workman, C. T., Rau, M. H., Hansen, S. K., Folkesson, A., Johansen, H. K., Ciofu, O., et al. (2011). Evolutionary dynamics of bacteria in a human host environment. *Proc Natl Acad Sci U S A*, 108 (18): 7481-6. doi: 10.1073/pnas.1018249108.

Yero, D., Vipond, C., Climent, Y., Sardinas, G., Feavers, I. M. & Pajon, R. (2010). Variation in the *Neisseria meningitidis* FadL-like protein: an evolutionary model for a relatively low-abundance surface antigen. *Microbiology*, 156 (Pt 12): 3596-608. doi: 10.1099/mic.0.043182-0.

Yin, W., Li, H., Shen, Y., Liu, Z., Wang, S., Shen, Z., Zhang, R., Walsh, T. R., Shen, J. & Wang, Y. (2017). Novel plasmid-mediated colistin resistance gene mcr-3 in *Escherichia coli*. *MBio*, 8 (3). doi: 10.1128/mBio.00543-17.

Young, B. C., Golubchik, T., Batty, E. M., Fung, R., Larner-Svensson, H., Votintseva, A. A., Miller, R. R., Godwin, H., Knox, K., Everitt, R. G., et al. (2012). Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. *Proc Natl Acad Sci U S A*, 109 (12): 4550-5. doi: 10.1073/pnas.1113219109.

Zhang, B., Tang, C., Liao, M. & Yue, H. (2014). Update on the pathogenesis of *Haemophilus parasuis* infection and virulence factors. *Vet Microbiol*, 168 (1): 1-7. doi: 10.1016/j.vetmic.2013.07.027.

Zhang, L., Xie, J., Patel, M., Bakhtyar, A., Ehrlich, G. D., Ahmed, A., Earl, J., Marrs, C. F., Clemans, D., Murphy, T. F., et al. (2012). Nontypeable *Haemophilus influenzae* genetic islands

associated with chronic pulmonary infection. *PLoS One*, 7 (9): e44730. doi: 10.1371/journal.pone.0044730.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9: 40. doi: 10.1186/1471-2105-9-40.

Fagerberg, L., Hallstrom, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., et al. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics, 13 (2): 397-406. doi: 10.1074/mcp.M113.035600.

GENERAL DISCUSSION

General Discussion

This Thesis work focused on unraveling adaptation features by two bacterial species belonging to the *Haemophilus* spp. genus, which are swine and human host-restricted colonizing opportunistic pathogens at the airways, in terms of the consequences of antibiotic administration, the use bacterial natural transformation for genetic screening purposes, and the occurrence of within-host lung evolutionary adaptation.

The **First Chapter** tackles the presence, distribution and features by an antimicrobial resistance determinant as a consequence of antibiotic prophylactic administration. Setting up a collection of *H. influenzae* strains isolated from healthy human volunteers was not possible in the context of this work. As an alternative, we employed *Haemophilus parasuis*, isolated from colonizing the nasopharynx of healthy piglets. *H. parasuis*, the causative agent of Glässer's disease, is an early colonizer of the nasal mucosa of piglets. Antibiotics are commonly used in disease control, and resistance to several antibiotics has been described in this bacterium. Both microorganisms (*H. influenzae* and *H. parasuis*) belong to the same bacterial genus despite their different host tropism and share some common features; i) both bacteria share the same type of niche within their respective hosts (Nedbalcova, 2006; Siegel & Weiser, 2015), ii) they are naturally competent and have great genomic diversity among isolates (De Chiara et al., 2014; Howell et al., 2014; Maughan, 2018) iii) they can reach the lower airways and cause disease, with the potential of causing invasive disease (Langereis & de Jonge, 2015; Nedbalcova, 2006), iv) they share virulence factors and strategies (Kostyanev & Sechanova, 2012; Zhang et al., 2014) , v) antibiotic administration in a prophylactic and/or therapeutic manner occurs in both hosts, thus generating similar selective pressures (Chattopadhyay, 2014; Herath & Poole, 2013). In this work, we isolated and characterized a representative number of *H. parasuis* strains from three distant regions in Spain. We acknowledge that sampling a higher number of farms, including more geographically distant ones, would have been desirable to get more robust observations and check for the presence of pJMA-1 and other resistance traits at a wider extent. On the other hand, and taking into account that animals play important roles as reservoirs of antimicrobial-resistant pathogens (Kruse & Sorum, 1994), and that the related plasmid pB1000 has been found in *H. influenzae* clinical isolates in Spain, Italy, Australia and North America, we unsuccessfully checked a collection of 77 colonizing and clinical isolates of *H. influenza*e for the presence of bla_{ROB-1}. Again, wider sampling is likely to increase the chances of identifying genetic cues associated to acquisition and/or spread of antibiotic resistance. Despite these limitations, results presented in Chapter 1 show

