Comparative and transcriptional analysis of the predicted secretome in the lignocellulose-degrading basidiomycete fungus *Pleurotus ostreatus*

**Manuel Alfaro,[1] Raúl Castanera,[1] José L. Lavín,[1,2]
Igor V. Grigoriev,[3] José A. Oguiza,[1]
Lucía Ramírez[1] and Antonio G. Pisabarro[1]***

[1]*Department of Agrarian Production, Genetics and Microbiology Research Group, Public University of Navarre, Pamplona 31006, Spain.*

[2]*Genome Analysis Platform, CIC bioGUNE & CIBERehd, Bizkaia Technology Park, Derio 48160, Spain.*

[3]*US Department of Energy Joint Genome Institute, Walnut Creek, CA, 94598, USA.*

Summary

Fungi interact with their environment by secreting proteins to obtain nutrients, elicit responses and modify their surroundings. Because the set of proteins secreted by a fungus is related to its lifestyle, it should be possible to use it as a tool to predict fungal lifestyle. To test this hypothesis, we bioinformatically identified 538 and 554 secretable proteins in the monokaryotic strains PC9 and PC15 of the white rot basidiomycete *Pleurotus ostreatus*. Functional annotation revealed unknown functions (37.2%), glycosyl hydrolases (26.5%) and redox enzymes (11.5%) as the main groups in the two strains. When these results were combined with RNA-seq analyses, we found that the relative importance of each group was different in different strains and culture conditions and the relevance of the unknown function proteins was enhanced. Only a few genes were actively expressed in a given culture condition in expanded multigene families, suggesting that family expansi on could increase adaptive opportunities rather than activity under a specific culture condition. Finally, we used the set of *P. ostreatus* secreted proteins as a query to search their counterparts in other fungal genomes and found that the secretome profiles cluster the tested basidiomycetes into lifestyle rather than phylogenetic groups.

Introduction

Fungi in the *phylum* Basidiomycota occupy a large variety of ecological niches: ectomycorrhizal (ECM) fungi form symbiotic relationships with the roots of most tree species (Plett and Martin, 2011; Vincent *et al.*, 2012; Hess *et al.*, 2014); saprotrophic fungi (SAP), such as white rot (WR) (Martinez *et al.*, 2004) and brown rot (BR) (Eastwood *et al.*, 2011), use decaying matter as carbon sources (Stajich *et al.*, 2010); plant pathogens (PPT) invade living tissues of plants to obtain nutrients (Kämper *et al.*, 2006); and animal pathogens (APT) (Loftus *et al.*, 2005) infect animals, including humans. Furthermore, some fungi can switch between different lifestyles over their life cycles (e.g. *Moniliophthora perniciosa*) (Mondego *et al.*, 2008). These different lifestyles are associated with the ability of the fungi to produce high amounts of secreted enzymes that are crucial tools to obtain nutrients, a characteristic that has been widely exploited by industry (Conesa *et al.*, 2001; Shoji *et al.*, 2008). The secretome is defined as the set of proteins secreted by a cell or an organism at a given time (Tjalsma *et al.*, 2000; Alfaro *et al.*, 2014). This definition includes proteins released into the surrounding medium and also proteins that persist anchored to the membrane or cell wall, including integral membrane proteins. The secretome is a very dynamic fungal tool that varies over time depending on the growth substrate, temperature or growth phase. One of the main objectives of fungal secretome studies is to understand secretome modulation due to environmental changes.

Lignocellulose is the largest reservoir of organic carbon on Earth and is a renewable resource that serves not only as the raw material of the pulp and paper industry but also as a promising potential source of second-generation biofuels. Basidiomycetes are the only organisms known to degrade lignocellulose on a global scale by producing a large amount of lignocellulolytic secreted enzymes and playing a key role in the global carbon cycle. From a geological time frame perspective, lignocellulose could have

served as a major carbon sink until the appearance of SAP lignin-degrading fungi at the end of the Carboniferous period (Floudas *et al.*, 2012). Lignin degradation can be achieved using two strategies: WR fungi secrete enzymes that remove the lignin polymer before they start to enzymatically degrade the cellulose, whereas BR fungi barely degrade the lignin and attack the cellulose layer via a Fenton reaction-based mechanism. However, these two lignocellulose-degrading strategies do not seem to be mutually exclusive because the *Jaapia argillacea* and *Botryobasidium botryosum* genome sequences simultaneously show similarities to the WR (presence of gene families coding for carbohydrate and lignin active enzymes) and BR (absence of genes coding for the class II peroxidases) fungi (Riley *et al.*, 2014).

The degradation of the lignin moiety of lignocellulose is achieved by a set of principal enzymes that include manganese peroxidases (MnP, EC 1.11.1.13), versatile peroxidases (VP, EC 1.11.1.16), lignin peroxidases (LiP, EC 1.11.1.14) and phenol oxidases (benzenediol:oxygen oxidoreductase Pox, laccases, EC 1.10.3.2) that are accompanied by other accessory enzymes, such as cytochrome c peroxidases (EC 1.11.1.5), chloroperoxidases (EC 1.11.1.10), dye decolourizing peroxidases (DyP, EC 1.11.1.19), glyoxal oxidases (GLOX), aryl-alcohol oxidases (AAO, EC 1.1.3.7), pyranose dehydrogenases (EC 1.1.99.29), and methanol oxidases (EC 1.113.13). The cellulose moiety is attacked by secreted carbohydrate-active enzymes (CAZy) (Cantarel *et al.*, 2009) such as exocellulases (cellobiohydrolases, CBH, EC 3.2.1.91) that are subclassified as types I and II depending on their attack of the reducing or non-reducing cellulose ends, endocellulases (EC 3.2.1.4), and cellobiases (beta-glucosidases, EC 3.2.1.21). Finally, the cementing role of the hemicellulose is disrupted by the action of endoxylanases (EC 3.2.1.8), α-glucuronidases (EC 3.2.1.131), acetyl-xylan esterases (EC 3.1.1.72), arabinofuranosidases (EC 3.2.1.55), ferulic acid esterases (feruloyl esterases, EC 3.1.1.73) and β-xylosidases (EC 3.2.1.37), among other enzymes (Sun *et al.*, 2012). All of these enzymes are frequently encoded by multi-genic families whose members are differentially regulated [see (Castanera *et al.*, 2012) as an example in *P. ostreatus*].

The identification of the set of secretable proteins encoded in the genome of a fungal species should permit the development of hypotheses about its lifestyle. Bioinfosecretomes are *in silico*-predicted secretomes based on the identification of secretion signals in the predicted proteins using the automatically annotated gene models of a genome sequence (Alfaro *et al.*, 2014). Therefore, the quality of a bioinfosecretome depends on the quality of the genome sequence and annotation. Fungi secrete proteins using either the conventional (CSP) or the unconventional (USP) secretion pathways. CSPs require that secretion-

oriented proteins contain an amino-terminal peptide sequence (N-terminal sequence, signal peptide, SP) that targets them to the endoplasmic reticulum (ER), where they will be correctly folded and post-translationally modified (Conesa *et al.*, 2001). Subsequently, proteins pass through the Golgi apparatus and are transported in secretory vesicles that fuse to the plasma membrane to perform the secretion process (Conesa *et al.*, 2001; Shoji *et al.*, 2008; Fonzi, 2009). Alternatively, proteins secreted by the UPS pathways do not require the presence of an SP sequence to guide the protein through the ER/Golgi pathway (Nickel and Seedorf, 2008). The occurrence of UPS pathways in fungi has been well documented in the case of yeasts (Nombela *et al.*, 2006). The computational prediction of the bioinfosecretome is based on the initial screening of the required SP sequences in the secreted proteins (Caccia *et al.*, 2013) using programs such as SignalP (Bendtsen *et al.*, 2004) and TargetP (Emanuelsson *et al.*, 2007), followed by a subsequent scan to determine the presence of transmembrane motifs. This approach was used to identify 168 secreted proteins, including 39 glycosylphosphatidyl inositol (GPI)-anchored proteins, encoded in the genome of the basidiomycete maize smut *Ustilago maydis* (Mueller *et al.*, 2008). Additionally, the first bioinfosecretome of a WR basidiomycete was constructed for *Phanerochaete chrysosporium* (Vanden Wymelenberg *et al.*, 2005) using the first genome version. It consisted of 268 secretion-predicted proteins (Martinez *et al.*, 2004), although this number could be an underestimation due to inaccurate and incomplete gene model annotation. Subsequently, 769 secreted proteins were identified based on the second assembled version of this genome (Vanden Wymelenberg *et al.*, 2006). Many predicted secreted proteins were classified as glycosyl hydrolases (GH), oxidoreductases, peptidases and esterases-lipases, as expected for the SAP lifestyle of *P. chrysosporium*. Finally, Jain and colleagues (2008) focused on non-classically secreted proteins using the SecretomeP method (Bendtsen *et al.*, 2004) to analyse genome sequences of *Laccaria bicolor, Botrytis cinerea, P. chrysosporium*, and five other fungal species and found proteins involved in carbohydrate metabolism, lectins and proteases. There are currently more than 135 basidiomycete genome sequences available in the MycoCosm database (Grigoriev *et al.*, 2014) from the Joint Genome Institute (JGI, October 2014) (Grigoriev *et al.*, 2012). These sequences offer the opportunity to predict the bioinfosecretomes of these basidiomycetes and correlate them with their corresponding lifestyles in order to identify characteristics that permit the deduction of the set of secreted proteins.

