19

Basidiomycetes Telomeres – A Bioinformatics Approach

Ramírez Lucía, Pérez Gúmer, Castanera Raúl,
Santoyo Francisco and Pisabarro Antonio G.
Genetics and Microbiology Research Group, Public University of Navarre, Pamplona, Spain

1. Introduction

1.1 The telomere: A complex nucleoprotein complex with a broad range of functions

The telomeres (from the Greek *télos* far and *-meros* part) are the genetic structures found at the physical ends of linear chromosomes. They are nucleoprotein complexes composed of DNA repeats and a myriad of telomere and non-telomere associated proteins aimed to protect the ends of eukaryotic chromosomes from being recognized as double strand breaks, and to avoid chromosome end degradation by nucleases and non-canonical chromosome end fusions. Thus, telomeres are essential for chromosome integrity (Hande, 2004; Paeschke et al., 2010; Zakian, 1995). The fascinating story about telomere biology comes from the pioneering work of Elizabeth H. Blackburn who discovered that *Tetrahymena* telomeres consisted of a short DNA sequence motif that was repeated several times at the chromosomal end (Blackburn & Gall, 1978). This pattern is conserved in lower eukaryotes and in mammalian cells (Greider, 1998). Notable exceptions are *Drosophila* and some other dipterans, which instead possess tandem arrays of retrotransposons at their chromosome ends (Abad et al., 2004).

Telomere DNA consists of tandem arrays of short repeated sequences forming a cap. Telomere length is species-specific and small cell type variations were observed. For reviews, see Fisher & Zakian, (2005) and Sanchez-Alonso & Guzman (2008). The basic telomere DNA repeat unit is the hexamer TTAGGG in which the strand running 5´→ 3´ outwards the centromere is usually guanine-rich. This G-rich strand protrudes its complementary end and bends on itself to form a telomere DNA loop (T-loop) (Griffith et al., 1999) which protects the structure from being recognized as a double-stranded break by sequestering the 3′-overhang into a high order DNA structure. The G-rich strand also serves as an anchor for a telomere-dedicated reverse transcriptase, called telomerase, that compensates for the inability of DNA polymerases to replicate the 5´ ends of linear chromosomes (Blackburn & Gall, 1978). The telomerase binds the G-rich strand by complementary pairing of the protruding DNA sequence to the telomerase RNA subunit and, as a result, telomerase elongates the overhang by adding telomere sequence repeats (Masutomi et al., 2003; Morin, 1989; Zhao et al., 2009). The T loop structure is maintained by a complex of telomere and non-telomere proteins called shelterins which repress the DNA repair machinery at telomeres, and regulate telomere length (for review see de Lange, 2005; Palm & de Lange, 2008; Rhodes et al., 2002; Vega et al., 2003; Zhao et al., 2009). The shelterin

complex is formed by a core of six proteins including the Myb-type homeodomain TRF proteins in mammals, Rap1 in *Saccharomyces cerevisiae* and Taz1 in *Schizosaccharomyces pombe* which bind the duplex form of the telomere repeats, the OB-fold containing protein POT1 in mammals and *S. pombe* and Cdc13 in *S. cerevisiae* which bind the single-stranded telomere 3'overhang and by other proteins associated via protein–protein interactions with them (Rhodes et al., 2002; Vega et al., 2003; Zhao et al., 2009). This shelterin complex is evolutionary conserved although some differences between species appear in protein numbers and in its higher order structure (Linger & Price, 2009). The G-rich single stranded telomere tail is also able to form a secondary DNA order structure resulting from intra and intermolecular G-quadruplex (Fry, 2007; Maizels, 2006). G-quadruplexes are stacked associations of G-quartets, which are themselves planar assemblies of four Hoogsteen-bonded guanines, with the guanines derived from one or more nucleic acid strands (De Cian et al., 2008; Johnson et al., 2008). These structures have been observed in lower eukaryotes (Paeschke et al., 2008; Paeschke et al., 2005; Schaffitzel et al., 2001) and have the potential to regulate telomerase activity (Oganesian et al., 2006; Zahler et al., 1991).

Secondary DNA structures, G-quadruplex structures and T-loop may contribute to telomere function but pose an obstacle for semi-conservative and telomerase-mediated replication, a problem which should be solved to avoid telomere shorten. Telomeres become shortened during every cell cycle due to incomplete replication of the lagging strand (the so called ''end replication problem") resulting in cumulative telomere attrition during aging. In addition, a loss of telomere DNA occurs due to post-replicative degradation of the 5´ strand that generates long 3´ G-rich overhangs (Wellinger et al., 1996; Wellinger et al., 1993). In most species, the loss of telomere DNA is counteracted by the action of telomerase that carries its own RNA template coding for the telomere repeat sequence (Chan & Blackburn, 2004). The complementary C-rich strand is then synthesized by conventional RNA-primed DNA replication (Gilson & Geli, 2007; Verdun & Karlseder, 2007). Following replication, the telomeres created by the synthesis of the leading strand are either blunt-ended or left carrying a small 50 bp overhang whereas those created by the lagging-strand synthesis have a 3′ overhang with a length determined by the position of the outermost RNA primer (de Lange, 2009). This fact supports the importance of the telomerase activity for the genome integrity. When telomeres reach a critical minimal length they become uncapped. This leads to a permanent cell cycle arrest (termed cellular senescence) or to apoptosis, depending on the cellular context in which the uncapping occurs (Aubert & Lansdorp, 2008; Blasco, 2005). Extreme telomere shortening leads to chromosome instability, end-to-end fusions, and checkpoint-mediated cell cycle arrest and/or apoptosis (reviewed in Aubert & Lansdorp (2008) and in Shore & Bianchi (2009). The whole processes are related in mammals not only to aging, but also to several age associated diseases such as tumorigenesis, coronary artery disease, and heart failure (Donate & Blasco, 2011; Ogami et al., 2004; Sherr & McCormick, 2002; Starr et al., 2007). In addition to the role of telomerase in maintaining telomere length, it has been described that homologous recombination (HR) constitute an alternative method (ALT ''alternative lengthening of telomeres'') to maintain telomere DNA in telomerase-deficient cells with telomeres highly heterogeneous in length. This mechanism was described in *S. cerevisiae* and consists in two pathways depending on different recombination proteins that use different telomere sequences as substrates for recombination. Cancer and immortalized cells can utilize the ALT mechanism to maintain telomere length (Lundblad & Blackburn, 1993; Teng et al., 2000; Teng & Zakian, 1999).

Telomeres were considered as regions where the transcription of structural genes is repressed (Mondoux & Zakian, 2005), although it has been recently reported that telomere repeats and subtelomere regions can be transcribed (Azzalin et al., 2007; Luke & Lingner, 2009; Luke et al., 2008; Schoeftner & Blasco, 2008). The telomere transcribed region, called TERRA (telomere repeat-containing RNA), forms an integral component of telomere heterochromatin, and produces non-coding G-rich RNAs transcribed from the telomere C-rich strand in mammals and fungi (Azzalin et al., 2007; Luke et al., 2008; Sanchez-Alonso & Guzman, 2008; Schoeftner & Blasco, 2008). TERRA transcription occurs at most or all chromosome ends and it is regulated by RNA surveillance factors and in response to changes in telomere length. The accumulation of TERRA at telomeres can also interfere with telomere replication (Azzalin & Lingner, 2007; Luke & Lingner, 2009; Schoeftner & Blasco, 2009).
The particular sequence organization of telomere makes difficult its genetic mapping and sequencing due to they are cloning recalcitrant and underrepresented in mapping and final assembled genomes. Consequently, their cloning and characterization must be made by dedicated molecular and bioinformatics strategies (Perez et al., 2009; Sanchez-Alonso & Guzman, 2008).

1.2 Subtelomere chromosome regions

Human subtelomere chromosome regions contain complex and dynamic stretches of DNA which, together with their associated proteins, are essential for genome stability and proper chromosome replication (Riethman et al., 2005). Subtelomere DNA repeats are a complex region of variable size segmentally duplicated containing low copy DNA repetitive tracts adjacent to the telomere. These duplications could be found only at the subtelomere regions, although it is common to find them also at pericentromeric and interstitial chromosomal loci (Riethman et al., 2001).
In humans, subtelomere DNA regions are operationally defined as the terminal 500 Kbp of each euchromatic chromosome arm. These regions contain subtelomere repeats (Srpts), segmental duplications, satellite sequences, and internal $(TTAGGG)_n$-like sequences (Riethman et al., 2004). The organization of the subtelomere region is structurally conserved across eukaryotes (Anderson et al., 2008; Brown et al., 1990a; Brown et al., 1990b; Chan & Tye, 1983a; 1983b; Flint et al., 1997a; Flint et al., 1997b; Karpen & Spradling, 1992; Levis, 1993; Louis, 1995; Louis & Borts, 1995; Mefford & Trask, 2002; Pryde et al., 1997; Pryde & Louis, 1997; Walter et al., 1995; Wilkie et al., 1991). This region, susceptible to hypermethylation has been recently shown to have a central function in mammalian telomere-length homeostasis (Blasco, 2007). Subtelomere repeats are characterized by their high level of polymorphism among different chromosome ends and among individuals of the same species. This polymorphism, possibly indicative of a quick and dynamic sequence turnover, leads to a lack of relationship among subtelomere repeats across species. Nonhomologous or ectopic exchange between subtelomere regions of different chromosomes has been reported as a possible reason of polymorphisms in both yeast and humans (Linardopoulou et al., 2005; Louis & Haber, 1990; Mefford & Trask, 2002). In humans, there are more re-arrangements at the sub-telomere regions than in the rest of the genome. This is also true in some lower eukaryotes such as *Plasmodium falciparum* (Freitas-Junior et al., 2000), *Magnaporthe oryzae* (Rehmeyer et al., 2006) and *Neurospora crassa* (Wu et al., 2009), among others. In all these cases genes involved in niche adaptation (species-

specific genes) were found in the subtelomeric regions. It could be that the high evolutive potential of the subtelomeric regions were used by these organisms to create variability aimed to avoid the detection by the host. In fungi the genes more frequently found in subtelomere regions are transposons, telomere-linked RecQ helicases, clusters of secondary-metabolites, cytochrome oxidases, hydrolases, molecular transporters, and genes encoding secreted proteins (Perez et al., 2009). These genes could undergo transcriptional silencing (subtelomere silencing) due to its close proximity to telomeres.

1.3 Interstitial telomere repeats

The various chemical modifications occurring at the amino terminal end of the histones affect the structure of chromatin and help establishing the functional and structural domains known as euchromatin and heterochromatin. Euchromatin is an open form of chromatin that allows transcription factors access to and transcriptionally activate their target genes. It is largely occupied by housekeeping genes, condensed during metaphase and decondensed during interphase. Heterochromatin, on the contrary, differs from euchromatin in that it is condensed during interphase.

Heterochromatin has been often said to be "poor in genes" and mainly constituted by repetitive DNA sequences. Moreover, since it is highly condensed and inaccessible to transcription factors, heterochromatin is generally transcriptionally silent (Hernandez-Rivas et al., 2010). Heterochromatin appears as blocks spread over the chromosomes when they are stained with Giemsa dye. The molecular analysis of heterochromatic blocks reveals sequences similar to the telomere repeats that are called in this case interstitial telomere repeat sequences or ITRs. These sequences include those repeats located close to the centromere and those found at interstitial sites, i.e., between the centromere and the telomeres (Meyne et al., 1990; Slijepcevic et al., 1996). ITRs were described in plants, animals and humans (Bolzan & Bianchi, 2006; Uchida et al., 2002; Welchen & Gonzalez, 2005). At the chromosome level, ITRs can be detected either by using the Fluorescence *in situ* hybridization (FISH) technique with a DNA or a peptide nucleic acid (PNA) pan-telomere probe (i.e., a probe that identifies simultaneously all of the telomeres in a metaphase cell), or by the primed *in situ* labeling (PRINS) reaction using an oligonucleotide primer complementary to the telomere DNA repeated sequence (Bolzan & Bianchi, 2006).