that commensal *H. parasuis* represents a reservoir of β-lactam resistance genes, with potential to be transferred to pathogens or other bacteria. The natural plasmid pJMA-1, bearing the bla_{ROB-1} β-lactamase was detected, shown to share a backbone with other small plasmids described in the *Pasteurellaceae*, to be stable, and to have a biological cost lower than previously described ones (San Millan et al., 2010).

The **Second Chapter** presents a genetic screening strategy that takes advantage of the inducible natural competent status of NTHi, its high inter-strain phenotypic variation, and deep sequencing technology. In this work, TREP has been shown to be a useful tool to unravel genetic traits of NTHi epithelial invasion. HMW plays an important role as a bacterial adhesin (St Geme et al., 1993), but the present study earmarks its role as an invasin based on the newly described self-aggregation + intracellular bacterial invasion phenotype. Gain of $hmw1_{86\text{-}028NP}$ by a poorly invading strain naturally lacking both *hmw1* and *hmw2* enhanced both adhesion and invasion, showing that $hmw1_{86\text{-}028NP}$ is responsible of conferring high adhesion and invasion levels. Moreover, one strain already possessing both *hmw1* and *hmw2* alleles and presenting an intermediate invasion rate, may be also able of allele exchange therefore becoming highly invasive. This fact points to the importance of not only HMW presence for epithelial invasion, but also to specific *hmw1* alleles as responsible for hyperinvasion. In any case, it would be interesting to assess the involvement of $hmw1_{86\text{-}028NP}$ in bacterial invasion in other cell types, since we cannot exclude the possibility of cell type-specific interactions.

An intriguing question arising from specific HMW alleles starring role in NTHi epithelial invasion is the frequency and distribution of this type of highly proficient alleles among NTHi isolates. De Chiara and co-workers showed that 49 out of 97 NTHi strains in study possessed an *hmw* allele, ~50% of a bacterial population encompassing isolates from different pathologic origins (De Chiara et al., 2014). Dawid and co-workers showed ~75% prevalence in an independent set of clinical isolates (Dawid et al., 1999); similar results have been obtained among invasive (75%) (Rodriguez et al., 2003) and otitis media (82%) strains (Krasan et al., 1999). We checked HMW distribution among strains belonging to the NTHi collection presented in Chapter 3 of this Thesis work, and observed significantly lower prevalence, in ~35% of strains. Of note, we observed a trend to decreased *hmw* expression linked to phase variation based on increasing numbers of heptanucleotide repeats in its promoter region. This suggests that hyperinvasion could be a phenotypic trait associated to early infection stages or acute infection processes such as those occurring at the ear. In contrast, lower airways long-term colonization may

involve progressive reduction of high invasion rates as a means of adaptive evolution. This observation is in high contrast with the notion of epithelial intracellular location associated to NTHi persistence (Clementi & Murphy, 2011; Clementi et al., 2014; Morey et al., 2011; Salcedo & Cid, 2011).