However, the bioinfosecretome is a static list of secretable proteins that does not produce a complete picture of how the toolbox of secreted proteins is actually used. To approximate to its actual use, we propose here to colour
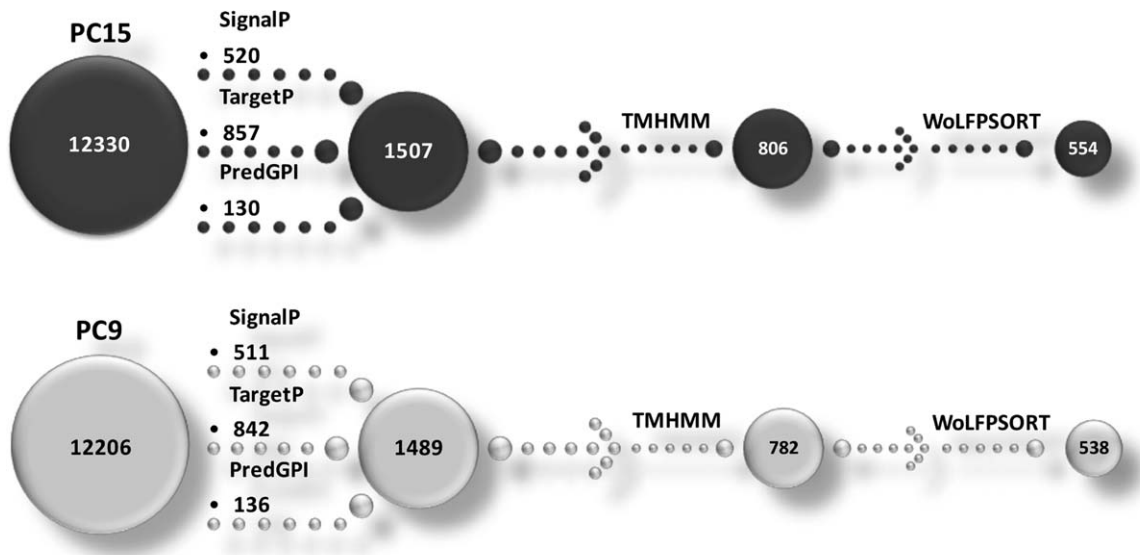
Fig. 1. SECRETOOL pipeline for the prediction of secreted proteins in the *P. ostreatus* PC15 and PC9 monokaryons.

the bioinfosecretome with data derived from whole transcriptome studies of the same individual cultured under different conditions as a proxy for the actual secretome. Moreover, we hypothesize that because the secretome reflects the fungal lifestyle, the analysis of the bioinfosecretome should permit the prediction of the fungal lifestyle. To test this hypothesis, we studied the secreted proteins released into the surrounding medium (membrane-bound proteins were excluded) and compared the bioinfosecretomes of different fungi and their expression under different culture conditions. We show that although the secretome is conserved in different strains of the same species, its expression varies among strains and culture conditions. Moreover, the secretome's transcriptional effort is more similar within strains, even when they are cultured under different conditions, than between different strains cultured under the same conditions. We also show that the expansion of gene families does not imply the expression of more genes in the expanded family. We identify new families of secreted proteins with unknown functions and suggest a function for some of these groups. Finally, we show that fungal secretomes cluster according to the fungal lifestyles rather than according to their taxonomic or phylogenetic classifications.

Results and discussion

The *P. ostreatus* dikaryotic strain N001 is a hybrid strain that has been used as a source for commercial cultures for many years. The two nuclei present in this dikaryotic strain were separated, and two monokaryotic strains (PC9 and PC15) were produced and maintained. The genomes of these two strains have been sequenced. In this paper, we use the term 'strain' to refer to these two monokaryons that are genetically compatible and that reconstitute the dikaryotic strain N001 when mated.

Pleurotus ostreatus *PC9 and PC15 bioinfosecretomes*

The protein models of the *P. ostreatus* PC9 (12,206 proteins) and PC15 (12,330 proteins) monokaryons were analysed with SECRETOOL to identify putatively secreted proteins. A total of 659 gene models were identified as encoding secretable proteins. These included 538 (PC9) and 554 (PC15) models that represented 4.41% and 4.49% of the total number of protein models, respectively (Fig. 1). A total of 433 of these proteins were identified as secretable in both PC9 and PC15 (Fig. 2A and Supporting Information File 1, Table S1), whereas 105 and 121 were predicted to be secreted exclusively by PC9 or PC15 respectively. There was no corresponding allele predicted in PC15 for 23 of the 105 PC9-exclusive proteins (i.e. they were encoded for genes solely identified in the PC9 genome); an allele for the remaining 82 proteins was found in PC15, although it encoded a protein that did not match the criteria for secretion used in the SECRETOOL pipeline. Conversely, an allele in PC9 was not found for 30 of the 121 PC15-exclusive proteins; for the remaining 91 proteins, an allele was found in PC9 that did not encode a protein meeting the secretion criteria established in the SECRETOOL pipeline.

Next, we took a closer look at the genes for which a secretable allele was found in one of the two monokaryons whereas the corresponding allele was annotated as nonsecretable in the other monokaryon (91 PC9 alleles and 82 PC15 alleles = 173 gene models, see above). To
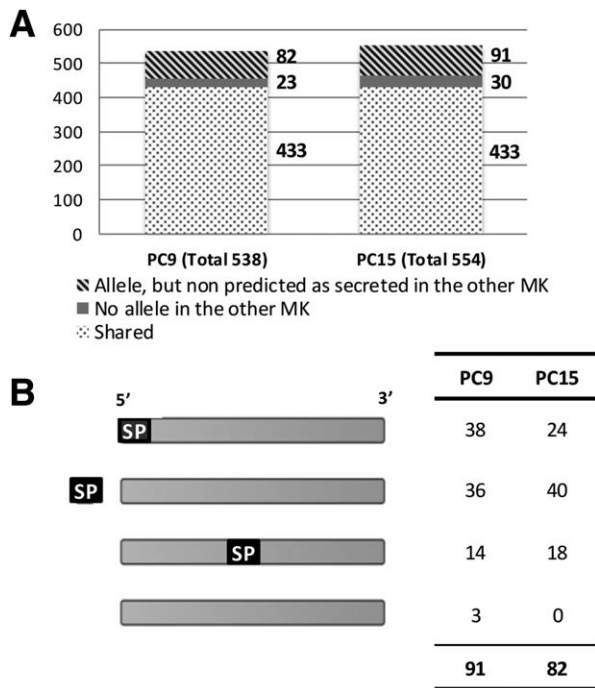
Fig. 2. A. Number of predicted proteins in the *P. ostreatus* monokaryons PC9 and PC15. The complete list arranged by allelic pairs is available in Supporting Information File 1, Table S1. B. Signal peptide position in allelic protein models not predicted to be secreted.

accomplish this, we used the first 210 bp of each model (starting at the initial ATG) as a query in a BLASTN search (Camacho *et al.*, 2009) performed in the genome of the monokaryon in which the corresponding protein was not predicted to be secreted. The rationale was to search for the corresponding signal peptide in the non-secreted alleles. For 172 out of the 173 searches, a sequence similar to the signal peptide query was found in the model not predicted to be secretable (Fig. 2B). There was an exception that corresponded to a gene model identified in PC15 as carrying a signal peptide whose sequence was not found in the entire PC9 genome; thus, this gene model could correspond to a protein secreted by one of the monokaryons but not the other (PC9/PC15 gene model 91216/1109867).

The 172 (82 + 90) gene models could be grouped into three categories (Fig. 2B): (i) gene models in which a sequence similar to the corresponding signal peptide was found in the non-secreted allele at the expected position for a signal peptide (31 PC9 and 22 PC15 gene models); (ii) gene models in which the sequence similar to the corresponding signal peptide was found at different positions upstream of the expected position (36 in PC9 and 39 in PC15); and (iii) gene models in which the sequence similar to the corresponding signal peptide was found at different positions downstream of the start of the corresponding

non-secreted model (23 in PC9 and 21 in PC15). In the case of genes in which the putative signal peptide was found upstream of the predicted gene model, the distance between the signal and the model varied from a few base pairs to 11 Kbp. The signal appeared in the inverted orientation in two of the PC9 models in which the putative signal peptide was far upstream from the gene model. Similarly, there was one PC15 gene model in which the sequence corresponding to the signal peptide was found downstream of the gene model and in an inverted orientation. Finally, in one of the models in which the signal peptide was far upstream from the protein start (PC15 gene model 1103939), the intervening sequence was identified as an LTR-*Gypsy* transposable element.

In summary, 53 gene models (31 PC9 and 22 PC15 gene models under the first category of the preceding paragraph) were discarded as secretable by the SECRE-TOOL pipeline although they had a *bona fide* signal peptide at the expected position. This discard is due to the stringent conditions of the SECRETOOL algorithm. For the remaining 120 gene models, the sequence differences between the alleles could reflect the dynamic nature of the genomes due to identifiable insertion of a transposon-related sequence and inversion events. This result empha-sizes the need for refined automatic annotation systems that prevent artifactual biases in the annotation of genomes.

Because the 173 gene models described above had dif-ferent types of alterations that prevented them from qualifying as encoding secretable proteins according to the standard established by SECRETOOL, we discarded them for the remainder of the studies described in this paper. Moreover, there could be protein models for which neither of the two alleles fulfill the conditions imposed by the SECRETOOL algorithm and have been undetected in this analysis. Consequently, the list of secreted proteins described in this work should not be taken as exhaustive, since stringency has prevailed over sensitivity in the process.

Functional categories of the proteins identified in the predicted P. ostreatus *bioinfosecretome*

Functional characteristics were associated with the *P. ostreatus* putative secreted proteins after the individual manually curated annotation. Figure 3 shows that the func-tional profiles of the PC9 and PC15 predicted secretomes were highly similar. A protein function, being enzymatic or structural (i.e. hydrophobins), could not be predicted for 197 putative secreted proteins in PC9 (36.6% of the SECRETOOL positive models) and 209 in PC15 (37.7%). Most of these proteins with unknown function (145 gene models corresponding to 26% of the PC15 putative secreted protein set) appeared to be conserved in other
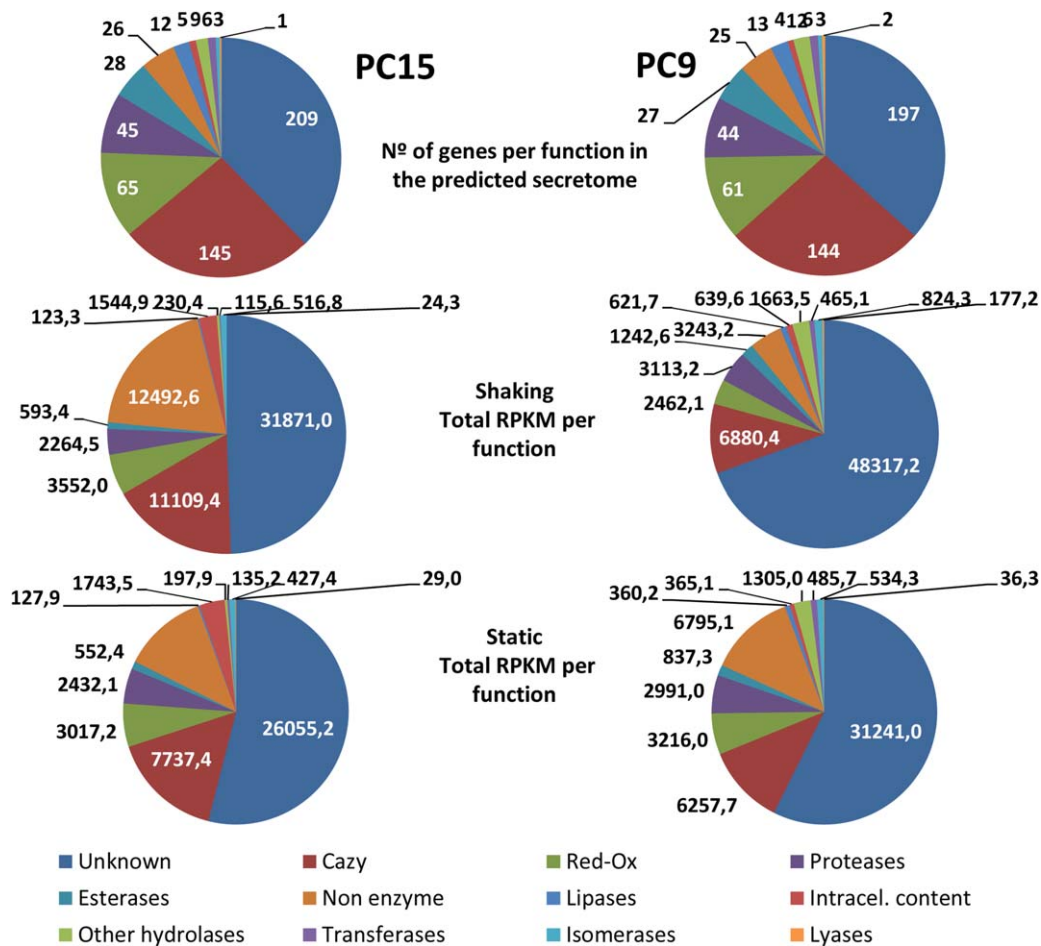
Fig. 3. Function of proteins predicted to be secreted by the *P. ostreatus* monokaryons PC9 and PC15 and a comparison between the number of proteins predicted to be secreted and the number of RPKM by protein function.