The length and the locations of the heterochromatic blocks in chromosomes are variable (Azzalin et al., 2001; Faravelli et al., 1998; Weber et al., 1990) as well as their origin. However, the presence of ITRs in the heterochromatic blocks is interpreted as the result of tandem telomere–telomere fusions during evolution (Hastie & Allshire, 1989; Holmquist & Dancis, 1979; Meyne et al., 1990) or the insertion of telomere DNA within genome unstable sites (recombination hotspots) during the repair of double strand DNA breaks (DSB) (Azzalin et al., 2001). The presence of some relatively small ITRs flanked by unstable AT-rich DNA sequences could support this last hypothesis (Faravelli et al., 2002). On the other hand, telomere associations and fusions are common cytogenetic findings that have been implicated in the initiation of chromosome instability and tumorigenesis (Callen & Surralles, 2004; Murnane & Sabatier, 2004; Soler et al., 2005). Telomere fusions are the result of telomere dysfunction due to attrition of chromosome ends (Maser & DePinho, 2004). They are usually found in repair- and/or telomerase-deficient cells (Bailey et al., 1999; Blasco et al., 1997; Hande, 2004; Hande et al., 1999; Lo et al., 2002; Samper et al., 2000) with a variety of mutations affecting telomere function, including those occurring in proteins of the

shelterin complex (Bailey et al., 1999; Goytisolo et al., 2001). Mammal cells show a high frequency of telomere fusions (end-to-end) and chromosome instability (Bailey et al., 2004; Bailey et al., 1999; Espejel et al., 2002; Hsu et al., 2000; Smogorzewska et al., 2002; Takai et al., 2003; van Steensel et al., 1998). The FISH technique allows the identification of metacentric-submetacentric and acrocentric-telocentric chromosome telomere fusions also known as Robertsonian-like configurations (Al-Wahiby et al., 2005; Hande, 2004). The occurrence of telomere–telomere associations has been suggested to play a role in nuclear organization (Nagele et al., 2001). In fact, telomere associations were seen in metaphases of human cells with shorten telomeres suggesting that a minimal telomere length is required for a proper chromosome function during mitosis.

2. Fungal telomeres - A bioinformatics approach

The basic and conserved telomere unit sequence in most filamentous fungi is TTAGGG. This sequence has been described in *Aspergillus nidulans* (Bhattacharyya & Blackburn, 1997), *Beauveria bassiana* (Padmavathi et al., 2003), *Botrytis cinerea* (Levis et al., 1997), *Cladosporium fulvum* (Coleman et al., 1993), *Fusarium oxysporum* (Inglis et al., 2000), *Glomus intraradices* (Hijri et al., 2007), *Magnaporthe grisea* (Gao et al., 2002), *Metarrhizium anisopliae* (Inglis et al., 2005), *N. crassa* (Wu et al., 2009), *Pestalotiopsis microspora* (Long et al., 1998), *Pleurotus ostreatus* (Perez et al., 2009), *Pneumocystis carinii* (Keely et al., 2001), and *Ustilago maydis* (Sanchez-Alonso & Guzman, 2008). However, variations of this sequence can be found in other fungi such as *A. oryzae* that has dodecanucleotide telomere repeats (Kusumoto et al., 2003) Incomplete and imperfect telomere units have been reported in: *A. oryzae, Candida albicans, Kluyveromyces lactis, S. cerevisiae* and *S. pombe*.

Two DNA sequence domains can be found adjacent to the telomere repeats. One of them, distal, is placed next to the telomere and contains tandem repeat motifs. The other, proximal, is interstitial, contains less repeated sequences and ferries clusters of related genes (Pryde et al., 1997). In several fungi, it has been observed an increased number of proteins involved in interactions with the environment coded for by genes mapping close to the telomeres. These genes are called 'contingency genes', they are dispensable for survival and are highly variable in populations. The accumulation of these genes near the telomeres is a strategy that allows fungi to afford new environments. In fact, it has been observed in *S. cerevisiae* that the genes located near the telomeres display variation in gene amplification and/or expression depending on the growing niche of the yeast. Some of these genes belong to the PAU family, the largest gene family in *S. cerevisiae* (23 members), whose regulation depends on the environmental growing conditions (anaerobiosis) (Rachidi et al., 2000). Other telomere associated genes in *S. cerevisiae* are MAL and MEL that participate in maltose and melibiose fermentation, used in the baking and brewing industries (Gibson et al., 1997; Teunissen & Steensma, 1995), and FLO that encodes cell-wall glycoproteins which participate in the regulation of cellular adhesion (Gibson et al., 1997; Halme et al., 2004; Teunissen & Steensma, 1995). Similarly, the TLO family unique in *C. albicans* consists of 15 members present on every chromosome, 14 of which are located at chromosome ends. Genome comparisons between *C. albicans* and *Candida dublinienesis* showed that the principal disparity in gene content between both species resides in the lack of the TLO genes in this last one. CdTLO1 null strains show a major reduction in hyphal formation in response to serum that can be reversed by complementation with either of two *C. albicans* TLO genes (Jackson et al., 2009).

In human parasites it has been described several families of well characterized or putative virulence factors at chromosome ends. In *Candida glabrata* (De Las Penas et al., 2003) nearly 24 adhesin encoding genes were located at telomere regions, and, in *P. carinii* clusters of major surface antigen genes have been bioinformatically predicted at every chromosome end (Keely et al., 2005). A similar situation was reported in *P. falciparum* and *Trypanosoma brucei* where families of antigen surface proteins inducing immune responses are encoded in telomere regions as part of a pathogen's mechanism to amplify and diversify surface antigen genes to avoid host recognition, as it has been hypothesized (Barry et al., 2003).

To determine if plant pathogenic fungi use a similar mechanism to avoid host defenses, Farman's group analyzed and characterized the telomere organization in the rice blast pathogen *M. oryzae* (Rehmeyer et al., 2006) and in *N. crassa* (Wu et al., 2009). The molecular and bioinformatics approach allowed them to identify 14 chromosomes in both of them. In *M. oryzae,* the analysis of these sequences reveals the presence of a clearly defined distal subtelomere domain that contains a telomere-linked helicase (TLH) gene. No gene duplication near the chromosome termini is observed. Thus it is impossible to detect a proximal subtelomere domain. The sequenced *N. crassa* genome (Galagan et al., 2003) contains very little intact, duplicated DNA due to the repeat-induced point mutation (RIP) process (Selker, 1990). This situation made unlikely that *N. crassa* would possess intact subtelomere domains or terminal gene duplications. The search for tandem repeats at the ends of some chromosomes whose sequences extend out of the telomere repeats shows the lack of conserved subtelomere tandem repeats. A similar situation is observed when the sequences immediately adjacent to the TTAGGG repeats were compared between strains. Consistent with the absence of distinct subtelomere elements, *N. crassa* lacks the TLH genes that are present in the subtelomere regions of diverse fungi (Gao et al., 2002; Inglis et al., 2005; Louis & Haber, 1992; Mandell et al., 2005; Perez et al., 2009; Rehmeyer et al., 2006; Sanchez-Alonso & Guzman, 1998). As in other fungi, the terminal regions of *N. crassa* chromosomes ferry genes related to secondary metabolism such as a monooyygenase (CYP450), a FAD-binding domain containing protein, a second CYP450, an O-methyltransferase, a polyketide synthase, a major facilitator superfamily efflux pump, a putative transcription factor, and an oxidoreductase. Apart from the clusters of secondary metabolite genes, there were no genes overrepresented in the chromosome ends, except those predicted to code for enzymes related to plant cell-wall degradation activity.

In basidiomycetes, despite the genomes of an ever growing number of genera involved in lignin degradation, enzyme production, and bio pulping have been sequenced and annotated, the information on the characteristics of their telomere and sub-telomere regions is rather limited. As it was discussed above, this is due to the particular characteristic of telomeres that make them refractory to cloning and difficult to sequence and assemble in whole genome sequencing projects. A consequence of the difficulty in cloning telomeres is that, for most organisms, there is limited information on the organization of chromosome ends. Data about basidiomycete telomeres are available for *U. maydis* (Sanchez-Alonso & Guzman, 1998), *M. anisopliae* (Inglis et al., 2005) and *P. ostreatus* (Perez et al., 2009). In the three cases, the conserved telomere repetitive unit is TTAGGG, and the number of tandem repetitions varies among species: 37 times in U. *maydis*, from 18 to 26 in *M. anisopliae*, and from 25 to 150 in *P. ostreatus*. In all these cases, genes coding for RecQ helicases have been found adjacent to the telomere regions.

In *P. ostreatus*, a lignin degrader edible mushroom, the analysis of telomere organization was carried out with a combination of genetic, molecular, and bioinformatics tools. This

approach allowed our group to map 19 out of the 22 chromosome ends expected in its linkage map (Perez et al., 2009), as well as to study the telomere adjacent regions. Similar strategies have been described by different authors (Rehmeyer et al., 2006; Sanchez-Alonso & Guzman, 2008; Wu et al., 2009). The search for telomere regions was performed in whole genome sequence draft assemblies. These preliminary genome sequence versions appeared as incomplete, contained some genes truncated and were misassembled. The rationale for using them is that the genome final assembling strategies very often alter or eliminate the telomere and subtelomere repetitive sequences of the fully assembled genomes**.** We used the open-access Tandem Repeats Finder program (TRFp, http://tandem.bu.edu/trf/trf.html) to screen for repetitive telomere sequences in more than 6,200 contigs of the 4X coverage draft sequence assembly of *P. ostreatus* PC15 produced by the Joint Genome Institute (http://genome.jgi-psf.org/PleosPC15_1/PleosPC15_1.home.html). The TRFp locates and displays tandem repeats in a DNA sequence file submitted in FASTA format without need of specifying the repetitive pattern, its size or any other parameter. The TRFp output consists of two files: a repeat table and an alignment. The repeat table contains information about each repeat, its size, copy number, nucleotide content and location. Clicking on the location indices for this table's entries opens a second web browser that shows an alignment of the copies against a consensus pattern. TRFp is a very fast program that permits the analysis of up to 5 Mb sequence length. Repeats with pattern size in the range from one to 2000 bases could be detected (Benson, 1999). We identified, in each telomere sequence containing scaffold, the filtered gene models computer and manually annotated within 50 Kbp of the telomere repetitive sequence. This was done by manual inspection of each one of the predicted genes and using them as query in the non-redundant NCBI gene database using the BlastX program (Altschul et al., 1997). The BlastX results were considered significant if their expected value (e-value) was <e-20. The identified protein sequences were also used to query the online Pfam database using default parameters (Bateman et al., 2004). A similar approach has also been used by Wu et al. (2009) to search for subtelomere regions containing gene in *N. crassa.* This type of multiple analysis (genetic, molecular, and bioinformatics) allows the characterization of most of the *P. ostreatus* telomeres as well as several subtelomere regions that show high nucleotide similarity. The highly polymorphic subtelomere region of *P. ostreatus* chromosome six contains genes similar to those described in other eukaryotic organisms (RecQ helicases), apart from a species-specific laccase gene cluster (six out of 12 genes annotated in the genome (Perez et al., 2009).

In conclusion, the assemblage of telomere regions by bioinformatics strategies is a powerful tool to determine the arrangement of genomes in putative linkage groups in species with no genetic maps available, to establish synteny among different basidiomycete genomes, and to determine the presence of genes and gene clusters conserved in the subtelomere regions of different genomes.

2.1 Analysis of the basidiomycetes' telomere regions

In the following sections we will review the composition and structure of the telomere and subtelomere regions of the different basidiomycetes using the bioinformatics approach described above. We will use the genome sequence data that are publicly available at the Joint Genome Institute (DOE-JGI, http://www.jgi.doe.gov/). This institute has been developing an intensive sequencing effort on different fungi related with the biological lignocellulose degradation. Lignin is a complex recalcitrant macromolecule that hinders the access of enzymes to cellulose. The enzymatic removal of lignin will permit the access to this

large carbon reservoir for its use in different energy-related applications. There are two types of fungi according to their strategy for making cellulose accessible: white rot fungi degrade lignin, and brown rot fungi minimally modify the lignin and attack the cellulose using a different chemical approach (Lundell et al., 2010; Martinez et al., 2005; Ruiz-Duenas & Martinez, 2009). Among all the sequenced basidiomycetes, we will concentrate here on the white rot degraders *Ceriporiopsis subvermispora, Phanerochaete chrisosporium* and *P. ostreatus*; the brown rot *Postia placenta* and the tree pathogen *Heterobasidion annosum*.

2.1.1 Analysis of the telomere regions of *Ceriporiopsis subvermispora* B

C. subvermispora is a white rot basidiomycete that rapidly depolymerizes lignin with relatively little cellulose degradation when growing on wood (Martinez et al., 2005). The chromosome number of this species has not been conclusively determined. The JGI has sequenced the strain B. Two type of sequence data (un-assembled reads and assembled scaffolds unmasked) were screened for the presence of telomere sequences in the genome of *C. subvermispora*: 297,269 unassembled and 740 assembled scaffolds. In 207 unassembled scaffolds, between five and 23 tandem repeats of the telomere motif TTAGGG were found. In most cases (82 % of these unassembled scaffolds) 17 and 19 repeats of the motif were found, and the mean repeat number was 18.7. On the other hand, 187 scaffolds carried between seven and 22 repeats of the telomere complementary sequence CCCTAA. In this case, the modal repetition number varied between 17 and 19 (84% of the scaffolds) and the mean number was 19.2. Taken together these data suggest that the average number of telomere repeats in *C. subvermispora* is 19.0.