The **Third Chapter** focuses on the evolutionary dynamics of NTHi when reaching the lower airways of COPD patients. We collected NTHi isolates from sputum samples of 13 COPD patients longitudinally, and analyzed genomic architecture variation through clonal type (CT) assignment and genome comparison. An important feature emerging from the outlined natural evolution research is the importance of good longitudinally collected clinical samples, to allow comprehensive studies on how selective pressures shape bacterial genomes. Clinical sampling procedure, identification of clonal and non-clonal isolates within each sample, complete clinical data, number of sampled time points, and sampled time windows are important attributes that define any isolate collection as a starting point to unravel genomic traits associated with bacterial evolution. We faced several limitations in this study. First, the impossibility to fully distinguish between NTHi strains that were colonizing the nasopharynx and those reaching the lower airways and causing the long-term colonization/acute exacerbation, since bacterial isolation was performed from sputum samples; second, we kept one single strain from each sputum sample (except in two cases, where two strains were isolated), probably missing other isolation time points and bypassing the polyclonal nature of NTHi infection (Lacross et al., 2008; Murphy et al., 1999); third, this study focused on NTHi, whereas COPD exacerbations have been associated with isolation of other microorganisms from the lower airways, including *Streptoccocus pneumonie, Moraxella catharratidis, Burkholderia cepacia, Stenotrophomonas maltophilia, Klebsiella pneumonie, Pseudomonas aeruginosa, Staphylococcus aureus*, *Chlamydophila pneumoniae* and *Mycoplasma pneumoniae* (Bari et al., 2010; Beasley et al., 2012; Erkan et al., 2008; Sethi, 2004) both from sputum and bronchoscopic samples. This suggests a possible implication of other microorganisms, despite having positive cultures of NTHi. Moreover, co-infection with different NTHi strains and other bacterial species, as well as respiratory viruses, and the interactions and implications of such social behavior is missing in our study.

Of note, our whole genome sequencing allowed us tracking genetic changes associated to increased resistance to antibiotics. We focused on cotrimoxazole, as four CTs showed transitions from sensitive to resistant phenotype. We found some SNPs previously shown to confer or increase resistance to this antibiotic through *folP* and *folH* genes (de

Groot et al., 1996; Enne et al., 2002), and also identified new ones likely to be associated or contribute to increase such antibiotic resistance. We also analyzed bacterial resistance to colistin, with no specific resistance mechanisms described, except for the *mcr* plasmid-encoded gene (Gao et al., 2016; Yin et al., 2017), absent in our isolates. In this case, polymorphisms in genes encoding heme uptake systems might be associated with the observed increased MIC to colistin, in agreement with previous observations from our research group (Garmendia et al., 2014).

We have tackled the high genetic diversity within our NTHi collection, with a polymorphism range from 40,000 to 70,000 SNPs, when comparing NTHi genomes from isolates not belonging to the same CT. Moreover, our phylogenomic analysis incorporated all the high quality NTHi genomes publicly available. Based on these data, at least two observations can be made. First, we observed a random distribution of our COPD isolates along the phylogenomic tree. This emphasizes the high genomic diversity previously pointed by other researchers and strain collections (Berrens et al., 2007; Farjo et al., 2004; Lacross et al., 2008; Smith-Vaughan et al., 1998), and the unlikely possibility to cluster NTHi genomes by pathological sources of isolation, as previously pointed (De Chiara et al., 2014). Second, we unexpectedly found examples of high genomic closeness for NTHi isolates belonging to different patients. As an example, seventeen strains belonging to four patients clustered together within the same CT, and consequently within the same branch of the tree. As a further matter, these four patients personal addresses were between 100 and 800 meters apart, leading us to speculate possible inter-patient transmission.