basidiomycetes (Supporting Information File 1, Tables S2 and S3). For the rest of the models, the functional groups of carbohydrate active enzymes (CAZy), redox and proteases were the more abundant, accounting for more than 75% of the models with an assigned function.

A total of 166 gene models that could be grouped into 53 CAZy groups were identified as secretable (Supporting Information File 2). The most abundant groups corresponded to AA9 (copper-dependent lytic polysaccharide monooxygenases or LPMOs, formerly classified as group GH61, 20 gene models), GH16 (endoglucanase, 13 gene models) and GH7 (cellobiohydrolase I, 10 gene models). The most abundant Carbohydrate Binding Module (CBM domain) was the ricin-like CBM13 (8 gene models). A total of 70 gene models related to redox reactions were identified. These models could be further classified into 17 functional groups of which those corresponding to the GMC oxidoreductases (17 gene models), proteins containing a cupredoxin motif (11 gene models) and laccases (11 gene models) were the most abundant. A total of 53

gene models encoding different types of peptidases could be classified into 18 classes, among which peptidase type A1 (aspartic endopeptidases) and S8 (subtilases) were the most abundant (6 gene models each). A total of 12 different groups of secretable esterases were identified within 33 gene models that included the ribonucleases (7 gene models) and carboxylesterases (6 gene models, the most abundant). Finally, a total of 16 lipase genes were identified, of which 14 were classified as lipase 3. Four PC15 and three PC9 lipase 3 gene models were identified in only one of the two strains without an allele in the other strain.

Given the relevance of the group of gene models encoding proteins with unknown functions, we used a complementary approach to generate hypotheses about their functions. We used the Phyre2 web server (Kelley and Sternberg, 2009), which uses the more sensitive PSI-BLAST algorithm, to predict the structures and functions of these proteins, and then we used MCL (Enright *et al.*, 2002) to cluster them into families. This approach allowed us to identify two protein clusters
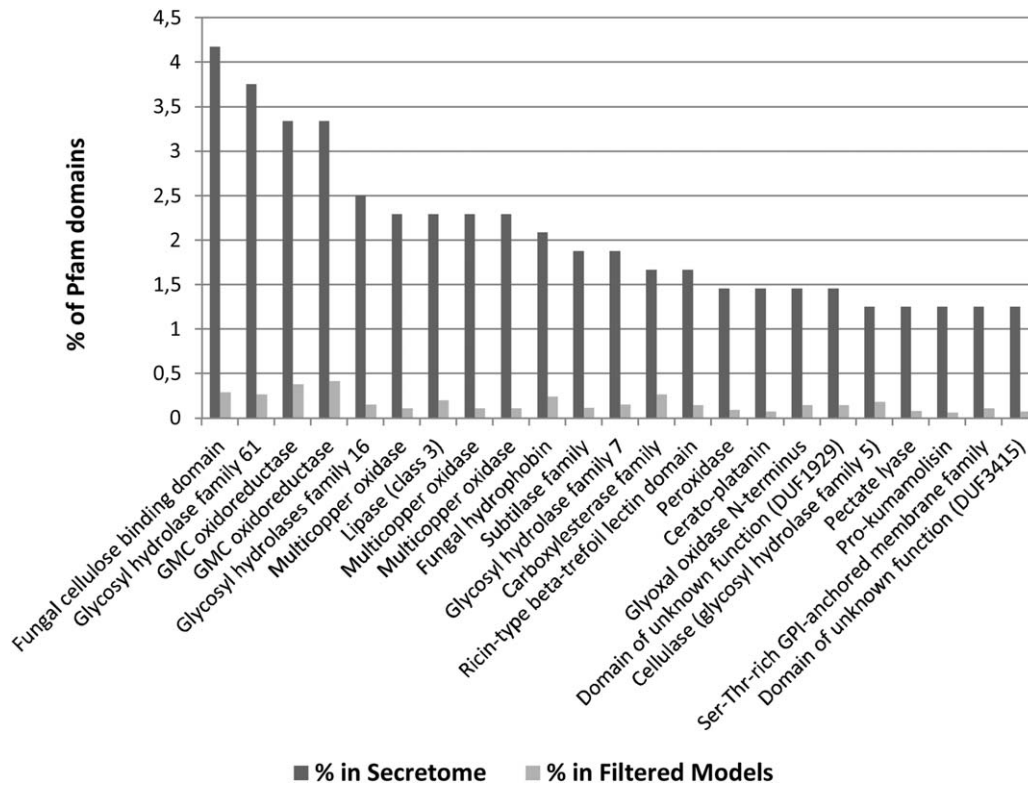
Fig. 4. Percentage of domains from a Pfam category in the bioinfosecretome and in all filtered models of the genome sequence of *P. ostreatus* PC15. Only Pfam categories with at least six domains found in the secretome are shown. The complete list is available in Supporting Information File 1, Table S7.

with a common predicted function among the family members (Supporting Information File 1, Table S4). The first cluster was formed by 16 proteins (10 of which could be folded with a level of confidence higher than 90%) as putative cell invasion proteins with a MAC/perforin domain (MACPF) (Lukoyanova *et al.*, 2015). A protein with this domain that was involved in transmembrane pore formation (Pleurotolysin) was previously described as a cytolysin present in the basidiocarp of *P. ostreatus* (Bernheimer and Avigad, 1979; Tomita *et al.*, 2004). Similarly, pore-forming proteins have been described in other eukaryotes and bacteria, and many play key roles in immunity and pathogenesis (Rosado *et al.*, 2007). The second cluster was formed by four proteins whose predicted structures and functions matched a bacterial xylan esterase (confidence > 95%) of the soil bacterium *Cellvibrio japonicum*. This saprophytic bacterium has been experimentally shown to degrade all of the major plant cell wall polysaccharides, including xylan (DeBoy *et al.*, 2008). Both groups have similar proteins that are present in many other basidiomycete genomes, where they have also been annotated as proteins without a predicted function (Supporting Information File 1, Tables S5 and S6).

These results emphasize the need to use structure-based algorithms for the prediction of functions in proteins

with unknown functions because they greatly diverge in sequence but maintain a more conserved global structure.

Pfam enrichment in the bioinfosecretome of P. ostreatus

To elucidate whether there was an enrichment of functional categories in the secretome *versus* the whole genome, both the bioinfosecretome protein models and all the filtered models of the *P. ostreatus* PC15 genome were subjected to a Pfam domain search using the InterPro standalone program (Apweiler *et al.*, 2000; Bateman *et al.*, 2004). All of the Pfam categories overrepresented in the secretome are listed in Supporting Information File 1, Table S7, and those with an enrichment higher than sixfold are shown in Fig. 4. Pfam domains involved in lignocellulose degradation were overrepresented in the *P. ostreatus* PC15 bioinfosecretome. The most overrepresented Pfam domains were the multi-copper oxidase domain (which can be found in laccases and other proteins), the Pfam domains related to small secreted proteins hydrophobins and ceratoplatanins, that are about 100 ± 25 amino acids long, cysteine-rich, hydrophobic fungal surface proteins able to self-assemble *in vitro* (Peñas *et al.*, 2002; Sbrana *et al.*, 2007), and ricin type beta trefoil lectin domains]

which are always among the most highly expressed genes in the *P. ostreatus* cultures as shown below, and domains involved in cellulose or hemicellulose degradation (families of GH5 and GH16 endoglucanases, GH7 cellobiohydrolase I, AA9/GH61, pectate-lyases and fungal cellulose-binding domains). Overall, the prevalence of Pfam domains related to lignocellulose degradation in proteins predicted to be secreted compared with the complete genome proteins demonstrated the influence of the ecological niche on the fungal secretome composition.

Transcriptional evaluation of the PC9 and PC15 secretomes

The bare enumeration of the genes encoding putative secreted proteins does not provide information about the relevance of each model or functional group in the performance of the organism. To investigate this relevance, we used the data obtained in six transcriptome analyses of the PC9 and PC15 monokaryons and N001 dikaryon grown in rich medium (SMY) under static and shaking culture conditions. For this purpose, we scored the expression data values (RPKM) for each of the models identified with SECRETOOL. Expression was detected for 93.9% (505/538) and 91.5% (507/554) of the genes encoding secretable proteins in PC9 and PC15 respectively. Specifically, 99.3% (430/433) of the secretable proteins with alleles in the two protoclones were expressed by at least one of them. If we focused on the secretable proteins identified in only one of the two protoclones (105 and 121 for PC9 and PC15, respectively, Fig. 2), 100 (95.2%) and 109 (90.1%) showed expression signals in PC9 and PC15 respectively. In summary, among the 659 *P. ostreatus* proteins predicted to be secreted by at least one monokaryon, transcripts of 639 (96.97%) were expressed in liquid cultures (Supporting Information File 2).

The transcriptome pie chart of the expression values of the different gene models encoding secretable proteins differs significantly from the chart containing only the gene models (Fig. 3). From the perspective of the large groups of secreted proteins (i.e., those described in the pie-chart), in all cases the weight of the genes without an assigned function increased to more than 50%. Moreover, the transcriptome profiles of the two monokaryons differed more when they were cultured under shaking conditions than when they were cultured under static conditions. Although liquid cultures do not represent physiological conditions, the cake-like growth observed in the static cultures could be more similar to the solid condition than the pellet growth of the shaken cultures. From this standpoint, the transcriptomic response to the environmental conditions of both strains was more similar in the static than the shaken cultures.