The unmasked analysis of 740 assembled scaffolds revealed that 42 of them contained telomere sequences: 22 harbored the TTAGGG sequence and 20 the complementary CCCTAA. The telomere region TTAGGG was placed at the bottom (3′ telomere) end of the chromosome in 21 out of the 22 scaffolds. An exception to this was observed in scaffold 3. The sequence was located at an interstitial position. The telomere region CCCTAA was placed at the upper (5′ telomere) end of the chromosome in 19 out of 20 scaffolds. As it was described above, this sequence was also placed at an interstitial location in scaffold 3. This suggest a missassembling scaffold 3 because both the direct $TTAGGG_{21}$ (TTAGGG, scaffold 3: 1205371-1205497) and the reverse $CCCTAA_{21}$ (CCCTAA, scaffold 3: 1206763-1205497) telomere sequences have been found flanking a 1265 bp gap 3 (Fig. 1).

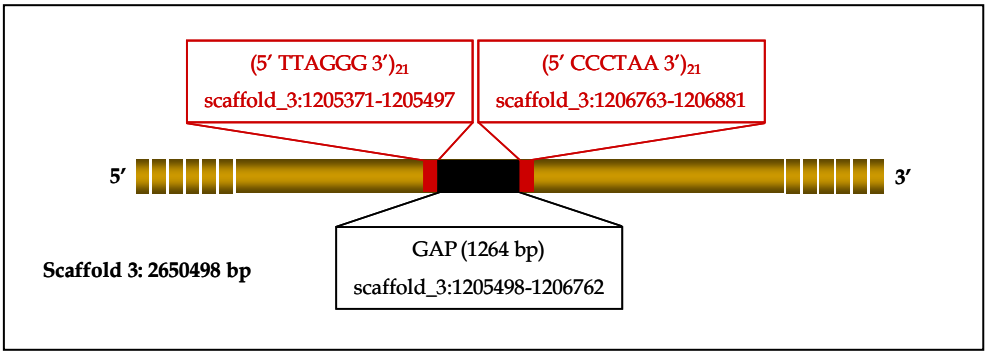


Fig. 1. Location of interstitial telomere sequences in scaffold 3 of the genome sequence of *C. subvermispora.*

The telomeres found in scaffolds 5 and 7 have a complex structure. The telomere of scaffold 5 contained 22 and 20 copies of the CCCTAA sequence separated by a gap of 193 bp. In the case of scaffold 7, 20 and 18 copies of this sequence appeared separated by a gap of 1680 bp (Fig. 2). We presume that these particular structures could reflect misassemblages of the telomere sequences in both scaffolds.

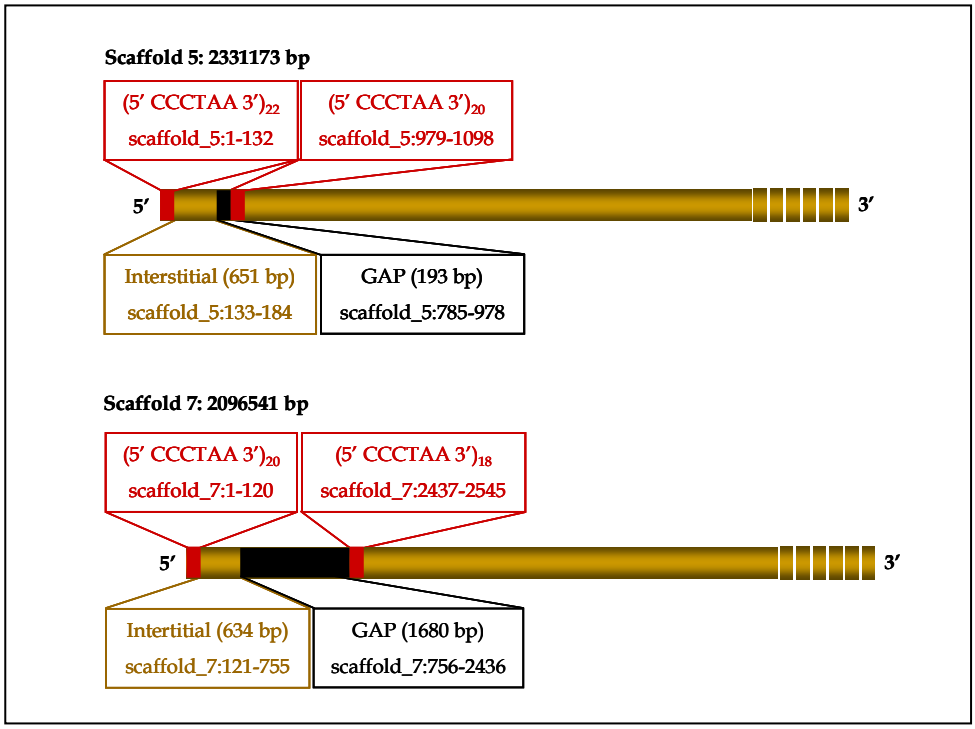


Fig. 2. Location of telomere regions in scaffolds 5 and 7 of the genome sequence of *C. subvermispora.*

In summary this analysis has allowed the identification of 42 telomere repeat containing regions. 22 out of them fit with direct telomere sequences (TTAGGG) present in 22 scaffolds and 22 reverse telomere regions (CCCTAA) present in 20 scaffolds. Taking into account that the *C. subvermispora* strain sequence was a dikaryon, these results suggest that the *Ceriporiopsis* genome would consist of 11 chromosomes. Because scaffold 7 showed telomere repeats at both ends, the sequence contained in this scaffold would be the only fully assembled chromosome of *C. subvermispora.*

2.1.2 Analysis of the telomere regions of *Phanerochaete chrysosporium* strain RP78

P. chrisosporium is a model white rot basidiomycete that has been extensively used because of it interest as lignin degrader (Kersten & Cullen, 2007; Tien, 1987). The draft genome of the homokaryotic strain of *P. chrysosporium* strain RP78 was assembled into 232 scaffolds and contains 35.1 Mbp of non-redundant sequences (Martinez et al., 2004). 90% of the assembly was found in 21 scaffolds, while 50% was found in eight scaffolds larger than 1.9 Mbp.

The screening of the telomere sequences was performed as described above using the scaffolds of the assemblage v2.0. The basic telomere unit of *P. chrysosporium* is the heptamer TTTAGGG. 16 out of the 232 scaffolds contained telomere sequences, eight of them with the TTTAGGG motif. In four of these scaffolds (numbers 7, 9, 23 y 159) the repetitive unit was located at an interstitial position at 3200 bp from the 3′ end flanked by a gap. This arrangement could be the result of a wrong assemblage as it can be seen in Fig. 3.

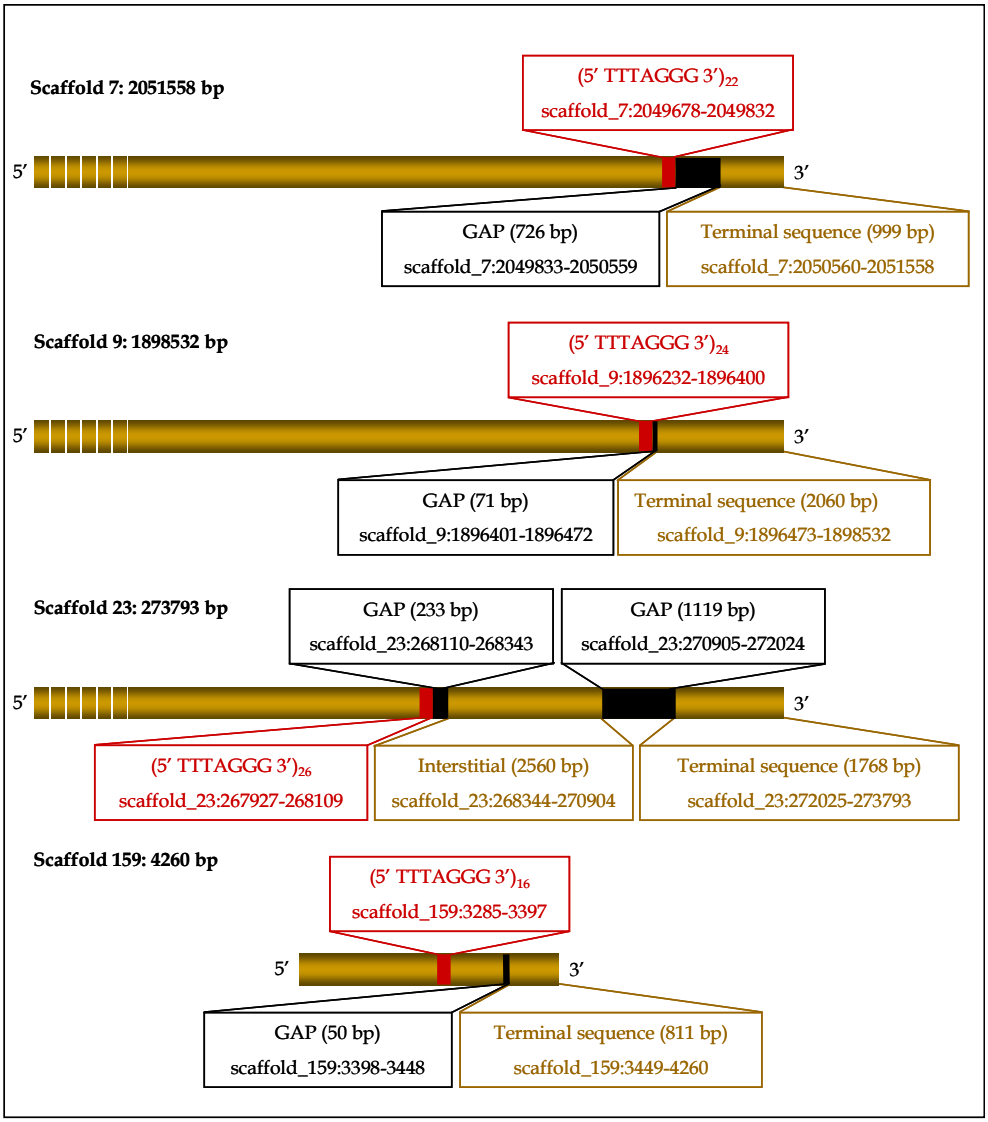


Fig. 3. Location of telomere regions in scaffolds 5, 9, 23 and 159 of the genome sequence of *P. chrisosporium*.

Another telomere-like interstitial region was found at about 300000 bp of the 3′ end of scaffold 10. The position of the interstitial telomere block within the scaffold would suggested that it could be the result of an ancestral intra-chromosomal rearrangement (inversions and/or fusions), from differential crossing-over or from the repair of double-strand break during evolution (Lin & Yan, 2008) (Fig. 4).

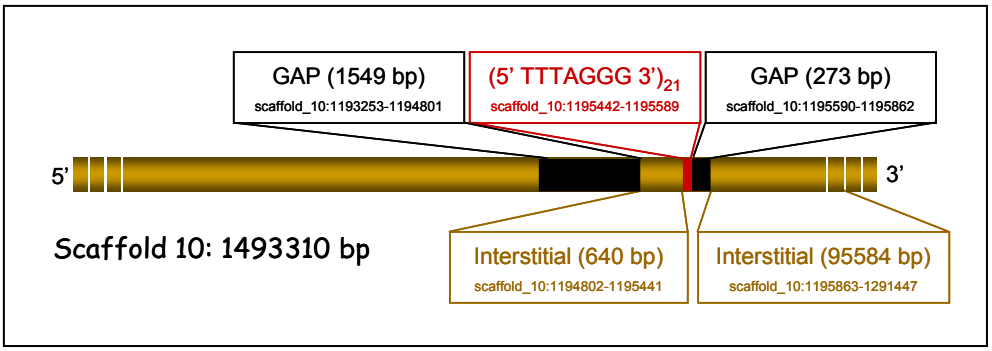


Fig. 4. Location of telomere regions in scaffold 10 of the genome sequence of *P. chrysosporium.*

It was observed that in eight of the 16 telomere-containing scaffolds identified, the repetitive unit was the complementary CCCTAAA. In five of them, the heptamer unit was placed at the 5′ end. A telomere unit in an internal region at 8772 bp of the 5′ end was present in scaffold 8 while scaffold 5 ferried a telomere motif at 2123492 bp of the 3′ end and another one at 40552bp of the 3′ end (Fig. 5).

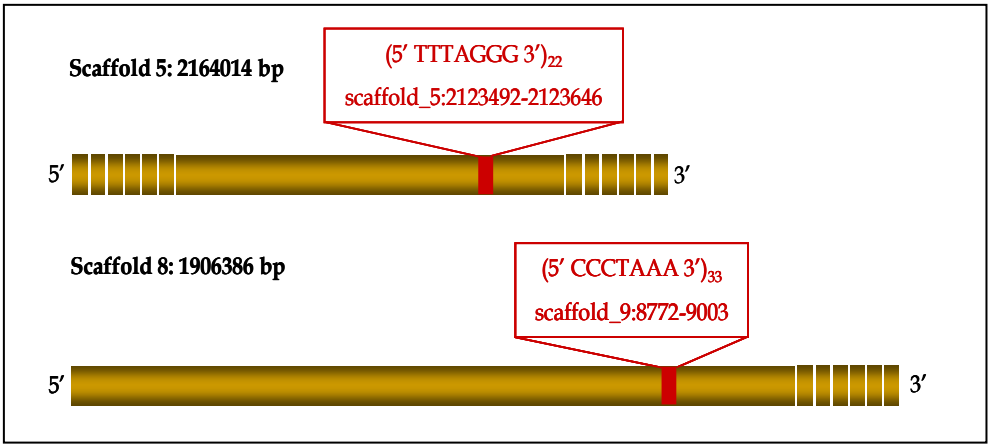


Fig. 5. Location of telomere regions in scaffolds 5 and 8 of the genome sequence of *P. chrysosporium.*

The analysis of scaffold 28 revealed two arrays of interstitial copies of the heptamer CCCCTAAA (23 and 18 repeats, respectively) placed at 11417 bp and 13853 bp from the 5′

end. An interstitial fragment of 1342 bp and a gap of 933 bp were found between them suggesting the occurrence of missassemblage (Fig. 6).