This study also allowed analyzing evolution dynamics based on recurrent polymorphisms detected in the same genes among diverse CTs, thus suggesting parallel evolution during NTHi patho-adaptation within the lower airways of COPD patients. We focused on the *fadL* gene because it showed recurrent polymorphisms classified as high impact by SnpEff software. This tool annotates variants and predicts their effect on genes by using an interval forest approach, and *fadL* was shown to be the highest hit due to frame shift and non-sense mutations, unveiling strong selective pressures for FadL loss-of-function. In an attempt to assign biological significance to the observed genetic variation, we established the role of $FadL_{NTHi}$ as a bacterial ligand of the human cell receptor hCEACAM1 through *in vitro* invasion assays in a cell line expressing hCEACAM1, confirming previous observations (Tchoupa et al., 2015). Nevertheless, $FadL_{NTHi}$ loss-of-function prompted us to speculate some deleterious function for $FadL_{NTHi}$. Since FadL is

also a fatty acid transporter, and NTHi lacks the β-oxidation machinery necessary to use fatty acids as a sole carbon and energy source, we wondered the implications of fatty acid uptake without a metabolic benefit. In fact, a bactericidal effect for AA on *H. influenzae* has been previously described (Knapp & Melly, 1986). This previous observation, together with the frequent FadL loss-of-function in this NTHi strain collection, prompted us to speculate that these two observations may be related. Thus, a AA bactericidal assay by comparing isogenic strains with or without a functional *fadL* gene revealed that *fadL* inactivation increases bacterial viability in the presence of AA. Trying to merge the biological meaning of FadL functionality as a bacterial ligand for a host cell receptor and as a bacterial fatty acid transporter, we conclude that this may be a case of antagonistic pleotropic behavior at the bacterial membrane. We speculate that, at a first stage of COPD airway colonization, $FadL_{NTHi}$ could facilitate the bacterial interplay with host cells resulting beneficial; in contrast and at a subsequent stage, loss of $FadL_{NTHi}$ function may result in a decreased bactericidal effect by free fatty acids. This is likely to be physiologically relevant at the COP lung because AA metabolites are key players in COPD-related airway inflammation (Malhotra et al., 2012). Interestingly, this speculation is reinforced by the fact that, when looking at all publicly available NTHi genomes and their pathological origins, pulmonary infection-causing NTHi strains lack a functional FadL protein at a higher percentage than strains from a non-pulmonary infectious origin do so. This strongly suggests that $FadL_{NTHi}$ loss-of-function is an adaptive trait of NTHi evolutionary dynamics within the lung of chronic respiratory patients.

CONCLUSIONS

Conclusions:

1.- Colonizing *Haemophilus parasuis* isolates recovered from the nasal cavities of healthy piglets showed a heterogeneous ability to form biofilms of proteinaceous nature *in vitro*. Such isolates induced comparable levels of IL-8 pro-inflammatory cytokine release by cultured swine epithelial cells.

2.- The screen of antibiotic resistance determinants in a collection of 86 *H. parasuis* isolates recovered from the nasal cavity of 88 healthy piglets revealed the existence of a novel plasmid (named pJMA-1, size 2,661 bp), with ColE1 replication origin, and bearing the bla_{ROB-1} gene, which may confer β-lactam antibiotic resistance.

3.- The pJMA-1 plasmid is stably maintained through serial passages in the absence of antibiotic pressure when introduced in a heterologous host of the *Pasteurellaceae* family (*Haemophilus influenzae* strain Rd KW20).

4.- Introduction of the pJMA-1 plasmid into a heterologous host of the *Pasteurellaceae* family (*Haemophilus influenzae* strain Rd KW20) has a low impact on this bacterial fitness.

5.- Transformed Recombinant Enriched Profiling (TREP) is a useful strategy for gain-of-function genetic mapping, applicable to naturally transformable bacterial species with high genomic diversity, like *Haemophilus influenzae*.

6.- The use of TREP to screen *H. influenzae* invasin-encoding genes identified that natural transformation-based acquisition of the *hmw* gene is a genetic trait conferring high epithelial adhesion and invasion rates.