If we focus our study to the gene or gene family level, it is possible to study the conservation of gene expression within strains cultured under different conditions and between different strains cultured under the same conditions (Fig. 5). This analysis revealed a high correlation of gene expression between samples of the same strain cultured under different conditions (static or shaking, determination coefficients higher than 85%), whereas there was a low correlation of gene expression between samples of different strains cultured under the same conditions (determination coefficient lower than 50%). This result suggests that the gene by gene expression profiles of the secretome differ more among strains than between the two conditions tested within a given strain and emphasizes the convenience of studying the transcriptome of different independent genetically related individuals to identify genes whose expression is environmentally regulated.

The study of the accumulated gene expression values indicated that between 11 and 16 genes were responsible for 50% of the total expression under each condition. In PC15, 61 (static) and 65 (shaken) genes scored RPKM values higher than 100, whereas these values were exhibited by 79 genes in the two culture conditions in PC9. The genes with transcription signals higher than 100 represented nearly 90% of the total expression signals of the bioinfosecretome. In summary, most genes encoding secretable proteins were expressed, although this expression was concentrated in less than 25% of them.

Four genes were among the most highly expressed genes in the four analysed transcriptomes (Supporting Information File 2). Highly conserved counterparts for these four genes were found in other fungal genomes available. The first one (PC9/PC15 models 115427/1089609) encoded a small secreted protein (SSP, 128 amino acids) with unknown function that contains an 80 amino acid region well conserved in proteins found in other fungal genomes. Within this conserved region, there are some residues compatible with a putative Pfam10342 domain (glycosyl-phosphatidyl-inositol-anchored, GPI-anchored) suggesting that this protein can remain trapped in the cell wall when exported. The second gene (PC9/PC15 models 117301/1089988) was automatically annotated and KOG-described as β-1,6-acetylglucosaminyltransferase. The third (PC9/PC15 models 114382/172522) encoded a barwin-like protein that was distantly related to plant expansins containing a double-$\varphi\beta$-barrel fold; incidentally, this fold was also present in cerato-platanins, whose genes were also among the top expressed genes in PC15 under both culture conditions. The fourth (PC9/PC15 models 116525/1067734) encoded a protein conserved across the fungal kingdom for which no functional clues are available. In summary, the highest transcriptional effort in all strains and culture conditions was focused on the synthesis of proteins associated with cell wall constituents with different putative activities.
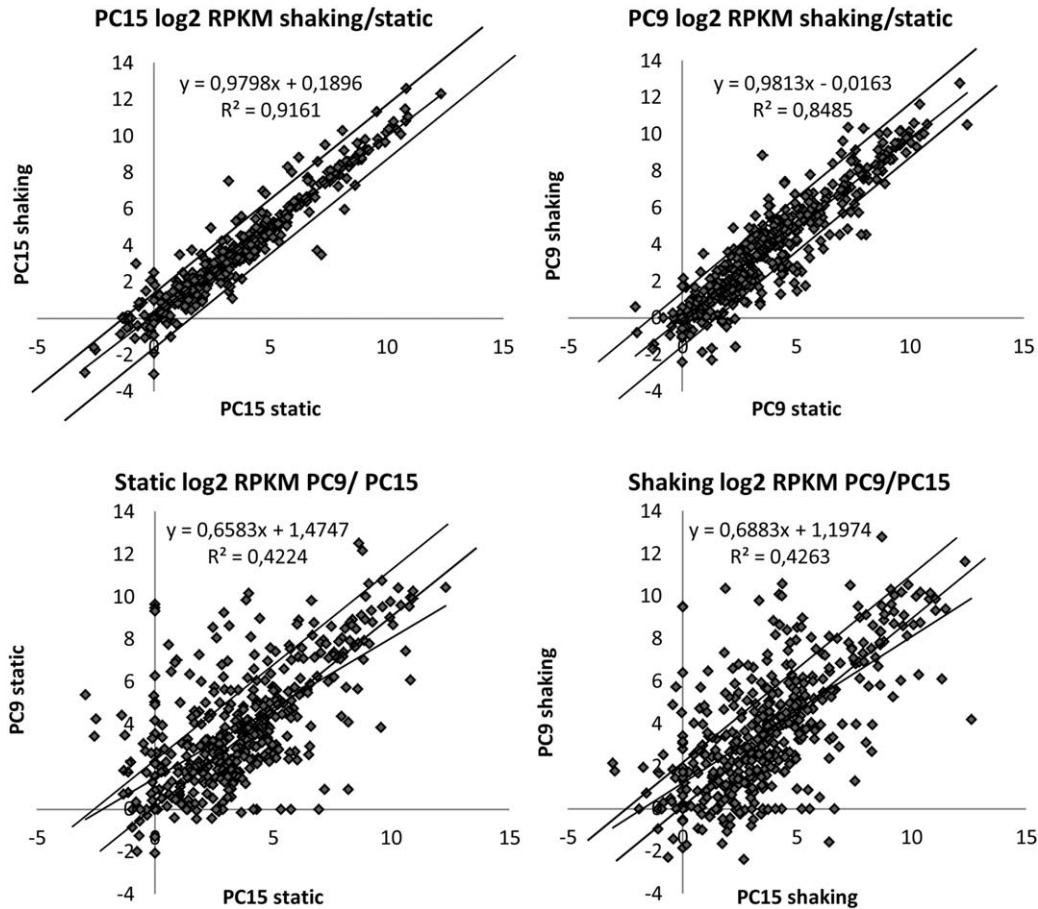
Fig. 5. Gene expression (log2 RPKM) comparison within strains cultured under different conditions (static or shaking) and between different strains (PC15 and PC9) cultured under the same conditions.

If we study the top expressed genes encoding secreted proteins in PC15 under both conditions, we can see that these genes are also among the most expressed in PC9 (both conditions). However, the four more expressed genes in PC9 (static growth) were not among the most expressed genes in PC15 (both conditions). These four PC9 expression-enhanced genes encoded hydrophobin Vmh3 (Peñas *et al.*, 2002) (PC9/PC15 models 80078/1114379) which was the most expressed gene encoding for secretable proteins in PC9sta, and three SSPs (192, 256 and 145 aminoacids) with unknown functions.

CAZy-encoding genes

The expression of the genes encoding secreted CAZy was biased towards a limited number of families. The five most highly expressed families accounted for more than 66% of the total CAZy expression level, and the top ten represented nearly 90% of the total expression of this gene group. In all cases, the gene-rich families of the lytic polysaccharide monooxigenase (LPMO, AA9/GH61), the glucanase of the GH16 family and

CBM13-containing proteins were among the top five most highly expressed CAZy genes. Surprisingly, two CAZy families contained few members with expression among the top three in all cases: the barwin-like endoglucanases (GH45) and the PL14 (glucuronan lyase) families contained only two genes each. The correlation coefficient of the global transcriptomes of the four secretomes was always higher compared to the corresponding coefficients for the top ten most highly expressed gene families (Supporting Information File 2). Finally, we observed that one gene model was preferentially expressed in each gene family, that the same gene model was the most highly expressed in all conditions and, that the expression of this predominant gene represented more than 75% of the gene family expression in most cases. The most obvious exception to this pattern was the gene family GH16, for which two gene models were preferentially expressed at similar levels under each condition; their combined expression accounted for between 58 and 74% of the expression of this gene family (Supporting Information Fig. S2).

In summary, the genes encoding secreted CAZy included 53 gene families and 166 gene models that represented

roughly a quarter of the *P. ostreatus* bioinfosecretome. However, if we consider these genes weighted by their expression values, most of the CAZy expression was concentrated in a much smaller number of gene families and furthermore in only one or two genes per family. Notwithstanding, expression was detected for most of the CAZy-coding genes, albeit at a rather low level.

Redox enzymes

Among the 17 gene families classified as redox, genes encoding proteins containing the cupredoxin motif (cupredoxin family, 11 gene models in addition to the 11 laccase models that also contain the cupredoxin fold, see below) were the most highly expressed in all of the strains and conditions tested and accounted for 44–45% and 58–63% of the family expression in PC9 and PC15 respectively (Supporting Information File 2). The second and third most highly expressed gene families in all tested conditions were the glyoxal oxidase and the GMC oxidoreductase families respectively. Within the cupredoxin family (11 members), more than half of the family's gene expression corresponded to a single gene model in PC15, whereas two genes (one that was the most expressed allele in PC15) were needed to accumulate this expression level in PC9. In the GMC oxidoreductases (17 genes), between two (PC9 shaken) and four genes (PC15 both conditions) were needed to accumulate 50% of the family expression. Gene expression within the glyoxal oxidase family was strongly biased towards the expression of a single gene that accounted for more than 90% of the family's gene expression under all conditions. Finally, the phenol oxidase family (laccases, 11 genes) was especially relevant in the WR fungus *P. ostreatus*. In this family, 50% of the accumulated expression was due to the gene model encoding laccase 6 either alone (PC9 static) or accompanied by a second gene model: laccase 3 in PC15 and laccase 12 in PC9 shaken. These results were similar to those observed by Castanera and colleagues in glucose-based liquid media using RT-qPCR (Castanera *et al.*, 2012, 2013). The manganese-peroxidase and versatile-peroxidase encoding genes displayed a low expression level in all strains and conditions tested (Supporting Information File 2 and Fig. S2).

In summary, similar to the observations in the CAZy gene family, most of gene expression in the redox group of families depended on a limited number of gene families and a limited number of predominantly expressed genes within each family. Notably, the most highly expressed families in the CAZy and redox categories corresponded to proteins using copper as a cofactor.

Peptidases

The accumulated expression of the peptidase-encoding genes was similar under the four strains/conditions tested;

however, the detailed expression profiles were different. In PC15, only four out of 18 gene families displayed an expression level higher than 100 RPKM, whereas six families were over this threshold in PC9. Three of the gene families ranked high for expression under all four conditions (S8 serine endopeptidases, M36 metalloendopeptidases using Zn as a cofactor, and a protein annotated as similar to the disintegrin). Moreover, one family was selectively expressed in PC15 (serine-carboxypeptidase S10) and two in PC9 (peptidase A1 and S53).

Esterases

Of the 12 gene families classified as encoding esterases, two accounted for more than 50% of the expression in PC9 and PC15, although they differed between the two strains. The phospholipase C-like gene family was the most highly expressed in all strains and conditions except in PC9-Shk, where it ranked second. However, the next most highly expressed esterase family was different in the two strains: arylesterase in PC15 cultured under static and shaken conditions and carboxylesterases and genes classified as putative esterases/lipases in PC9 cultured under static and shaken conditions. The two richer esterase gene families (carboxylesterases, six gene models, and ribonucleases, seven gene models) exhibited higher expression in PC9 than in PC15 (217.9 vs. 84.96 for the carboxylesterase family and 137.24 vs. 20.73 for the ribonuclease family for PC9 and PC15 respectively). In all case, this increase in the relative transcription was due to only one of the family members.