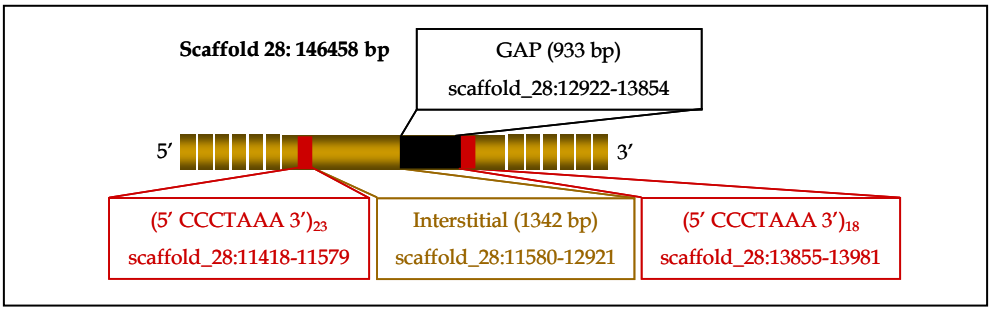


Fig. 6. Location of telomere regions in scaffold 28 of the genome sequence of *P. chrysosporium.*

In summary, the results presented here suggest that the genome of *P. chrysosporium* is arranged in at least eight linkage groups. Because scaffold 8 shows telomere repeats at both ends it is suggested that the sequence contained in this scaffold constitute the only fully assembled chromosome of *P. chrysosporium*.

2.1.3 Analysis of the telomere regions of *Pleurotus ostreatus* PC15

P. ostreatus PC15 is a monokarytic strain of an industrially-produced edible basidiomycete that has been also used as a model system for lignocellulose degradation. *P. ostreatus* differs from the other white rot model system (*P. chrysosporium*) in its enzymatic portfolio for lignin degradation. The structure of its genome was determined by linkage analysis (Larraya et al., 2000), and the complete genome of this strain has been sequenced. The assemblage v1.0 consists of 19 scaffolds of which 18 were larger than 2 Kbp. The screening of the genome for telomere sequences was carried out as described above and revealed that the elementary telomere unit of *P. ostreatus* is the hexamer TTAGGG. All scaffolds were screened for telomere regions and 19 telomere regions were recovered. In eight of them the motif TTAGGG was found, and the remaining 11 had the motif CCCTAA. The number of repetitions of the basic unit ranged from 19 to 38. As it was determined, scaffolds 1, 3, 4, 5, 6, 9, 10 and 11 show telomere repeats at both ends indicating that they are fully assembled.

2.1.4 Analysis of the telomere regions of *Postia placenta* MAD-698

P. placenta is a brown rot basidiomycete that rapidly depolymerizes the cellulose in wood without significant lignin removal. This type of decay differs sharply from white rot fungi such as *P. chrysosporium* and *P. ostreatus*. The genome of the dikaryotic strain of *P. placenta* MAD-698 revealed a genome of 90.9 Mbp assembled in 1243 scaffolds (Martinez et al., 2009). All the scaffolds of the assemblage were screened for telomere sequences as above. The basic telomere unit in this fungus is the pentamer TTAGG. The analysis of 1243 scaffolds revealed the presence of 23 regions containing telomere sequences. 12 of them carried the TTAGG sequence: in eight of them the sequence was found at the scaffold's 3′ end, three scaffolds ferried the sequence in an interstitial location (Fig. 7) and in one of them two regions with the pentamer sequence appeared at the end of the chromosome but separated by 2373 bp suggesting a missassemblage of scaffold 178 at its 3′ end (Fig. 8).

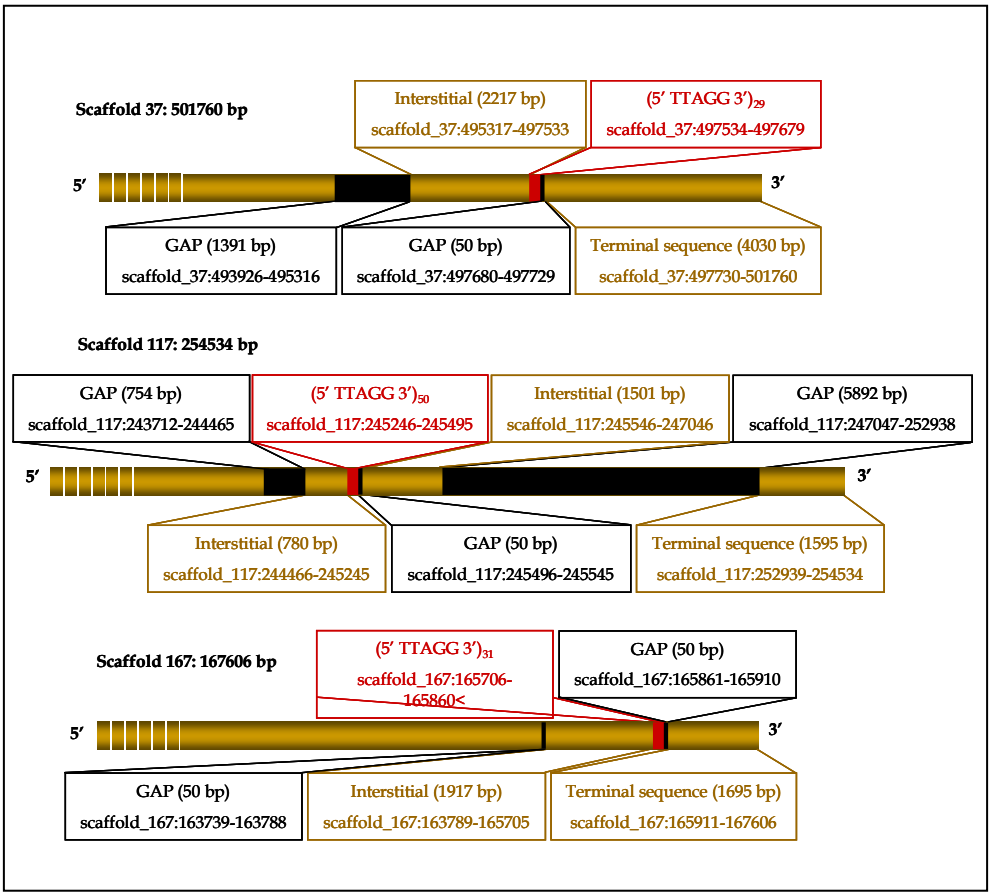


Fig. 7. Location of telomere regions in scaffolds 37, 117 and 167 of the genome sequence of *P. placenta.*

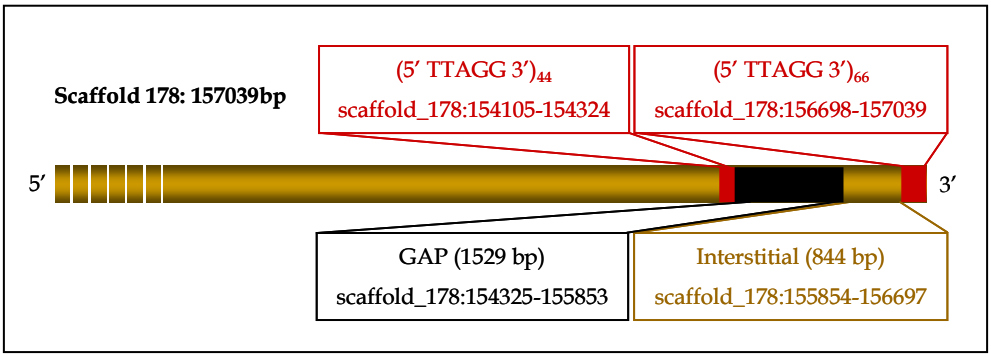


Fig. 8. Location of telomere regions in scaffold 178 of genome sequence of *P. placenta.*

The remaining 11 scaffolds carried the telomere unit CCTAA placed at the 5´end in seven of them. The scaffold 99 (Fig. 9) has a complex structure with two interstitial CCTAA regions containing the motif. One of them located towards the scaffold 5′ end, contains 40 copies of the telomere unit, and the other, located 16533 bp downstream, another interstitial region containing 19 repetitions of the unit. A gap of about 50 bp placed at the 5´end of this interstitial region suggests that this arrangement could be due to a missassemblage.

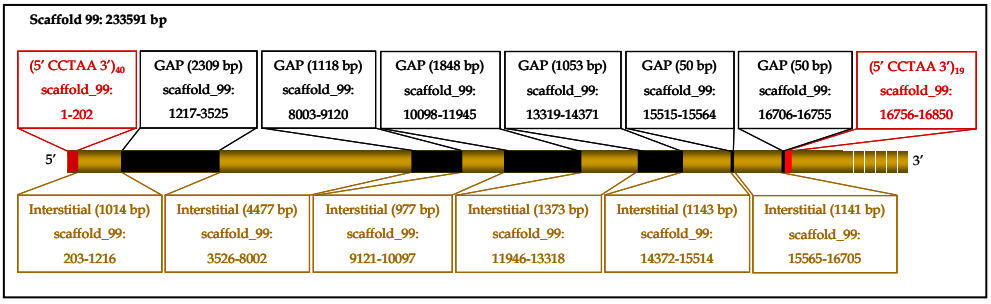


Fig. 9. Location of telomere regions in scaffold 99 of genome sequence of *P. placenta.*

The scaffold 33 also showed two interstitial regions with 40 and 30 repetitions of the of the CCTAA telomere motif. One of the regions was preceded by a gap of 701 bp suggesting that it could be a wrong assemblage. The other one (40 copies of the telomere unit) could very well represent and ITS (Interstitial Telomere Sequence) that can be produced by chromosome rearrangements as described above.

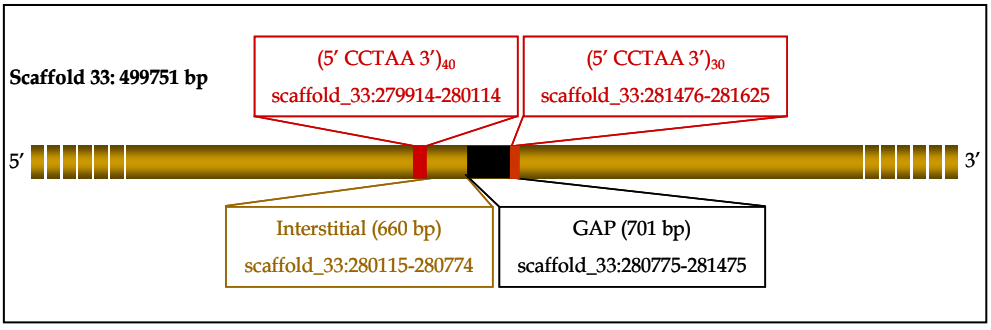


Fig. 10. Location of telomere regions in scaffold 33 of genome sequence of *P. placenta.*

Two other scaffolds with interstitial sequences were found. Scaffold 70 showed 26 repetitions of the CCTAA telomere unit at 38395 bp from the 5´ end preceded by a gap, and scaffold 144 showed 50 repetitions of the telomere unit at 181041 bp from the 5′ end preceded by another other gap of 50 bp (Fig. 11). We suggest that these structures are consequence of wrong assemblages.