7.- TREP specifically identified the $hmw_{86\text{-}028NP}$ gene allele as responsive for NTHi high epithelial adhesion-invasion, as shown by natural transformation-based allelic replacement.

8.- The hyperinvasive $hmw_{86\text{-}028NP}$ gene allele favors four-stage bacterial interaction dynamics with human epithelial cells, including self-aggregation, cell surface adhesion, invasion, and intracellular aggregated location in subcellular compartments co-localizing with the Lamp-1 endosomal marker.

9.- Whole genome sequencing of a collection of 94 NTHi strains longitudinally recovered from sputum samples of thirteen COPD patients revealed persistent, polyclonal and high genomically diverse infection by NTHi within the COPD lung.

10.- Antimicrobial susceptibility testing showed variation in NTHi resistance to cotrimoxazole and polymixin E among clonaly related strains isolated from the same patient. Variations within the *folP* and *folH* gene sequences likely to be associated to cotrimoxazole resistance were identified. Polymorphisms in genes encoding outer membrane proteins were identified in strains with different level of polymixin E resistance.

11.- The *hmw1A* gene phase-variation has been shown to represent a patho-adaptive trait among longitudinally collected COPD isolates. Increasing number of 7-bp tandem repeats within this gene promoter region correlated with its decreasing protein levels and epitelial invasion rate.

12.- The search of recurrent coding mutations in NTHi genomes belonging to the same clonal type identified a panel of bacterial genes showing parallel evolution.

13.- The *fadL* gene was top ranked after the analysis of high impact protein-coding polymorphisms in the WGS collection of NTHi COPD strains. Its variation consisted of both frameshift and nonsense mutations.

14.- The *fadL* gene was present in all isolates within the collection, showing high diversity; non-functional *fadL* alleles were found in 11 out of the 13 COPD patients, distributed among 14 out of 40 clonal types in study, and 24 out of 92 NTHi isolates.

15.- FadL is a NTHi bifunctional protein predicted to locate at the bacterial outer membrane. FadL is a bacterial ligand for the hCEACAM1 host cell receptor, and its loss-of-function impairs NTHi invasion of hCEACAM1-expressing epithelial cells. Moreover, FadL is an exogenous fatty acid transporter, and its loss-of-function increases NTHi survival in the presence of bactericidal fatty acids. FadL allelic variation and its phenotypic consequences imply antagonistic pleiotropy.

Conclusiones:

1.- Cepas de la bacteria *Haemophilus parasuis* que colonizan la cavidad nasal de lechones sanos mostraron heterogeneidad fenotípica respecto a su capacidad para formar biopelículas de naturaleza proteica *in vitro*. Asimismo, dichos aislados estimularon la secreción de niveles comparables de IL-8 por células epiteliales porcinas en cultivo.

2.- El escrutinio de determinantes de resistencia antibiótica en una colección de 86 cepas de *H. parasuis* aisladas de la cavidad nasal de 88 lechones sanos reveló la presencia de un nuevo plásmido (denominado pJMA-1, con un tamaño de 2.661 pb). pJMA-1 tiene origen de replicación ColE1, contiene un gen que codifica la β-lactamasa ROB-1 (bla_{ROB-1}), y su presencia confiere resistencia a antibióticos β-lactámicos.

3.- El plásmido pJMA-1 puede ser introducido en un hospedador heterólogo de la familia *Pasteurellaceae* (*Haemophilus influenzae* cepa Rd KW20), en el que se mantiene estable tras realizar pases seriados en ausencia de presión antibiótica.

4.- La introducción del plásmido pJMA-1 en un hospedador heterólogo de la familia *Pasteurellaceae* (*Haemophilus influenzae*, cepa Rd KW20) tiene bajo impacto en el *fitness* de esta bacteria.