Lipases

For the 16 genes identified as lipases, the largest gene family was lipase 3 (14 genes). This family was the most highly expressed in all strains and conditions, accounting for between 56.86 and 95.25% of the family expression. Within this family, the most expressed genes differed in the four strains and conditions tested, suggesting that each strain adapted to the different culture conditions in different ways.

Non-enzymes

The group of secreted proteins functionally classified as non-enzymes was overrepresented in terms of its transcription level compared with its representation in the number of proteins (Fig. 3). Hydrophobins are structural proteins widely studied in fungal systems. The nine hydrophobin genes identified by SECRETOOL have been found expressed in the *P. ostreatus* cultures. However, most of their expression corresponded to the gene coding for the Vmh3 protein, a glycosylated hydrophobin found detected
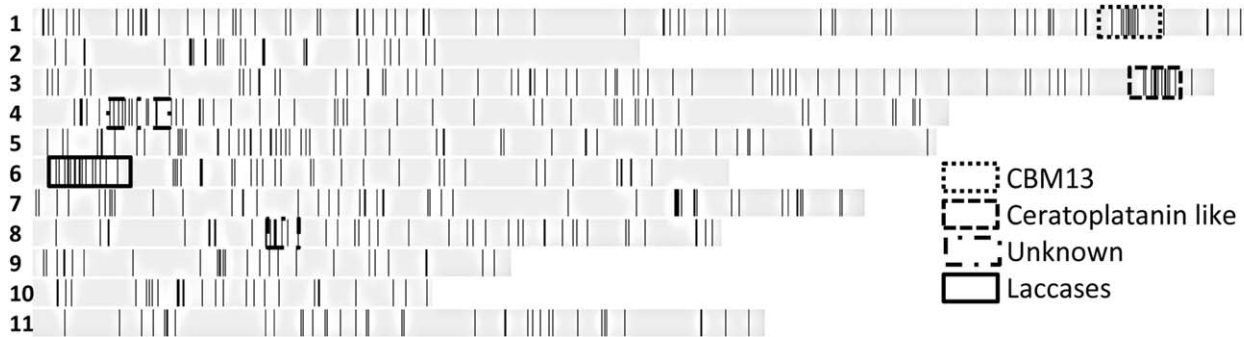
Fig. 6. Location of secreted proteins in the chromosomes, three selected clusters of *P. ostreatus* PC15 and clusters of secreted proteins (REEF program, FDR corrected *P*-value < 0.05).

during vegetative growth and in fruit bodies. This protein is supposed to play a role in development similar to that proposed for SC3 in *Schizophyllum commune*. (Peñas *et al.*, 2002). Vmh3 expression represented more than 94 and 69% of the expression of the identified hydrophobins in PC9 and PC15 respectively. The expression of these genes was increased approximately eightfold in PC9 compared with PC15; this difference was due to the high expression of the *vmh3* gene in PC9. In both strains, the total hydrophobin expression level was higher in the static compared to the shaken cultures. This difference in gene expression in the static cultures could be related to the aerial growth observed in the mycelium cakes produced in the static cultures compared with the pellet structure produced in the submerged shaken cultures.

Cerato-platanins are a group of small secreted proteins that were first identified in the fungal pathogen *Ceratocystis platani* as plant defense elicitors. There are seven genes encoding cerato-platanins in *P. ostreatus*. The expression of the cerato-platanin genes was concentrated in two genes that accounted for more than 50% of the family expression (one gene in PC9-Shk). Cerato-platanin genes were highly expressed in PC15 (especially in shaken cultures), where the two most highly expressed cerato-platanin genes ranked as the second and fifth most expressed genes encoding secretable proteins. The biochemical functions of cerato-platanins remain elusive. De Oliveira and colleagues (2011) showed that they had a structure similar to expansins and barwin-like endoglucanases, suggesting an activity involved in carbohydrate metabolism (Baccelli *et al.*, 2014; Gaderer *et al.*, 2014). However, they have no activity against a variety of carbohydrates. If we consider the *P. ostreatus* cerato-platanin-like proteins as CAZY enzymes and analyse them in conjunction with the genes encoding proteins annotated as Barwin-like endoglucanases (which are the five most highly expressed CAZy genes), the sum of these two groups represents the majority of all secreted genes.

Genome mapping of the P. ostreatus *secretome genes*

We studied the mapping positions of the genes encoding secretable proteins using the PC15 strain because it had a better genome assembly. The gene models encoding the 554 secreted proteins appeared to be distributed across the 11 scaffolds (chromosomes) of the *P. ostreatus* PC15 genome (Fig. 6). The number of genes encoding secretable proteins mapping to each chromosome increased linearly with the chromosome size ($R^2 = 0.92$) (Supporting Information File 1, Table S8). Using the Reef program (Coppe *et al.*, 2006), we identified five clusters of secreted proteins (Supporting Information File 1. Table S9): two clusters mapping to the subterminal region of chromosome 4 and to an internal region of chromosome 8 were mainly formed by proteins without a predicted function, a cluster formed by six laccase genes (out of 12 in PC15) appeared to be located at the subterminal region of chromosome 6, six CBM13 genes (out of eight in PC15) appeared to be clustered at a subterminal region of chromosome 1, and a cluster that included seven genes encoding cerato-platanin-related proteins was found at the subterminal region of chromosome 3. The occurrence of clusters of genes involved in environmental adaptation (species-specific) at subterminal chromosomal locations has been reported in several organisms and could be the result of the exploitation of the high variability associated with these chromosomal regions (Pérez *et al.*, 2009; Ramírez *et al.*, 2011).

Comparative analysis of the P. ostreatus *bioinfosecretome with the proteomes of 54 basidiomycetes*

The 554 proteins of the *P. ostreatus* PC15 bioinfosecretome were used as queries to search for similar proteins in other basidiomycete proteomes. The presence or absence of proteins similar to those of the *P. ostreatus* PC15
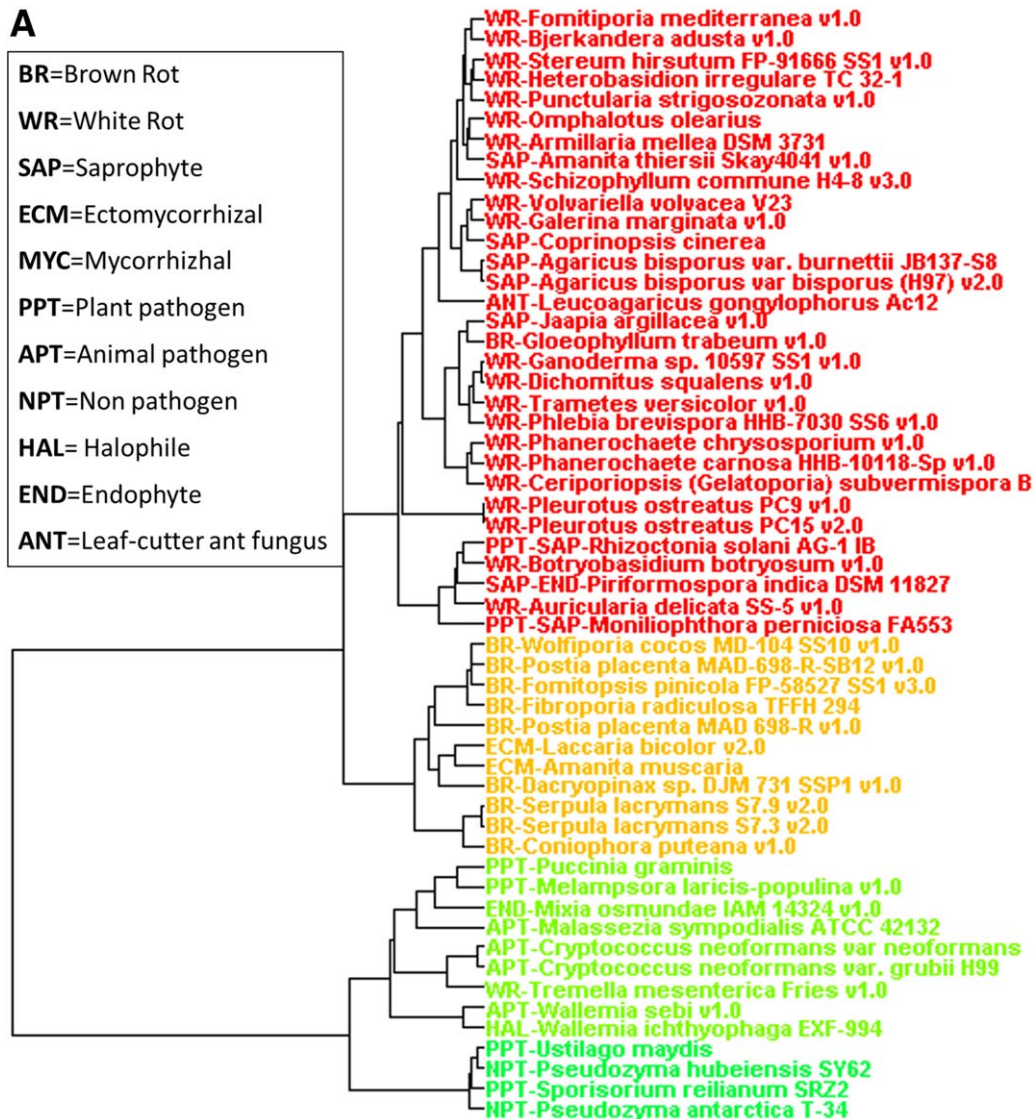
Fig. 7. A. Data from Supporting Information File 1, Table S10 arranged in a dendrogram (number of homologous proteins (e−20) to PC15 secreted proteins in basidiomycetes). Branches are coloured in four, starting from the left of the dendrogram.
B. Phylogeny of the fungal species analysed in this study. Branches are coloured following the pattern of colours in Fig. 7A.