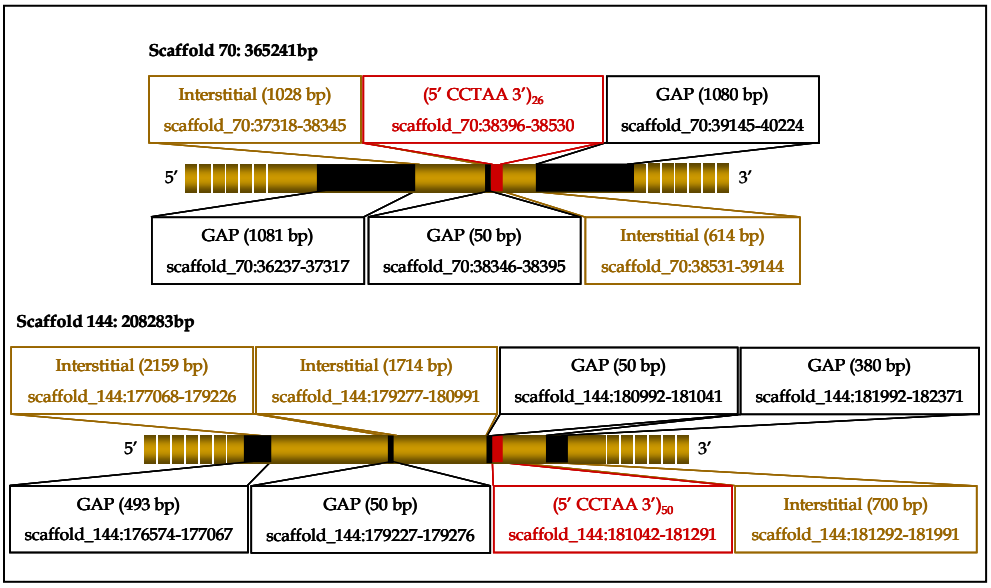


Fig. 11. Location of telomere regions in scaffolds 70 and 144 of genome sequence of *P. placenta.*

In summary, *P. placenta* has a telomere pentameric (TTAGG) basic repetitive unit. The number of copies present in the assembled genome ranged from 19 to 70. The analysis of data suggests that the minimum number of linkage groups of this species could be 12.

2.1.5 Analysis of the telomere regions of *Heterobasidion annosum*

Heterobasidion annosum is a root pathogen responsible for important losses in conifer plantations and natural forests throughout the northern hemisphere (Asiegbu et al., 2005). Genetic linkage analyses of this fungus had produced maps with 19 large linkage groups and 20 smaller ones (Lind et al., 2005), but the precise chromosome number for this species has not been conclusively determined. The v2.0 of the homokaryotic *H. annosum* genome assembly consists of 33.1 Mbp sequence assembled into 15 scaffolds at least 10 of which represent nearly complete chromosomes (http://genome.jgipsf.org/Hetan2/Hetan2.home.html).

The screening of the telomere sequences was performed as it was described above. The *H. annosum* telomere repetitive sequence is a TTAGG pentamer. The screening of the 15 scaffolds rendered 19 telomere regions. Six of them corresponded to the direct repeat sequence at the 3′ end of the scaffolds, and the remaining 13 carried the reverse sequence CCTAA at the scaffold's 5′ end. These results suggest that the genome of *H. annosum* is arranged in at least 13 linkage groups. Taking into account that scaffolds 5, 6, 9, 10, 11 and 12 contained telomere repeats at both ends, it can be concluded that they could correspond to fully assembled chromosomes.

2.1.6 Summary of the telomere regions of different basidiomycetes

A summary of the structural characteristics of the telomeres studied in this paper can be found in Table 1.

Species	Genome length assembled	Scaffold number	Telomere repetition	Average number of telomere	Minimum number of linkage groups analyzed
C. subvermispora	39.0 Mb	740	TTAGGG	19 copies	11
P. chrysosporium	35.1 Mb	232	TTTAGGG	22 copies	8
P. ostreatus	34.3 Mb	12	TTAGGG	24 copies	11
P. placenta	90.0 Mb	1243	TTAGG	25 copies	12
H. annosum	33.1 Mb	15	TTAGG	25 copies	13

Table 1. Structural characteristics of the telomeres in the basidiomycetes analyzed.

2.2 Analysis of the basidiomycetes' subtelomeric regions

The analysis of the subtelomeric regions is aimed at answering two questions: are the genes sitting at the subtelomeric a representative sample of the genes of each species or is there enrichment in sub-telomere specific genes? If this were the case, which are these telomere-enriched genes and are they conserved across species? In order to address these questions, we have recorded the genes automatically annotated in 50 Kbp regions adjacent to the different telomeres identified in the species analyzed, we have checked them manually and we have recorded and classified the Gene Ontology (GO) terms related to the genes identified in these regions (Ashburner et al., 2000).

As the number of telomere-containing scaffolds differed in the various species, (from 12 in *P ostreatus* to 1243 in *P. placenta*) the length of genomic sequence screened also varied, although in a much smaller degree (from 800 Kbp in *P. chrysosporium* to 2100 Kbp in *C. subvermispora*). The gene density of the analyzed regions was found to be related to the degree of finishing of the genome: those assembled as draft (*C. subvermispora*, *P. placenta* and *P. chrysosporium*) have gene densities lower than 0.20 genes per Kbp, whereas the gene density in the finished genomes is much higher (0.34 and 0.49 genes per Kbp). The gene density in the draft genomes seems to be significantly lower than the global gene density in the corresponding genomes. This can be due to deficiencies in the annotation of these draft genomes since the global gene density in all the species analyzed (with the exception of *P. placenta*) is very similar. In all cases, the most of the genes automatically annotated at the subtelomeric regions had no homology with others of the gene databases (cutoff criterion e-value<e-20 for BlastX) (Table 2).

The Gene Ontology annotation is aimed at standardizing the representation of gene and gene products in such a way that they can be compared among databases. This approach project provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data (Ashburner et al., 2000). There are three classification categories that are provided by the consortium: Biological Process (BP), Cellular Component (CC) and Molecular Function (MF). Each identified gene product is labeled with all the GO terms in each category that can define it. By this way, a list of GO terms provides a kind of picture describing the specific condition of the gene subset that is under study. We have studied the three categories of GO annotation in the genes annotated in 50 Kbp subtelomeric regions in the five genomes studied and recorded their general statistics (Table 3). Because of the low number of gene models identified in the subtelomeric regions and because not all of them can be labeled with a GO term, the numbers of terms in the categories of Biological

Feature	*C. subvermispora*	*P. chrysosporium*	*P. ostreatus*	*P. placenta*	*H. annosum*
Total number of scaffolds	740	232	12	1243	15
Number of subtelomeric regions analyzed	42	16	19	23	19
Scaffolds with direct motif	22	8	8	12	6
Scaffolds with reverse motif	20	8	11	11	13
Total length analyzed (Kbp)	2100	800	950	1150	950
Total number filtered model genes	283	134	319	177	470
Unknown genes (no homology)	134 (47.3 %)	56 (41.8 %)	87 (27.3 %)	64 (36.2 %)	279 (59.4 %)
Predicted protein	104 (36.7 %)	57 (42.5 %)	182 (57.1 %)	79 (44.6 %)	131 (27.9 %)
Known/annotated/putative genes	45 (15.9 %)	21 (15.7 %)	50 (15.7 %)	34 (19.2 %)	60 (12.8 %)
Subtelomere gene density (genes / Kbp)	0.13	0.17	0.34	0.10	0.49
Whole genome gene density (genes / Kbp)	0.31	0.39	0.34	0.19	0.40

Table 2. Density and homology types in the subtelomere regions.

Processes and Cellular Component are rather low, whereas the number of Molecular Function terms is much higher and produces a clearer picture of what are the subtelomeric regions coding for (Table 3).

	C. subvermispora	*P. chrysosporium*	*P. ostreatus*	*P. placenta*	*H. annosum*
GO category					
Biological Process (BP)	46	22	24	20	48
Cellular Component (CC)	15	9	10	7	12
Molecular Function (MF)	127	88	161	65	186

Table 3. GO terms richness in the subtelomeric regions of the five basidiomycetes analyzed.

In order to determine if the genes found at the subtelomere constitute a representative sample of the genes of each species, we can perform a simple statistical analysis to calculate the numbers that would be expected for each one of the GO terms in the subtelomeric regions using the whole genome data as frequency. If we do this type of study, we conclude that the distribution of the subtelomeric GO terms for each one of the categories is not a representative sample of the total gene set for each one of the species (data not shown) and, consequently, we can conclude that there are sets of genes that are found more frequently at the telomere regions. For identifying these sets, we must discuss the GO term distribution in each one of the species.

2.2.1 Analysis of the subtelomere regions of *Ceriporiopsis subvermispora* B

The analysis of the 283 gene models annotated in the subtelomeric regions of *C. subvermispora* revealed 46 BP terms in which transport, protein amino acid phosphorylation,

metabolic process, electron transport, carbohydrate metabolic process, and proteolysis are the more represented ones. The terms transport, protein amino acid phosphorylation and carbohydrate metabolism seem to be overrepresented in this region whereas the terms metabolic processes and electron transport seem to be underrepresented in comparison with the total genome. The analysis revealed 15 CC terms annotated in this region. Out of which, the terms intracellular, integral to membrane, membrane and nucleus are the most represented ones in this category. All of them, except the term nucleus, seem to be overrepresented in the subtelomeric region. Finally, out of the 127 MF terms being the more represented were ATP binding, zinc ion binding, nucleic acid binding protein, binding protein, kinase activity, protein serine/threonine kinase activity, transporter activity, oxidoreductase activity, and protein-tyrosine kinase activity. All of them seem to be overrepresented in the subtelomeric regions in comparison to the whole genome.

2.2.2 Analysis of the subtelomere regions of *Phanerochaete chrysosporium*

The analysis of the 134 gene models annotated in the subtelomeric regions of *P. chrysosporium* revealed 22 BP terms being the most represented terms proteolysis and peptidolysis, electron transport, metabolism, methionine biosynthesis, protein transport, small GTPase mediated signal transduction, and transport. The analysis revealed 9 CC terms out of which, the terms membrane, nucleus and integral to membrane are the most represented ones in this category. Finally, out of the 88 MF terms, the more represented terms are those of aspartic-type endopeptidase activity, nucleic acid binding, zinc ion binding, ATP binding, and oxidoreductase activity

2.2.3 Analysis of the subtelomere regions of *Pleurotus ostreatus*

The analysis of the 319 gene models annotated in the subtelomeric regions of *P. ostreatus* revealed 24 BP terms being the most represented terms protein amino acid phosphorylation, proteolysis, metabolic process, electron transport, and transport. The analysis revealed 10 CC terms out of which, the terms intracellular, integral to membrane, nucleus, cell wall and ribosome are the most represented ones in this category. Finally, out of the 161 MF terms, the more represented terms are those of ATP binding, nucleic acid binding, oxidoreductase activity, protein-tyrosine kinase activity and zinc ion binding.

2.2.4 Analysis of the subtelomere regions of *Postia placenta*

The analysis of the 177 gene models annotated in the subtelomeric regions of *P. placenta* revealed 20 BP terms being the most represented terms proteolysis, electron transport, protein amino acid phosphorylation and regulation of transcription DNA-dependent. The analysis revealed 7 CC terms out of which, the terms intracellular, membrane, and nucleus are the most represented ones in this category. Finally, out of the 65 MF terms, the more represented terms are those of ATP binding, nucleic acid binding and zinc ion binding.

2.2.5 Analysis of the subtelomere regions of *Heterobasidion annosum*

The analysis of the 470 gene models annotated in the subtelomeric regions of *H. annosum* revealed 48 BP terms being the most represented terms metabolic process, transport, electron transport, regulation of transcription DNA-dependent, carbohydrate metabolic process, and proteolysis. The analysis revealed 12 CC terms out of which, the terms membrane, integral to membrane, intracellular, nucleus and cytoplasm are the most

represented ones in this category. Finally, out of the 186 MF terms, the more represented terms are those of zinc ion binding, oxidoreductase activity, ATP binding, binding and heme binding.

2.2.6 Comparative analysis of the subtelomere regions of the five basidiomycetes

The record of the GO terms associated to genes found in the 50 Kbp adjacent to the telomere sequences in the five basidiomycetes analyzed reveals common patterns that permit to determine some telomere-enriched GO term families. As a preliminary study, we have taken into account the terms that are always present in among the most represented ones in each of the categories and we have extracted those of them that are present in all or most of the species analyzed. If we consider the BP category, the term electron transport is among the most represented in the five species studied and the terms transport, protein aminoacid phosphorylation, metabolic process and proteolysis are present in four of the five species. So, we can conclude that the subtelomeric regions are enriched in these processes. The species *C. subvermispora*, *H. annosum*, and *P. ostreatus* have subtelomeric regions where the more abundant BP-GO terms are highly similar whereas the subtelomeric regions in *P. chrysosporium* are the most dissimilar.

If we consider the CC category, the terms nucleus and intracellular are present among the more represented ones in all the species studied, the terms membrane and integral to membrane are present in four of the species and the terms ribosome and cytoplasm are present in three of the five species. In this case the more different subtelomeric regions in terms of the CC-GO terms are those of *C. subvermispora* and *P. placenta*, being the other three species in intermediate positions.

Finally, in the case of the MF terms, as their number is much higher, a deeper comparison can be made among the species (Table 4). The five species analyzed share the most frequent MF terms associated to the genes in the subtelomere regions supporting the idea of a preference for certain gene of gene families at these chromosome locations.