5.- La evaluación del enriquecimiento de recombinantes generados mediante transformación (*Transformed Recombinant Enriched Profiling*, TREP) es una estrategia de mapeo genético mediante ganancia de función aplicable a especies bacterianas que son competentes naturales y presentan alta diversidad genómica entre aislados, como es el caso de *Haemophilus influenzae*.

6.- El uso de TREP para la búsqueda de genes que codifican invasinas de *H. influenzae* identificó que la adquisición del gen *hmw* puede aportar altas tasas de adhesión e invasión epitelial a este patógeno bacteriano. Esta ganancia de función se mostró mediante transformación natural y adquisición alélica.

7.- TREP identificó el gen *hmw* de la cepa HiNT 86-028NP (alelo $hmw_{86\text{-}028NP}$) como un factor específicamente asociado a altos niveles de adhesión e invasión epitelial por HiNT. Esta especificidad alélica se mostró mediante transformación natural e intercambio alélico.

8.- El alelo hiperinvasivo $hmw_{86\text{-}028NP}$ puede favorecer la interacción de HiNT con células epiteliales humanas a través de una dinámica con cuatro etapas: auto-

agregación bacteriana, adhesión epitelial, invasión, localización intracelular formando grupos de bacterias alojadas en compartimentos subcelulares que co-localizan con el marcador de endosoma tardío Lamp-1.

9.- La secuenciación de genomas completos de una colección de 94 cepas de HiNT aisladas longitudinalmente a partir de muestras de esputo de 13 pacientes EPOC a lo largo del tiempo reveló que la infección pulmonar por HiNT asociada a la EPOC es persistente, policlonal y presenta alta diversidad genómica entre aislados.

10.- El análisis de susceptibilidad antibiótica en una colección de aislados HiNT EPOC mostró variación en los niveles de resistencia a cotrimoxazol y a polimixina E entre cepas clonales aisladas de un mismo paciente. Identificamos variaciones en los genes *folH* y *folP* que pueden estar asociadas con la adquisición de resistencia a cotrimoxazol. Identificamos polimorfismos en genes que codifican proteínas de membrana externa en aislados con diferente nivel de resistencia a polimixina E.

11.- La variación de fase del gen *hmw1A* es un rasgo pato-adaptativo observado en aislados de HiNT recogidos longitudinalmente. El aumento progresivo del número de repeticiones de un hepta-nucleótido presente en el promotor de este gen conlleva un descenso progresivo de sus niveles de proteína y de invasión epitelial.

12.- La búsqueda de mutaciones recurrentes en regiones codificantes de los genomas clonales de HiNT generados y analizados en este trabajo identificó un panel de genes bacterianos que muestran rasgos de evolución paralela.

13.- El análisis del impacto de polimorfismos en regiones codificantes de los genomas de la colección de cepas de HiNT analizados en este trabajo identificó el gen *fadL* como un candidato con fuertes rasgos de evolución paralela. La variación en este gen consistió en rupturas en el marco de lectura, y en mutaciones sin sentido por la aparición de codones de parada prematuros.

14.- El gen *fadL* está presente en todos los aislados HiNT de la colección analizada en este trabajo, en la que mostró una gran diversidad. Encontramos variantes alélicas no funcionales de *fadL* en 24 genomas, distribuidos en 11 de los 13 pacientes EPOC muestreados, y en 14 de los 40 tipos clonales determinados.

15.- FadL es una proteína de HiNT con localización predicha en su membrana externa, y con doble función. Por una parte, es un ligando bacteriano del receptor

eucariota hCEACAM1, y su pérdida de función reduce la tasa de invasión de HiNT en células epiteliales que expresan hCEACAM1. Por otra parte, FadL es un transportador de ácidos grasos exógenos, y su pérdida de función aumenta la supervivencia bacteriana en presencia de ácidos grasos con efecto bactericida. La variación alélica de FadL y las consecuencias fenotípicas de dicha variación presentan rasgos de antagonismo pleiotrópico durante la infección del pulmón EPOC por HiNT.