bioinfosecretome was used as a characteristic to classify the corresponding fungal species in a dendrogram. For a comparison, we built up a phylogenetic tree based on the predicted proteomes of all the fungal species used in this work (using *Aspergilus nidulans* as outgroup). The comparison of both trees revealed that the studied basidiomycetes appeared to be grouped according to their lifestyle (Fig. 7, Supporting Information File 1 (Table S10) and Fig. S1 for the bootstrap data). Two major branches were observed in the dendrogram: the first branch could be subdivided into a group that included WR and SAP species and a second group of BR and ECM species, and the second branch included PPT, APT and endophytes (END) (Fig. 7). The cluster of BR and ECM supports previous

studies showing similarities in their organic matter degradation mechanisms (Rineau *et al.*, 2012) and a parallel contraction of the lignin-degrading peroxidase gene family (Floudas *et al.*, 2012; Kohler *et al.*, 2015). The groups formed in this analysis reflected the fungal lifestyle rather than their evolutionary relationships, supporting the hypothesis that the fungal secretome could be used to predict the lifestyle because it is targeted to the ecological niche of the organism (Martinez *et al.*, 2009; Lowe and Howlett, 2012; Alfaro *et al.*, 2014; Krijger *et al.*, 2014). Two species of the genus *Amanita* [*A. thiersii* (Hess *et al.*, 2014) *and A. muscaria* (Kohler *et al.*, 2015)] with different lifestyles were included in this analysis; they appeared to be grouped according to their lifestyle rather than their
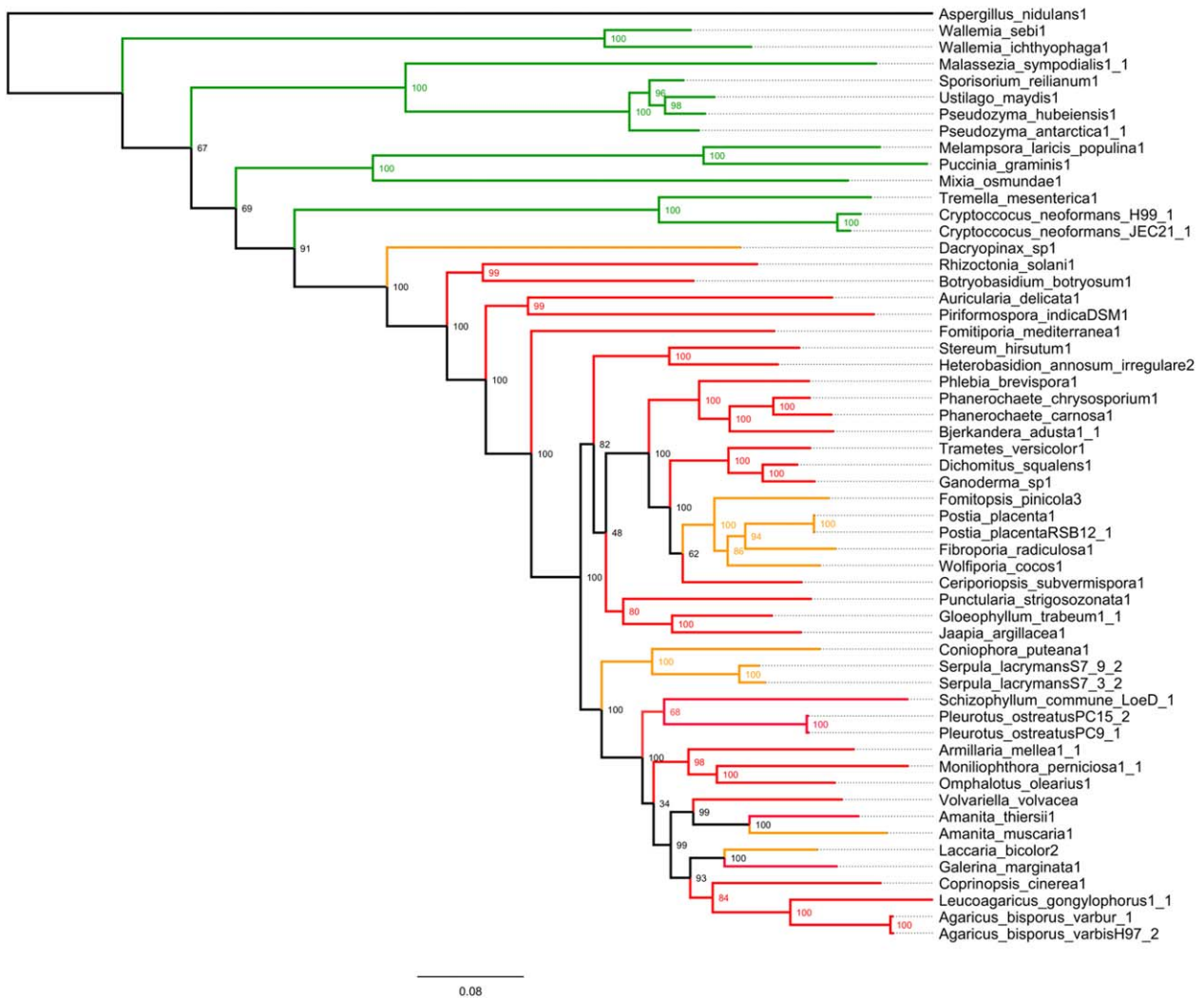
B

Aspergillus_nidulans1
Wallemia_sebi1
Wallemia_ichthyophaga1
Malassezia_sympodialis1_1
Sporisorium_reilianum1
Ustilago_maydis1
Pseudozyma_hubeiensis1
Pseudozyma_antarctica1_1
Melampsora_laricis_populina1
Puccinia_graminis1
Mixia_osmundae1
Tremella_mesenterica1
Cryptoccocus_neoformans_H99_1
Cryptoccocus_neoformans_JEC21_1
Dacryopinax_sp1
Rhizoctonia_solani1
Botryobasidium_botryosum1
Auricularia_delicata1
Piriformospora_indicaDSM1
Fomitiporia_mediterranea1
Stereum_hirsutum1
Heterobasidion_annosum_irregulare2
Phlebia_brevispora1
Phanerochaete_chrysosporium1
Phanerochaete_carnosa1
Bjerkandera_adusta1_1
Trametes_versicolor1
Dichomitus_squalens1
Ganoderma_sp1
Fomitopsis_pinicola3
Postia_placenta1
Postia_placentaRSB12_1
Fibroporia_radiculosa1
Wolfiporia_cocos1
Ceriporiopsis_subvermispora1
Punctularia_strigosozonata1
Gloeophyllum_trabeum1_1
Jaapia_argillacea1
Coniophora_puteana1
Serpula_lacrymansS7_9_2
Serpula_lacrymansS7_3_2
Schizophyllum_commune_LoeD_1
Pleurotus_ostreatusPC15_2
Pleurotus_ostreatusPC9_1
Armillaria_mellea1_1
Moniliophthora_perniciosa1_1
Omphalotus_olearius1
Volvariella_volvacea
Amanita_thiersii1
Amanita_muscaria1
Laccaria_bicolor2
Galerina_marginata1
Coprinopsis_cinerea1
Leucoagaricus_gongylophorus1_1
Agaricus_bisporus_varbur_1
Agaricus_bisporus_varbisH97_2

0.08

Fig. 7. Continued.

phylogenetic classification. Moreover, this analysis revealed that the secretomes of *J. argillacea* (classified as an uncertain WR) and *Gloeophyllum trabeum* (classified as a BR fungus) were highly similar, supporting the finding that both fungi belong to sister groups (Riley *et al.*, 2014). Finally, the analysis placed *B. botryosum*, which is a species with enzymes typical of both WR and BR (Riley *et al.*, 2014), into a small sister group of the WR fungi. Interestingly, this group included two PPT (*Rhizoctonia solani* and *M. perniciosa*) and a root endophyte and growth promoting fungus (*Piriformospora indica*) that exhibited a dual lifestyle, with a switch to a SAP lifestyle during their infection cycles (Mondego *et al.*, 2008; Zuccaro *et al.*, 2011; Wibberg *et al.*, 2013). Moreover, a mycorrhizal association of some species of the genus *Botryobasidium* with the orchid genus *Apostasia* has been suggested (Yukawa *et al.*,

2009). The data presented here support a secretome profile for this fungus that is more closely related to other plant-interacting fungi than to pure WR or BR.

It is worth emphasizing that 209 predicted secreted proteins of *P. ostreatus* with unknown enzymatic function were included in this comparative analysis with other species of basidiomycetes. These proteins have been ignored in previous studies of lifestyle-associated proteins. Many of these proteins were present in nearly all the basidiomycetes compared in this study, and other proteins were clearly less represented in some lifestyles compared with others. Sorting fungi according to their lifestyles is not an easy issue, especially because many fungal species can switch between several lifestyles over the course of their life cycle. Our results suggest that the study of the secretome composition would provide a certain insight and add

valuable data that could enable us to infer the roles of proteins with unknown functions.

Conclusions

The quality of a fungal bioinfosecretome will depend on the quality of the genome annotation. Thus, these in silico-predicted secretomes will evolve along with new versions of genome assemblies and annotations. The accuracy of bioinfosecretomes can be improved, but *in vivo* analyses will always be necessary to determine the pool of enzymes that fungi secrete to adapt to their ecological niches. Because mass spectrometry and transcriptomic analyses cannot identify every single enzyme involved in this process, a combination of *in silico* and *in vivo* techniques is still the best approach to analyse fungal secretomes. In all cases, the transcriptomic analyses of the data associated with bioinfosecretomes provides important information about the actual use of the available gene repertoire. This transcriptome analysis revealed that the secretome's transcriptome was more dependent on the strain genotype than on the environmental conditions. This finding indicates that the transcriptomes of a strain cultured under different conditions are more similar than the transcriptomes of other strains cultured under similar conditions.

Classical genome analyses emphasize the expansion of gene families. However, gene family expansion does not correlate with an increase in the number of actively expressed genes. We suggest that gene family expansions are associated with an increase in potential rather than with an increase in activities in a given environmental condition.

Among the *P. ostreatus* PC15 predicted secreted proteins that were conserved in other basidiomycetes, more than 20% have an unknown enzymatic function. Transcriptome analysis noted the importance of these proteins, underlining the need for further studies to decipher their role in fungal biology. For example, we used domain structure prediction to predict the function of two new protein families: a putative xylanase and a putative MAC perforin domain protein.

The functional categories found in proteins predicted to be secreted showed that domains associated with lignocellulose degradation were overrepresented compared to the whole proteome. This finding reflects the matching between the secretome and the lifestyle. Indeed, a blind grouping of organisms using their secretomes as distinctive characteristics produces a dendrogram in which the fungi are grouped by their lifestyle rather than by their phylogeny. Taken together, these data emphasize the dynamic, environmentally dependent, and ecologically influenced characteristics of fungal secretomes (Martinez *et al.*, 2009; Lowe and Howlett, 2012; Alfaro *et al.*, 2014; Krijger *et al.*, 2014).