Molecular function	*Ceriporiopsis subvermispora*	*Phanerochaete chrysosporium*	*Pleurotus ostreatus*	*Postia placenta*	*Heterobasidion annosum*	Presence
zinc ion binding	4,76	3,60	2,02	4,96	4,72	5
oxidoreductase activity	2,60	2,70	2,69	2,48	3,14	5
ATP binding	5,63	2,70	5,05	6,61	2,83	5
nucleic acid binding	3,90	4,50	4,38	4,96	2,20	5
catalytic activity	2,16	1,80	1,35	2,48	2,20	5
DNA binding	1,30	0,90	1,68	2,48	1,57	5
transporter activity	3,03	1,80	1,35	ND	2,20	4
iron ion binding			1,35	1,65	1,57	3
protein-tyrosine kinase activity	2,60		2,02	2,48		3

Table 4. Frequency of Molecular Function GO terms in the subtelomere regions of the studied basidiomycetes.

3. Synteny

Synteny can be defined as the conservation of the relative positions and order of genes in different chromosomes. This definition implies that the conserved genes are related by their descent from an original ancestor (homologous genes). There are two types of homology: orthology and paralogy. We can call two genes belonging to different species as orthologous if they descent from a single gene present in the last common of the two species. On the other hand, two genes are called paralogous, if they derive from gene duplication events occurred in a given species. The orthology requires that speciation has occurred, whereas this is not necessary in the case of parology, which can occur only in individuals of the same species. As the evolutionary histories of different species may differ, groups of paralogous genes can be orthologous of a single gene in a different species. The preserved co-localization of genomic regions on chromosomes of different species is called shared synteny. This may involve relationships between genes within the syntenic regions involved, such as combinations of alleles that are advantegeous when inherited together, or shared regulatory mechanisms.

The problem of identifying syntenic regions in different genomes has been addressed using different strategies including the use of FASTA (Lipman & Pearson, 1985) and Blast (Altschul et al., 1997), and different bioinformatics approaches (Catchen et al., 2009; Grabherr et al., 2010; Tang et al., 2011). We have used a method based on the identification of synteny regions at the chromosome ends by means of a BLASTP analysis of genes in the two genomes using a cut-off threshold of e-20. Later, the Vista Synteny Viewer (http://genome.lbl.gov/vista/index.shtml) integrated into the JGI Genome Portal was used in the preliminary orthologous searching of each species. This tool enables pair-wise comparative analysis of genome assemblies at three levels of resolution. The use of synteny software programs is of particular interest to see the particular changes undergone by the subtelomeric regions during evolution (Housworth & Postlethwait, 2002). For instance, the chromosome 3 in *H. annosum* maintains the synteny with *P. ostreatus* chromosome 3, but the *H. annosum* subtelomeric region aligned with a central region of *P. ostreatus* chromosome 4 suggesting the occurrence of a translocation event after the divergence of the two species.

The different basidiomycetes were used as reference genomes and *P. ostreatus* PC15 v2.0 was the query genome. Focusing in the distal 50Kbp of each chromosome, we identified the putative gene orthologous in the subtelomeric regions of each basidiomycete. Then we used that gene sequences as query in a BlastP search of the *P. ostreatus* filtered model genes as subject. Two models were considered as orthologous if their alignment had an e-value lower than e-20 and they shared a minimum 60% in identity percentage.

The synteny between the subtelomeric regions was analyzed using *P. ostreatus* chromosomes as a reference. It was observed that seven *P. ostreatus* chromosomes (chromosomes 1, 3, 4, 5, 7, 8 and 10) harbored sequences homologous to subtelomere regions of the other basidiomycetes analyzed in this study (Table 5). 47 synteny regions were uncovered when the subtelomeric regions of *C. subvermispora* were compared to those of *P. ostreatus*. The highest number (16) corresponded to those regions placed at *P. ostreatus* chromosome 7. However, it should be mentioned that 12 *C. subvermispora* gene models were found within a 30 Kbp region of the *P. ostreatus* genome (data not shown). The lowest number of synteny regions was found when the subtelomeric regions of *P. placenta* were used as query. It

should be noticed that 10 out of 15 regions were placed on chromosome 5 of *P. ostreatus*. A similar situation was observed when synteny was analyzed between *H. annosum* and *P. ostreatus*. 10 out of 25 synteny regions of *H. annosum* mapped to chromosome 10 of *P. ostreatus*.

	P. ostreatus chromosome							
	Chr 1	Chr 3	Chr 4	Chr 5	Chr 7	Chr 8	Chr 10	Total
C. subvermispora	6	6	7	—	16	12	—	47
P. chrysosporium	4	—	4	—	4	—	—	12
P. placenta	—	—	5	10	—	—	—	15
H. annosum	—	—	6	9	—	—	10	25
Total	10	6	22	19	20	12	10	101

Table 5. Number of subtelomeric synteny regions in different basidiomycetes using *P. ostreatus* as reference.

The chromosome 4 of *P. ostreatus* can be defined as mosaic of modules of subtelomeres from the other basidiomycetes studied in this paper. The list below contains some gene models mapping to the subtelomeric regions in these basidiomycetes that were found at interstitial positions in *P. ostreatus* chromosome 4: from *C. subvermispora*, a cell cycle check point protein, a membrane transporter, a histone deacetylase, a histidine acid phosphatase, the ribosomal protein L1 and an ABC transporter; from *P. chrysosporium*, a haloacid dehalogenase-like hydrolase and a glycoside hydrolase; from *P. placenta*, an inositol polyphosphate phosphatase, a metal-dependent phosphohydrolase, and a monooxygenase; from *H. annosum*, a mitochondrial carrier transporter, a Golgi transporter, and zinc finger transcription factor (Fig. 12).

11 out of 12 gene models of *C. subvermispora* were syntenic to a 20 Kbp regions of *P. ostreatus* chromosome 8. These gene models corresponded to a nucleic acid binding protein, citrate synthase, a methyltransferase, an exoribonuclease, a phosphoribosyltransferase, a prenyltransferase, a homeobox transcription factor, a mitochondrial inner membrane protein importer, as DNA-J type heat shock protein, a RNA splicing protein, and a cytochrome c oxidase.

The genome of *P. falciparum* is organized in 14 compartmentalized chromosomes where the conserved regions form the central chromosomal domains and the polymorphic regions are at the terminal domains. In this way, housekeeping genes tend to be located at the central regions of the chromosomes, whereas the highly variable gene families responsible for the antigenic variation of the parasite are clustered towards the telomeres (Hernandez-Rivas et al., 2010). Our results suggest that a similar type of chromosomal organization would be expected to occur in basidiomycetes, although a larger number of genomes should be studied to fully support this hypothesis.

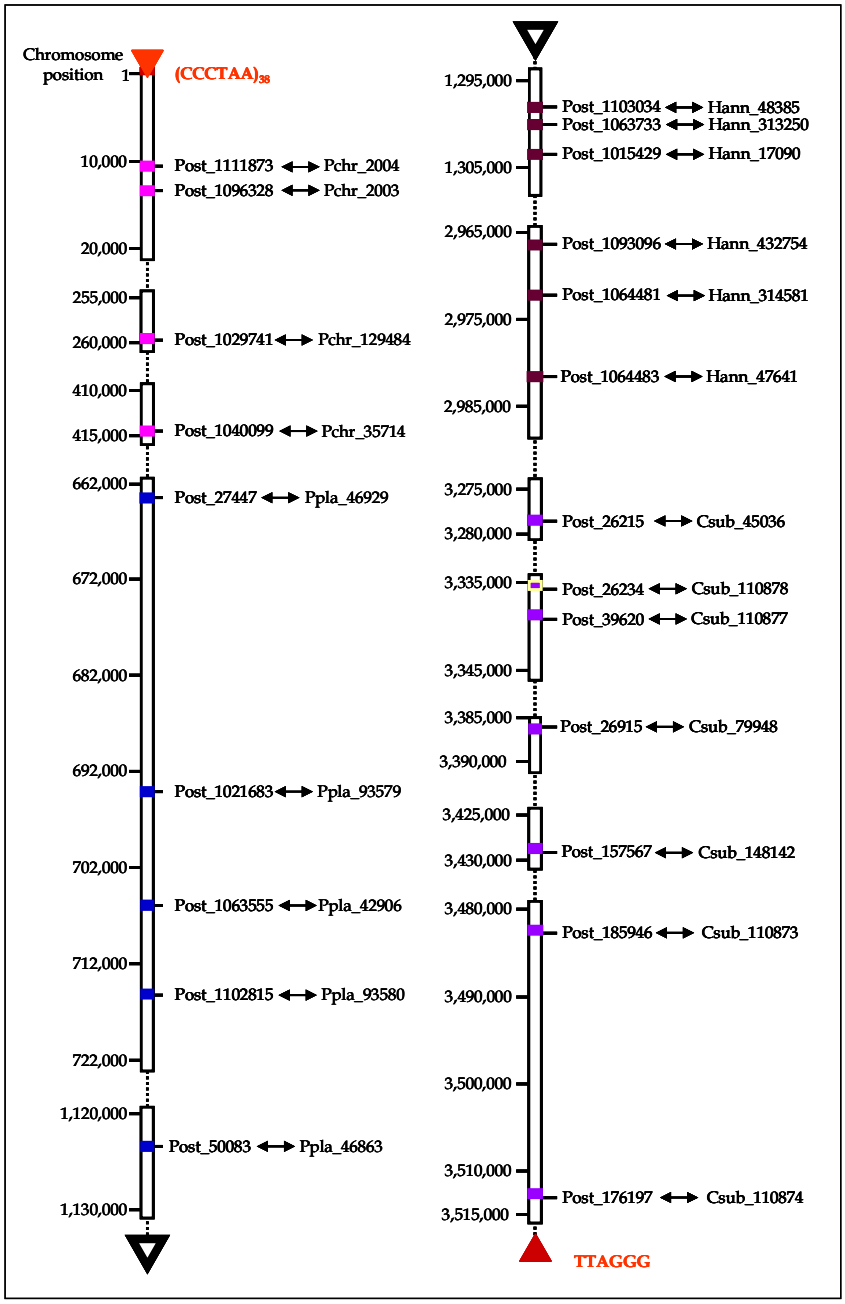


Fig. 12. Mosaic structure of *P. ostreatus* chromosome 4. The syntenic gene models of *P. ostreatus* (Post), *P. chrysosporium* (Pchr), *P. placenta* (Ppla) and *H. annosum* (Hann) are indicated along with their position on the *P. ostreatus* chromosome.

4. Conclusion

The bioinformatics analysis described in this paper allowed us to establish the type and the number of the telomere repeat unit in the basidiomycetes analyzed, to suggest the putative linkage groups in fungi where linkage maps are not available, to uncover misassembled telomere regions, and to reveal the preference for some gene models to be located at the subtelomeric regions and to uncover synteny among the subtelomere regions in the basidiomycetes analyzed.

5. Acknowledgements

This work has been supported by funds of the AGL2008-05608-C02-01 of the Spanish National Plan of Scientific Research, the Bioethanol Euroinnova project of the Goverment of Navarre (Spain), by additional institutional support from the Public University of Navarre. Some of the sequence data were produced in genome sequence projects developed at the JGI within the Community Sequence Program under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by its associate National Laboratories Lawrence Livermore and Los Alamos.
LR led and coordinated the project, GP determined bioinformatically the telomeres and subtelomere regions in the species. RC, FS and AGP made the GO analysis of the data. The manuscript was prepared by LR, and AGP.

6. References

Abad, J. P., De Pablos, B., Osoegawa, K., De Jong, P. J., Martin-Gallardo, A. & Villasante, A. (2004). TAHRE, a novel telomeric retrotransposon from Drosophila melanogaster, reveals the origin of Drosophila telomeres. *Mol Biol Evol.*, Vol. 21, No. 9, pp. 1620-1624.

Al-Wahiby, S., Wong, H. P. & Slijepcevic, P. (2005). Shortened telomeres in murine scid cells expressing mutant hRAD54 coincide with reduction in recombination at telomeres. *Mutat Res*, Vol. 578, No. 1-2, pp. 134-142.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, Vol. 25, No. 17, pp. 3389-3402.

Anderson, J. A., Song, Y. S. & Langley, C. H. (2008). Molecular population genetics of Drosophila subtelomeric DNA. *Genetics.*, Vol. 178, No. 1, pp. 477-487.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.*, Vol. 25, No. 1, pp. 25-29.

Asiegbu, F. O., Adomas, A. & Stenlid, J. (2005). Conifer root and butt rot caused by Heterobasidion annosum (Fr.) Bref. s.l. *Mol Plant Pathol.*, Vol. 6, No. 4, pp. 395-409.

Aubert, G. & Lansdorp, P. M. (2008). Telomeres and aging. *Physiol Rev*, Vol. 88, No. 2, pp. 557-579.

Azzalin, C. M. & Lingner, J. (2007). Molecular biology: damage control. *Nature.*, Vol. 448, No. 7157, pp. 1001-1002.

Azzalin, C. M., Nergadze, S. G. & Giulotto, E. (2001). Human intrachromosomal telomeric-like repeats: sequence organization and mechanisms of origin. *Chromosoma*, Vol. 110, No. 2, pp. 75-82.

Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*, Vol. 318, No. 5851, pp. 798-801.

Bailey, S. M., Cornforth, M. N., Ullrich, R. L. & Goodwin, E. H. (2004). Dysfunctional mammalian telomeres join with DNA double-strand breaks. *DNA Repair (Amst)*, Vol. 3, No. 4, pp. 349-357.

Bailey, S. M., Meyne, J., Chen, D. J., Kurimasa, A., Li, G. C., Lehnert, B. E. & Goodwin, E. H. (1999). DNA double-strand break repair proteins are required to cap the ends of mammalian chromosomes. *Proc Natl Acad Sci U S A*, Vol. 96, No. 26, pp. 14899-14904.

Barry, J. D., Ginger, M. L., Burton, P. & McCulloch, R. (2003). Why are parasite contingency genes often associated with telomeres? *Int J Parasitol.*, Vol. 33, No. 1, pp. 29-45.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C. & Eddy, S. R. (2004). The Pfam protein families database. *Nucleic Acids Res.*, Vol. 32, No. Database issue, pp. D138-141.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.*, Vol. 27, No. 2, pp. 573-580.

Bhattacharyya, A. & Blackburn, E. H. (1997). Aspergillus nidulans maintains short telomeres throughout development. *Nucleic Acids Res*, Vol. 25, No. 7, pp. 1426-1431.

Blackburn, E. H. & Gall, J. G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J Mol Biol.*, Vol. 120, No. 1, pp. 33-53.

Blasco, M. A. (2005). Telomeres and human disease: ageing, cancer and beyond. *Nat Rev Genet*, Vol. 6, No. 8, pp. 611-622.

Blasco, M. A. (2007). Telomere length, stem cells and aging. *Nat Chem Biol.*, Vol. 3, No. 10, pp. 640-649.

Blasco, M. A., Lee, H. W., Hande, M. P., Samper, E., Lansdorp, P. M., DePinho, R. A. & Greider, C. W. (1997). Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell*, Vol. 91, No. 1, pp. 25-34.

Bolzan, A. D. & Bianchi, M. S. (2006). Telomeres, interstitial telomeric repeat sequences, and chromosomal aberrations. *Mutat Res.*, Vol. 612, No. 3, pp. 189-214.

Brown, W. R., Dobson, M. J. & MacKinnon, P. (1990a). Telomere cloning and mammalian chromosome analysis. *J Cell Sci.*, Vol. 95, No. Pt 4, pp. 521-526.

Brown, W. R., MacKinnon, P. J., Villasante, A., Spurr, N., Buckle, V. J. & Dobson, M. J. (1990b). Structure and polymorphism of human telomere-associated DNA. *Cell.*, Vol. 63, No. 1, pp. 119-132.

Callen, E. & Surralles, J. (2004). Telomere dysfunction in genome instability syndromes. *Mutat Res*, Vol. 567, No. 1, pp. 85-104.

Catchen, J. M., Conery, J. S. & Postlethwait, J. H. (2009). Automated identification of conserved synteny after whole-genome duplication. *Genome Res.*, Vol. 19, No. 8, pp. 1497-1505.

Coleman, M. J., McHale, M. T., Arnau, J., Watson, A. & Oliver, R. P. (1993). Cloning and characterisation of telomeric DNA from Cladosporium fulvum. *Gene*, Vol. 132, No. 1, pp. 67-73.

Chan, C. S. & Tye, B. K. (1983a). A family of Saccharomyces cerevisiae repetitive autonomously replicating sequences that have very similar genomic environments. *J Mol Biol.*, Vol. 168, No. 3, pp. 505-523.

Chan, C. S. & Tye, B. K. (1983b). Organization of DNA sequences and replication origins at yeast telomeres. *Cell.*, Vol. 33, No. 2, pp. 563-573.

Chan, S. R. & Blackburn, E. H. (2004). Telomeres and telomerase. *Philos Trans R Soc Lond B Biol Sci*, Vol. 359, No. 1441, pp. 109-121.

De Cian, A., Lacroix, L., Douarre, C., Temime-Smaali, N., Trentesaux, C., Riou, J. F. & Mergny, J. L. (2008). Targeting telomeres and telomerase. *Biochimie*, Vol. 90, No. 1, pp. 131-155.

de Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev*, Vol. 19, No. 18, pp. 2100-2110.

de Lange, T. (2009). How telomeres solve the end-protection problem. *Science*, Vol. 326, No. 5955, pp. 948-952.

De Las Penas, A., Pan, S. J., Castano, I., Alder, J., Cregg, R. & Cormack, B. P. (2003). Virulence-related surface glycoproteins in the yeast pathogen Candida glabrata are encoded in subtelomeric clusters and subject to RAP1- and SIR-dependent transcriptional silencing. *Genes Dev.*, Vol. 17, No. 18, pp. 2245-2258.

Donate, L. E. & Blasco, M. A. (2011). Telomeres in cancer and ageing. *Philos Trans R Soc Lond B Biol Sci*, Vol. 366, No. 1561, pp. 76-84.

Espejel, S., Franco, S., Rodriguez-Perales, S., Bouffler, S. D., Cigudosa, J. C. & Blasco, M. A. (2002). Mammalian Ku86 mediates chromosomal fusions and apoptosis caused by critically short telomeres. *Embo J*, Vol. 21, No. 9, pp. 2207-2219.

Faravelli, M., Azzalin, C. M., Bertoni, L., Chernova, O., Attolini, C., Mondello, C. & Giulotto, E. (2002). Molecular organization of internal telomeric sequences in Chinese hamster chromosomes. *Gene*, Vol. 283, No. 1-2, pp. 11-16.

Faravelli, M., Moralli, D., Bertoni, L., Attolini, C., Chernova, O., Raimondi, E. & Giulotto, E. (1998). Two extended arrays of a satellite DNA sequence at the centromere and at the short-arm telomere of Chinese hamster chromosome 5. *Cytogenet Cell Genet*, Vol. 83, No. 3-4, pp. 281-286.

Fisher, T. S. & Zakian, V. A. (2005). Ku: a multifunctional protein involved in telomere maintenance. *DNA Repair (Amst)*, Vol. 4, No. 11, pp. 1215-1226.

Flint, J., Bates, G. P., Clark, K., Dorman, A., Willingham, D., Roe, B. A., Micklem, G., Higgs, D. R. & Louis, E. J. (1997a). Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum Mol Genet.*, Vol. 6, No. 8, pp. 1305-1313.

Flint, J., Thomas, K., Micklem, G., Raynham, H., Clark, K., Doggett, N. A., King, A. & Higgs, D. R. (1997b). The relationship between chromosome structure and function at a human telomeric region. *Nat Genet.*, Vol. 15, No. 3, pp. 252-257.

Freitas-Junior, L. H., Bottius, E., Pirrit, L. A., Deitsch, K. W., Scheidig, C., Guinet, F., Nehrbass, U., Wellems, T. E. & Scherf, A. (2000). Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P. falciparum. *Nature.*, Vol. 407, No. 6807, pp. 1018-1022.

Fry, M. (2007). Tetraplex DNA and its interacting proteins. *Front Biosci*, Vol. 12, No., pp. 4336-4351.

Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., FitzHugh, W., Ma, L. J., Smirnov, S., Purcell, S., Rehman, B., Elkins, T., Engels, R., Wang, S., Nielsen, C. B., Butler, J., Endrizzi, M., Qui, D., Ianakiev, P., Bell-Pedersen, D., Nelson, M. A., Werner-Washburne, M., Selitrennikoff, C. P., Kinsey, J. A., Braun, E. L., Zelter, A., Schulte, U., Kothe, G. O., Jedd, G., Mewes, W., Staben, C., Marcotte, E., Greenberg, D., Roy, A., Foley, K., Naylor, J., Stange-Thomann, N., Barrett, R., Gnerre, S., Kamal, M., Kamvysselis, M., Mauceli, E., Bielke, C., Rudd, S., Frishman, D., Krystofova, S., Rasmussen, C., Metzenberg, R. L., Perkins, D. D., Kroken, S., Cogoni, C., Macino, G., Catcheside, D., Li, W., Pratt, R. J., Osmani, S. A., DeSouza, C. P., Glass, L., Orbach, M. J., Berglund, J. A., Voelker, R., Yarden, O., Plamann, M., Seiler, S., Dunlap, J., Radford, A., Aramayo, R., Natvig, D. O., Alex, L. A., Mannhaupt, G., Ebbole, D. J., Freitag, M., Paulsen, I., Sachs, M. S., Lander, E. S., Nusbaum, C. & Birren, B. (2003). The genome sequence of the filamentous fungus Neurospora crassa. *Nature.*, Vol. 422, No. 6934, pp. 859-868.

Gao, W., Khang, C. H., Park, S. Y., Lee, Y. H. & Kang, S. (2002). Evolution and organization of a highly dynamic, subtelomeric helicase gene family in the rice blast fungus Magnaporthe grisea. *Genetics*, Vol. 162, No. 1, pp. 103-112.

Gibson, A. W., Wojciechowicz, L. A., Danzi, S. E., Zhang, B., Kim, J. H., Hu, Z. & Michels, C. A. (1997). Constitutive mutations of the Saccharomyces cerevisiae MAL-activator genes MAL23, MAL43, MAL63, and mal64. *Genetics.*, Vol. 146, No. 4, pp. 1287-1298.

Gilson, E. & Geli, V. (2007). How telomeres are replicated. *Nat Rev Mol Cell Biol*, Vol. 8, No. 10, pp. 825-838.

Goytisolo, F. A., Samper, E., Edmonson, S., Taccioli, G. E. & Blasco, M. A. (2001). The absence of the DNA-dependent protein kinase catalytic subunit in mice results in anaphase bridges and in increased telomeric fusions with normal telomere length and G-strand overhang. *Mol Cell Biol*, Vol. 21, No. 11, pp. 3642-3651.

Grabherr, M. G., Russell, P., Meyer, M., Mauceli, E., Alfoldi, J., Di Palma, F. & Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics*, Vol. 26, No. 9, pp. 1145-1151.

Greider, C. W. (1998). Telomeres and senescence: the history, the experiment, the future. *Curr Biol.*, Vol. 8, No. 5, pp. R178-181.

Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H. & de Lange, T. (1999). Mammalian telomeres end in a large duplex loop. *Cell.*, Vol. 97, No. 4, pp. 503-514.

Halme, A., Bumgarner, S., Styles, C. & Fink, G. R. (2004). Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast. *Cell.*, Vol. 116, No. 3, pp. 405-415.

Hande, M. P. (2004). DNA repair factors and telomere-chromosome integrity in mammalian cells. *Cytogenet Genome Res*, Vol. 104, No. 1-4, pp. 116-122. Epub 292010 Feb 290511.

Hande, M. P., Samper, E., Lansdorp, P. & Blasco, M. A. (1999). Telomere length dynamics and chromosomal instability in cells derived from telomerase null mice. *J Cell Biol*, Vol. 144, No. 4, pp. 589-601.

Hastie, N. D. & Allshire, R. C. (1989). Human telomeres: fusion and interstitial sites. *Trends Genet*, Vol. 5, No. 10, pp. 326-331.

Hernandez-Rivas, R., Perez-Toledo, K., Herrera Solorio, A. M., Delgadillo, D. M. & Vargas, M. (2010). Telomeric heterochromatin in Plasmodium falciparum. *Journal of Biomedicine & Biotechnology*, Vol. 2010, No., pp. 290501. Epub 292010 Feb 290511.

Hijri, M., Niculita, H. & Sanders, I. R. (2007). Molecular characterization of chromosome termini of the arbuscular mycorrhizal fungus Glomus intraradices (Glomeromycota). *Fungal Genet Biol*, Vol. 44, No. 12, pp. 1380-1386.

Holmquist, G. P. & Dancis, B. (1979). Telomere replication, kinetochore organizers, and satellite DNA evolution. *Proc Natl Acad Sci U S A*, Vol. 76, No. 9, pp. 4566-4570.

Housworth, E. A. & Postlethwait, J. (2002). Measures of synteny conservation between species pairs. *Genetics.*, Vol. 162, No. 1, pp. 441-448.

Hsu, H. L., Gilley, D., Galande, S. A., Hande, M. P., Allen, B., Kim, S. H., Li, G. C., Campisi, J., Kohwi-Shigematsu, T. & Chen, D. J. (2000). Ku acts in a unique way at the mammalian telomere to prevent end joining. *Genes Dev*, Vol. 14, No. 22, pp. 2807-2812.

Inglis, P. W., Aragao, F. J., Frazao, H., Magalhaes, B. P. & Valadares-Inglis, M. C. (2000). Biolistic co-transformation of Metarhizium anisopliae var. acridum strain CG423 with green fluorescent protein and resistance to glufosinate ammonium. *FEMS Microbiol Lett*, Vol. 191, No. 2, pp. 249-254.