Experimental procedures

Fungal strains and growth conditions

Two *P. ostreatus* monokaryotic strains were used in this work. PC9 and PC15 were produced by dedikaryotization of the original dikaryotic strain (N001). Details of their genetics and molecular biology have been published elsewhere (Larraya *et al.*, 1999a,b,). Their two genomes were sequenced by the Joint Genome Institute and are accessible at http://genome.jgi.doe.gov/PleosPC9_1 and http://genome.jgi.doe.gov/PleosPC15_2. Both strains were maintained in Petri dishes containing solid SMY medium (10 g/l sucrose, 10 g/l malt extract, 4 g/l yeast extract, and 15 g/l agar) at 24°C in the dark and subcultured every 8 days. The liquid cultures were performed using SMY (same composition as above but without agar) in the dark at 24°C under either static (Sta) or shaking (Shk, 200 rpm) conditions. The samples for the transcriptome experiments were harvested when the cultures were in the exponential growth phase (Parenti *et al.*, 2013), although the two growth conditions differed due to the inherent effects of the static (growth as a floating cake) or shaking (growth as pellet) conditions.

RNA-seq data analysis

Total RNA was purified from the corresponding cultures using a fungal E.Z.N.A. RNA kit (Omega Bio-Tek, Norcross, GA, USA) following the manufacturer's indications. Four samples were prepared corresponding to each of the two strains in the two culture conditions (static and shaking): PC9-Sta, PC9-Shk, PC15-Sta and PC15-Shk. The sequencing experiments were performed using the SOLiD platform at Sistemas Genomicos (Valencia, Spain). The RNA-seq reads were mapped to the *P. ostreatus* PC15 (assembled in 12 scaffolds) and PC9 (assembled in 572 scaffolds) genome sequences using TopHat (Trapnell *et al.*, 2009). HTseq-count (Anders *et al.*, 2014) was used to determine the number of reads mapped to every predicted gene model. Samtools (Li *et al.*, 2009) and custom Python scripts were used to handle the data and calculate the RPKM(Reads Per Kilobase of exon per Million reads mapped) values, which were used to evaluate the transcriptional levels of the secreted proteins.

Prediction of secreted proteins in P. ostreatus *PC9 and PC15*

The complete sets of proteins (best filtered models) predicted in PC9 and PC15 were obtained from MycoCosm (http://jgi.doe.gov/fungi) (Grigoriev *et al.*, 2012; 2014), and the prediction of the secreted proteins were performed using the web analysis tool SECRETOOL (http://genomics.cicbiogune.es/SECRETOOL/Secretool.php) (Cortázar *et al.*, 2014). This tool starts processing the data with TargetP (Emanuelsson *et al.*, 2007), SignalP (Bendtsen *et al.*, 2004) and PredGPI (Pierleoni *et al.*, 2008). Then, the proteins predicted by these three methods are merged into a common list that is evaluated for transmembrane domains (TMD) using the TMHMM tool (Krogh *et al.*, 2001). The candidate proteins with 0 or 1 TMDs are kept as input for WoLF PSORT (Horton *et al.*, 2007) (thereby eliminating them from our secretome analysis as

transmembrane proteins) to retain the sequences labelled as extracellular. When proteins were identified as allelic pairs in PC9 and PC15 but only one of the alleles was predicted to be secreted, we used Python scripts to extract the sequence 210 bp upstream from the 5′ region of the gene models used as queries to search for similar regions in the genome of the monokaryon with alleles not predicted to be secreted using the BLASTN standalone program (Camacho *et al.*, 2009). Alleles were defined as Blast Reciprocal Best Hits (RBH) between the PC15 and PC9 gene models (Santoyo *et al.*, unpublished).

Hypothetical functions for the predicted secreted proteins were assigned using the utilities available at the JGI webpage (Grigoriev *et al.*, 2014), CAZy database (Cantarel *et al.*, 2009), Pfam database (Finn *et al.*, 2006), Gene Ontology and Interpro Domain database (Apweiler *et al.*, 2000).

If we could not assign a hypothetical function using this procedure (proteins classified as unknown), we used the MCL program to cluster proteins (I = 2) (Enright *et al.*, 2002) and the Phyre2 server (Kelley and Sternberg, 2009) to predict the structure and function of the proteins. BLASTP (cut-off $E < 10^{-10}$) in JGI MycoCosm (Grigoriev *et al.*, 2014) was used to identify similar proteins in other basidiomycetes.

The Pfam domain search was performed using the Inter-ProScan standalone program (Apweiler *et al.*, 2000; Bateman *et al.*, 2004; Jones *et al.*, 2014) and database version. Fisher's exact test was used to reveal enriched Pfam domains ($P < 0.05$) in the bioinfosecretome compared to the genome data of *P. ostreatus*.

The OmnimapFree program was used to visualize the position of individual and cluster genes on the *P. ostreatus* PC15 chromosomes (Antoniw *et al.*, 2011). We used the REEF ('Regionally Enriched Features') program (Coppe *et al.*, 2006) to identify genomic regions enriched in specific features using a statistical test based on the hypergeometric distribution with a sliding window approach and adopting the false discovery rate to control multiplicity. A window width of 100 kb, a shift of 20 kb and an FDR-corrected *P*-value of < 0.05 were used. The minimum number of genes to form a cluster was 3.

Comparative analysis of the *P. ostreatus PC15* bioinfosecretome with the basidiomycetes' proteomes

For the comparison of the *P. ostreatus* PC15 bioinfosecretome with 54 basidiomycetes proteomes, the protein filtered models were obtained from the JGI webpage (Grigoriev *et al.*, 2014). BLASTP standalone (cut-off $E < 10^{-20}$) (Camacho *et al.*, 2009) was used to identify proteins similar to the PC15 secreted proteins. Python scripts were developed to arrange the results of the BLASTP search into a matrix of the presence or absence of similar proteins in the basidiomycete proteomes. The R program (R Core Team, 2013) with the function hclust was used to plot the dendrogram.

Phylogenetic analyses

The predicted proteomes of all species were downloaded from Mycocosm database (http://genome.jgi.doe.gov/programs/fungi/index.jsf) and proteins were clustered with MCL (Enright *et al.*, 2002) using an inflation value of 2. Clusters containing single copy genes of each genome were retrieved (allowing two missing taxa) and proteins were aligned with MAFFT (Katoh *et al.*, 2002). The alignments were concatenated after discarding poorly aligned positions with Gblocks (Castresana, 2000). Phylogeny was constructed using RaxML (Stamatakis, 2014) under PROTGAMMAWAGF substitution model and 100 rapid bootstraps.

References

Alfaro, M., Oguiza, J.A., Ramírez, L., and Pisabarro, A.G. (2014) Comparative analysis of secretomes in basidiomycete fungi. *J. Proteomics* **102:** 28–43.

Anders, S., Pyl, P.T., and Huber, W. (2014) HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31:** 166–169.

Antoniw, J., Beacham, A.M., Baldwin, T.K., Urban, M., Rudd, J.J., and Hammond-Kosack, K.E. (2011) OmniMapFree: a unified tool to visualise and explore sequenced genomes. *BMC Bioinformatics* **12:** 447.

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., *et al.* (2000) InterPro–an integrated documentation resource for protein families, domains and functional sites. *Bioinformatics* **16:** 1145–1150.

Baccelli, I., Luti, S., Bernardi, R., Scala, A., and Pazzagli, L. (2014) Cerato-platanin shows expansin-like activity on cellulosic materials. *Appl Microbiol Biotechnol* **98:** 175–184.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., et al. (2004) The Pfam protein families database. *Nucleic Acids Res* **32:** D138–141.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340:** 783–795.

Bernheimer, A.W. and Avigad, L.S. (1979) A cytolytic protein from the edible mushroom, Pleurotus ostreatus. *Biochim Biophys Acta* **585:** 451–461.

Caccia, D., Dugo, M., Callari, M., and Bongarzone, I. (2013) Bioinformatics tools for secretome analysis. *Biochim Biophys Acta* **1834:** 2442–2453.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10:** 421.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37:** D233–238.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17:** 540–552.

Castanera, R., Pérez, G., Omarini, A., Alfaro, M., Pisabarro, A.G., Faraco, V., *et al*. (2012) Transcriptional and enzymatic profiling of *Pleurotus ostreatus* laccase genes in submerged and solid-state fermentation cultures. *Appl Environ Microbiol* **78:** 4037–4045.

Castanera, R., Omarini, A., Santoyo, F., Pérez, G., Pisabarro, A.G., and Ramírez, L. (2013) Non-additive transcriptional profiles underlie dikaryotic superiority in *Pleurotus ostreatus* laccase activity. *PLoS One* **8:** e73282.

Conesa, A., Punt, P.J., van Luijk, N., and van den Hondel, C.A. (2001) The secretion pathway in filamentous fungi: a biotechnological view. *Fungal Genet Biol* **33:** 155–171.

Coppe, A., Danieli, G.A., and Bortoluzzi, S. (2006) REEF: searching REgionally Enriched Features in genomes. *BMC Bioinformatics* **7:** 453.

Cortázar, A.R., Aransay, A.M., Alfaro, M., Oguiza, J.A., and Lavín, J.L. (2014) SECRETOOL: integrated secretome analysis tool for fungi. *Amino Acids* **46:** 471–473.

DeBoy, R.T., Mongodin, E.F., Fouts, D.E., Tailford, L.E., Khouri, H., Emerson, J.B., et al. (2008) Insights into plant cell wall degradation from the genome sequence of the soil bacterium *Cellvibrio japonicus*. *J. Bacteriol* **190:** 5455–5463.

Eastwood, D.C., Floudas, D., Binder, M., Majcherczyk, A., Schneider, P., Aerts, A., *et al*. (2011) The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science* **333:** 762–765.

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2:** 953–971.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30:** 1575–1584.

Finn, R.D., Mistry, J., Schuster-Böckler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., *et al*. (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* **34:** D247–251.

Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R.A., Henrissat, B., *et al*. (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336:** 1715–1719.

Fonzi, W.A. (2009) The protein secretory pathway of *Candida albicans*. *Mycoses* **52:** 291–303.

Gaderer, R., Bonazza, K., and Seidl-Seiboth, V. (2014) Cerato-platanins: a fungal protein family with intriguing properties and application potential. *Appl Microbiol Biotechnol* **98:** 4795–4803.

Grigoriev, I.V, Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., *et al*. (2012) The genome portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Res* **40:** D26–32.

Grigoriev, I.V, Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., *et al*. (2014) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res* **42:** D699–704.

Hess, J., Skrede, I., Wolfe, B., LaButti, K., Ohm, R.A., Grigoriev, I. V, and Pringle, A. (2014) Transposable element dynamics among asymbiotic and ectomycorrhizal *Amanita* fungi. *Genome Biol Evol* **6:** 1564–1578.

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res* **35:** W585–587.

Jain, P., Podila, G.K., and Davis, M.R. (2008) Comparative analysis of non-classically secreted proteins in *Botrytis cinerea* and symbiotic fungus *Laccaria bicolor*. *BMC Bioinformatics* **9** (Suppl 10):O3.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., *et al*. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30:** 1236–1240.

Kämper, J., Kahmann, R., Bölker, M., Ma, L.J., Brefort, T., Saville, B.J., *et al*. (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444:** 97–101.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30:** 3059–3066.

Kelley, L.A. and Sternberg, M.J.E. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4:** 363–371.

Kohler, A., Kuo, A., Nagy, L.G., Morin, E., Barry, K.W., Buscot, F., *et al*. (2015) Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat Genet* **47:** 410–415.

Krijger, J.-J., Thon, M.R., Deising, H.B., and Wirsel, S.G. (2014) Compositions of fungal secretomes indicate a greater impact of phylogenetic history than lifestyle adaptation. *BMC Genomics* **15:** 722.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305:** 567–580.

Larraya, L., Peñas, M.M., Pérez, G., Santos, C., Ritter, E., Pisabarro, A.G., and Ramírez, L. (1999a) Identification of incompatibility alleles and characterisation of molecular markers genetically linked to the A incompatibility locus in the white rot fungus *Pleurotus ostreatus*. *Curr Genet* **34:** 486–493.

Larraya, L.M., Perez, G., Penas, M.M., Baars, J.J.P., Mikosch, T.S.P., Pisabarro, A.G., and Ramirez, L. (1999b) Molecular Karyotype of the White Rot Fungus *Pleurotus ostreatus*. *Appl Environ Microbiol* **65:** 3413–3417.

Larraya, L.M., Pérez, G., Ritter, E., Pisabarro, A.G., and Ramírez, L. (2000) Genetic linkage map of the edible

basidiomycete *Pleurotus ostreatus*. *Appl Environ Microbiol* **66:** 5290–5300.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Loftus, B.J., Fung, E., Roncaglia, P., Rowley, D., Amedeo, P., Bruno, D., *et al.* (2005) The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* **307:** 1321–1324.

Lowe, R.G.T. and Howlett, B.J. (2012) Indifferent, affectionate, or deceitful: lifestyles and secretomes of fungi. *PLoS Pathog* **8:** e1002515.

Lukoyanova, N., Kondos, S.C., Farabella, I., Law, R.H.P., Reboul, C.F., Caradoc-Davies, T.T., *et al.* (2015) Conformational changes during pore formation by the perforin-related protein pleurotolysin. *PLoS Biol* **13:** e1002049.

Martinez, D., Larrondo, L.F., Putnam, N., Gelpke, M.D., Huang, K., Chapman, J., *et al.* (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat Biotechnol* **22:** 695–700.

Martinez, D., Challacombe, J., Morgenstern, I., Hibbett, D., Schmoll, M., Kubicek, C.P., *et al.* (2009) Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proc Natl Acad Sci USA* **106:** 1954–1959.

Mondego, J.M.C., Carazzolle, M.F., Costa, G.G.L., Formighieri, E.F., Parizzi, L.P., Rincones, J., *et al.* (2008) A genome survey of *Moniliophthora perniciosa* gives new insights into Witches' Broom Disease of cacao. *BMC Genomics* **9:** 548.

Mueller, O., Kahmann, R., Aguilar, G., Trejo-Aguilar, B., Wu, A., and de Vries, R.P. (2008) The secretome of the maize pathogen *Ustilago maydis*. *Fungal Genet Biol* **45 (Suppl. 1):** S63–70.

Nickel, W. and Seedorf, M. (2008) Unconventional mechanisms of protein transport to the cell surface of eukaryotic cells. *Annu Rev Cell Dev Biol* **24:** 287–308.

Nombela, C., Gil, C., and Chaffin, W.L. (2006) Non-conventional protein secretion in yeast. *Trends Microbiol* **14:** 15–21.

De Oliveira, A.L., Gallo, M., Pazzagli, L., Benedetti, C.E., Cappugi, G., Scala, A., *et al.* (2011) The structure of the elicitor Cerato-platanin (CP), the first member of the CP fungal protein family, reveals a double ψβ-barrel fold and carbohydrate binding. *J Biol Chem* **286:** 17560–17568.

Parenti, A., Muguerza, E., Iroz, A.R., Omarini, A., Conde, E., Alfaro, M., *et al.* (2013) Induction of laccase activity in the white rot fungus *Pleurotus ostreatus* using water polluted with wheat straw extracts. *Bioresour Technol* **133:** 142–149.

Peñas, M.M., Rust, B., Larraya, L.M., Ramírez, L., and Pisabarro, A.G. (2002) Differentially regulated, vegetative-mycelium-specific hydrophobins of the edible basidiomycete *Pleurotus ostreatus*. *Appl Environ Microbiol* **68:** 3891–3898.

Pérez, G., Pangilinan, J., Pisabarro, A.G., and Ramírez, L. (2009) Telomere organization in the ligninolytic basidiomycete *Pleurotus ostreatus*. *Appl Environ Microbiol* **75:** 1427–1436.

Pierleoni, A., Martelli, P.L., and Casadio, R. (2008) PredGPI: a GPI-anchor predictor. *BMC Bioinformatics* **9:** 392.

Plett, J.M. and Martin, F. (2011) Blurred boundaries: lifestyle lessons from ectomycorrhizal fungal genomes. *Trends Genet* **27:** 14–22.

R Core Team, (R Foundation for Statistical Computing) (2013) R: a language and environment for statistical computing, Vienna, Austria.

Ramírez, L., Pérez, G., Castanera, R., Santoyo, F., and Pisabarro, A.G. (2011) Basidiomycetes telomeres- A bioinformatics approach. *In Bioinformatics - Trends and Methodologies*. Mahdavi, M.A., (ed). Rijeka, Croatia: InTech, pp. 393–424.

Riley, R., Salamov, A.A., Brown, D.W., Nagy, L.G., Floudas, D., Held, B.W., *et al.* (2014) Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proc Natl Acad Sci USA* **111:** 9923–9928.

Rineau, F., Roth, D., Shah, F., Smits, M., Johansson, T., Canbäck, B., *et al.* (2012) The ectomycorrhizal fungus *Paxillus involutus* converts organic matter in plant litter using a trimmed brown-rot mechanism involving Fenton chemistry. *Environ Microbiol* **14:** 1477–1487.

Rosado, C.J., Buckle, A.M., Law, R.H.P., Butcher, R.E., Kan, W.-T., Bird, C.H., *et al.* (2007) A common fold mediates vertebrate defense and bacterial attack. *Science* **317:** 1548–1551.

Sbrana, F., Bongini, L., Cappugi, G., Fanelli, D., Guarino, A., Pazzagli, L., *et al.* (2007) Atomic force microscopy images suggest aggregation mechanism in cerato-platanin. *Eur Biophys J* **36:** 727–732.

Shoji, J., Arioka, M., and Kitamoto, K. (2008) Dissecting cellular components of the secretory pathway in filamentous fungi: insights into their application for protein production. *Biotechnol Lett* **30:** 7–14.

Stajich, J.E., Wilke, S.K., Ahrén, D., Au, C.H., Birren, B.W., Borodovsky, M., *et al.* (2010) Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea (Coprinus cinereus)*. *Proc Natl Acad Sci USA* **107:** 11889–11894.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30:** 1312–1313.

Sun, J., Tian, C., Diamond, S., and Glass, N.L. (2012) Deciphering Transcriptional Regulatory Mechanisms Associated with Hemicellulose Degradation in *Neurospora crassa*. *Eukaryot Cell* **11:** 482–493.

Tjalsma, H., Bolhuis, A., Jongbloed, J.D., Bron, S., and van Dijl, J.M. (2000) Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol Mol Biol Rev* **64:** 515–547.

Tomita, T., Noguchi, K., Mimuro, H., Ukaji, F., Ito, K., Sugawara-Tomita, N., and Hashimoto, Y. (2004) Pleurotolysin, a novel sphingomyelin-specific two-component cytolysin from the edible mushroom *Pleurotus ostreatus*, assembles into a transmembrane pore complex. *J Biol Chem* **279:** 26975–26982.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25:** 1105–1111.

Vincent, D., Kohler, A., Claverol, S., Solier, E., Joets, J., Gibon, J., *et al.* (2012) Secretome of the free-living

mycelium from the ectomycorrhizal basidiomycete *Laccaria bicolor*. *J Proteome Res* **11:** 157–171.

Wibberg, D., Jelonek, L., Rupp, O., Hennig, M., Eikmeyer, F., Goesmann, A., *et al*. (2013) Establishment and interpretation of the genome sequence of the phytopathogenic fungus *Rhizoctonia solani* AG1-IB isolate 7/3/14. *J Biotechnol* **167:** 142–155.

Wymelenberg, A. V, Sabat, G., Martinez, D., Rajangam, A.S., Teeri, T.T., Gaskell, J., *et al*. (2005) The *Phanerochaete chrysosporium* secretome: database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium. *J Biotechnol* **118:** 17–34.

Vanden Wymelenberg, A., Minges, P., Sabat, G., Martinez, D., Aerts, A., Salamov, A., *et al*. (2006) Computational analysis of the *Phanerochaete chrysosporium* v2.0 genome database and mass spectrometry identification of peptides in ligninolytic cultures reveal complex mixtures of secreted proteins. *Fungal Genet Biol* **43:** 343–356.

Yukawa, T., Ogura-Tsujita, Y., Shefferson, R.P., and Yokoyama, J. (2009) Mycorrhizal diversity in *Apostasia* (Orchidaceae) indicates the origin and evolution of orchid mycorrhiza. *Am J Bot* **96:** 1997−2009.

Zuccaro, A., Lahrmann, U., Güldener, U., Langen, G., Pfiffi, S., Biedenkopf, D., *et al*. (2011) Endophytic life strategies decoded by genome and transcriptome analyses of the mutualistic root symbiont *Piriformospora indica*. *PLoS Pathog* **7:** e1002290.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Figure 6 Dendrogram with bootstrap values ($n = 1000$).

Fig. S2. Comparison between the transcriptional activity (RPKM) of the highest expressed gene versus others genes with the same protein function.

Supp. Inf. File 1, Table S1. Protein models predicted to be secreted arranged by allele and function.

Supp. Inf. File 1, Table S2. Unknown homologs.

Supp. Inf. File 1, Table S3. Unknown homologs n° basidios.

Supp. Inf. File 1, Table S4. Unknown clusters.

Supp. Inf. File 1, Table S5. MAC perforin similar proteins.

Supp. Inf. File 1, Table S6. Xylan esterases similar proteins.

Supp. Inf. File 1, Table S7. Number and percentage of Pfam domains in the predicted to be secreted filtered models and the genome filtered models. Non-correlated secretome-genome by fisher exact test.

Supp. Inf. File 1, Table S8. N° of predicted to be secreted proteins Chromosome lenght (Mbp). N° of predicted to be secreted proteins/Mbp Chromosome.

Supp. Inf. File 1, Table S9. Chromosome position clusters.

Supp. Inf. File 1, Table S10. Basidiomycete fungi ordered by number of proteins homologous (e−20) to PC15 proteins predicted to be secreted.