Inglis, P. W., Rigden, D. J., Mello, L. V., Louis, E. J. & Valadares-Inglis, M. C. (2005). Monomorphic subtelomeric DNA in the filamentous fungus, Metarhizium anisopliae,contains a RecQ helicase-like gene. *Mol Genet Genomics*, Vol. 274, No. 1, pp. 79-90.

Jackson, A. P., Gamble, J. A., Yeomans, T., Moran, G. P., Saunders, D., Harris, D., Aslett, M., Barrell, J. F., Butler, G., Citiulo, F., Coleman, D. C., de Groot, P. W., Goodwin, T. J., Quail, M. A., McQuillan, J., Munro, C. A., Pain, A., Poulter, R. T., Rajandream, M. A., Renauld, H., Spiering, M. J., Tivey, A., Gow, N. A., Barrell, B., Sullivan, D. J. & Berriman, M. (2009). Comparative genomics of the fungal pathogens Candida dubliniensis and Candida albicans. *Genome Res.*, Vol. 19, No. 12, pp. 2231-2244.

Johnson, J. E., Smith, J. S., Kozak, M. L. & Johnson, F. B. (2008). In vivo veritas: using yeast to probe the biological functions of G-quadruplexes. *Biochimie*, Vol. 90, No. 8, pp. 1250-1263.

Karpen, G. H. & Spradling, A. C. (1992). Analysis of subtelomeric heterochromatin in the Drosophila minichromosome Dp1187 by single P element insertional mutagenesis. *Genetics.*, Vol. 132, No. 3, pp. 737-753.

Keely, S. P., Renauld, H., Wakefield, A. E., Cushion, M. T., Smulian, A. G., Fosker, N., Fraser, A., Harris, D., Murphy, L., Price, C., Quail, M. A., Seeger, K., Sharp, S., Tindal, C. J., Warren, T., Zuiderwijk, E., Barrell, B. G., Stringer, J. R. & Hall, N. (2005). Gene arrays at Pneumocystis carinii telomeres. *Genetics.*, Vol. 170, No. 4, pp. 1589-1600.

Keely, S. P., Wakefield, A. E., Cushion, M. T., Smulian, A. G., Hall, N., Barrell, B. G. & Stringer, J. R. (2001). Detailed structure of Pneumocystis carinii chromosome ends. *J Eukaryot Microbiol*, Vol. Suppl, No., pp. 118S-120S.

Kersten, P. & Cullen, D. (2007). Extracellular oxidative systems of the lignin-degrading Basidiomycete Phanerochaete chrysosporium. *Fungal Genet Biol.*, Vol. 44, No. 2, pp. 77-87.

Kusumoto, K. I., Suzuki, S. & Kashiwagi, Y. (2003). Telomeric repeat sequence of Aspergillus oryzae consists of dodeca-nucleotides. *Appl Microbiol Biotechnol*, Vol. 61, No. 3, pp. 247-251.

Larraya, L. M., Perez, G., Ritter, E., Pisabarro, A. G. & Ramirez, L. (2000). Genetic linkage map of the edible basidiomycete Pleurotus ostreatus. *Appl Environ Microbiol.*, Vol. 66, No. 12, pp. 5290-5300.

Levis, C., Giraud, T., Dutertre, M., Fortini, D. & Brygoo, Y. (1997). Telomeric DNA of Botrytis cinerea: a useful tool for strain identification. *FEMS Microbiol Lett*, Vol. 157, No. 2, pp. 267-272.

Levis, R. W. (1993). Drosophila melanogaster does not share the telomeric repeat sequence of another invertebrate, Ascaris lumbricoides. *Mol Gen Genet.*, Vol. 236, No. 2-3, pp. 440-442.

Lin, K. W. & Yan, J. (2008). Endings in the middle: current knowledge of interstitial telomeric sequences. *Mutat Res*, Vol. 658, No. 1-2, pp. 95-110.

Linardopoulou, E. V., Williams, E. M., Fan, Y., Friedman, C., Young, J. M. & Trask, B. J. (2005). Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature.*, Vol. 437, No. 7055, pp. 94-100.

Lind, M., Olson, A. & Stenlid, J. (2005). An AFLP-markers based genetic linkage map of Heterobasidion annosum locating intersterility genes. *Fungal Genet Biol.*, Vol. 42, No. 6, pp. 519-527.

Linger, B. R. & Price, C. M. (2009). Conservation of telomere protein complexes: shuffling through evolution. *Crit Rev Biochem Mol Biol*, Vol. 44, No. 6, pp. 434-446.

Lipman, D. J. & Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science.*, Vol. 227, No. 4693, pp. 1435-1441.

Lo, A. W., Sprung, C. N., Fouladi, B., Pedram, M., Sabatier, L., Ricoul, M., Reynolds, G. E. & Murnane, J. P. (2002). Chromosome instability as a result of double-strand breaks near telomeres in mouse embryonic stem cells. *Mol Cell Biol*, Vol. 22, No. 13, pp. 4836-4850.

Long, D. M., Smidansky, E. D., Archer, A. J. & Strobel, G. A. (1998). In vivo addition of telomeric repeats to foreign DNA generates extrachromosomal DNAs in the taxol-producing fungus Pestalotiopsis microspora. *Fungal Genet Biol*, Vol. 24, No. 3, pp. 335-344.

Louis, E. J. (1995). The chromosome ends of Saccharomyces cerevisiae. *Yeast.*, Vol. 11, No. 16, pp. 1553-1573.

Louis, E. J. & Borts, R. H. (1995). A complete set of marked telomeres in Saccharomyces cerevisiae for physical mapping and cloning. *Genetics.*, Vol. 139, No. 1, pp. 125-136.

Louis, E. J. & Haber, J. E. (1990). Mitotic recombination among subtelomeric Y' repeats in Saccharomyces cerevisiae. *Genetics*, Vol. 124, No. 3, pp. 547-559.

Louis, E. J. & Haber, J. E. (1992). The structure and evolution of subtelomeric Y' repeats in Saccharomyces cerevisiae. *Genetics.*, Vol. 131, No. 3, pp. 559-574.

Luke, B. & Lingner, J. (2009). TERRA: telomeric repeat-containing RNA. *Embo J.*, Vol. 28, No. 17, pp. 2503-2510.

Luke, B., Panza, A., Redon, S., Iglesias, N., Li, Z. & Lingner, J. (2008). The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in Saccharomyces cerevisiae. *Mol Cell*, Vol. 32, No. 4, pp. 465-477.

Lundblad, V. & Blackburn, E. H. (1993). An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell*, Vol. 73, No. 2, pp. 347-360.

Lundell, T. K., Makela, M. R. & Hilden, K. (2010). Lignin-modifying enzymes in filamentous basidiomycetes--ecological, functional and phylogenetic review. *J Basic Microbiol*, Vol. 50, No. 1, pp. 5-20.

Maizels, N. (2006). Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat Struct Mol Biol*, Vol. 13, No. 12, pp. 1055-1059.

Mandell, J. G., Goodrich, K. J., Bahler, J. & Cech, T. R. (2005). Expression of a RecQ helicase homolog affects progression through crisis in fission yeast lacking telomerase. *J Biol Chem*, Vol. 280, No. 7, pp. 5249-5257.

Martinez, A. T., Speranza, M., Ruiz-Duenas, F. J., Ferreira, P., Camarero, S., Guillen, F., Martinez, M. J., Gutierrez, A. & del Rio, J. C. (2005). Biodegradation of lignocellulosics: microbial, chemical, and enzymatic aspects of the fungal attack of lignin. *Int Microbiol.*, Vol. 8, No. 3, pp. 195-204.

Martinez, D., Challacombe, J., Morgenstern, I., Hibbett, D., Schmoll, M., Kubicek, C. P., Ferreira, P., Ruiz-Duenas, F. J., Martinez, A. T., Kersten, P., Hammel, K. E., Vanden Wymelenberg, A., Gaskell, J., Lindquist, E., Sabat, G., Bondurant, S. S., Larrondo, L. F., Canessa, P., Vicuna, R., Yadav, J., Doddapaneni, H., Subramanian, V., Pisabarro, A. G., Lavin, J. L., Oguiza, J. A., Master, E., Henrissat, B., Coutinho, P. M., Harris, P., Magnuson, J. K., Baker, S. E., Bruno, K., Kenealy, W., Hoegger, P. J., Kues, U., Ramaiya, P., Lucas, S., Salamov, A., Shapiro, H., Tu, H., Chee, C. L., Misra, M., Xie, G., Teter, S., Yaver, D., James, T., Mokrejs, M., Pospisek, M., Grigoriev, I. V., Brettin, T., Rokhsar, D., Berka, R. & Cullen, D. (2009). Genome, transcriptome, and secretome analysis of wood decay fungus Postia placenta supports unique mechanisms of lignocellulose conversion. *Proc Natl Acad Sci U S A.*, Vol. 106, No. 6, pp. 1954-1959.

Martinez, D., Larrondo, L. F., Putnam, N., Gelpke, M. D., Huang, K., Chapman, J., Helfenbein, K. G., Ramaiya, P., Detter, J. C., Larimer, F., Coutinho, P. M., Henrissat, B., Berka, R., Cullen, D. & Rokhsar, D. (2004). Genome sequence of the lignocellulose degrading fungus Phanerochaete chrysosporium strain RP78. *Nat Biotechnol.*, Vol. 22, No. 6, pp. 695-700.

Maser, R. S. & DePinho, R. A. (2004). Telomeres and the DNA damage response: why the fox is guarding the henhouse. *DNA Repair (Amst)*, Vol. 3, No. 8-9, pp. 979-988.

Masutomi, K., Yu, E. Y., Khurts, S., Ben-Porath, I., Currier, J. L., Metz, G. B., Brooks, M. W., Kaneko, S., Murakami, S., DeCaprio, J. A., Weinberg, R. A., Stewart, S. A. & Hahn, W. C. (2003). Telomerase maintains telomere structure in normal human cells. *Cell*, Vol. 114, No. 2, pp. 241-253.

Mefford, H. C. & Trask, B. J. (2002). The complex structure and dynamic evolution of human subtelomeres. *Nat Rev Genet.*, Vol. 3, No. 2, pp. 91-102.

Meyne, J., Baker, R. J., Hobart, H. H., Hsu, T. C., Ryder, O. A., Ward, O. G., Wiley, J. E., Wurster-Hill, D. H., Yates, T. L. & Moyzis, R. K. (1990). Distribution of non-telomeric sites of the (TTAGGG)n telomeric sequence in vertebrate chromosomes. *Chromosoma*, Vol. 99, No. 1, pp. 3-10.

Mondoux, M. & Zakian, V. A. (2005). Telomere position effect: silencing near the end. In: *Telomeres* T. de Lange, V. Lundblad and E. H. Blackburn, pp. (261-316), CSHL Press, Cold Spring Harbor, New York.

Morin, G. B. (1989). The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell.*, Vol. 59, No. 3, pp. 521-529.

Murnane, J. P. & Sabatier, L. (2004). Chromosome rearrangements resulting from telomere dysfunction and their role in cancer. *Bioessays*, Vol. 26, No. 11, pp. 1164-1174.

Nagele, R. G., Velasco, A. Q., Anderson, W. J., McMahon, D. J., Thomson, Z., Fazekas, J., Wind, K. & Lee, H. (2001). Telomere associations in interphase nuclei: possible role in maintenance of interphase chromosome topology. *J Cell Sci*, Vol. 114, No. Pt 2, pp. 377-388.

Ogami, M., Ikura, Y., Ohsawa, M., Matsuo, T., Kayo, S., Yoshimi, N., Hai, E., Shirai, N., Ehara, S., Komatsu, R., Naruko, T. & Ueda, M. (2004). Telomere shortening in human coronary artery diseases. *Arterioscler Thromb Vasc Biol*, Vol. 24, No. 3, pp. 546-550.

Oganesian, L., Moon, I. K., Bryan, T. M. & Jarstfer, M. B. (2006). Extension of G-quadruplex DNA by ciliate telomerase. *Embo J*, Vol. 25, No. 5, pp. 1148-1159.

Padmavathi, J., UmaDevi, K., Rao, C. U. & Reddy, N. N. (2003). Telomere fingerprinting for assessing chromosome number, isolate typing and recombination in the entomopathogen Beauveria bassiana. *Mycol Res*, Vol. 107, No. Pt 5, pp. 572-580.

Paeschke, K., Juranek, S., Simonsson, T., Hempel, A., Rhodes, D. & Lipps, H. J. (2008). Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates. *Nat Struct Mol Biol*, Vol. 15, No. 6, pp. 598-604.

Paeschke, K., McDonald, K. R. & Zakian, V. A. (2010). Telomeres: structures in need of unwinding. *FEBS Lett*, Vol. 584, No. 17, pp. 3760-3772.

Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. & Lipps, H. J. (2005). Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat Struct Mol Biol*, Vol. 12, No. 10, pp. 847-854.

Palm, W. & de Lange, T. (2008). How shelterin protects mammalian telomeres. *Annu Rev Genet*, Vol. 42, No., pp. 301-334.

Perez, G., Pangilinan, J., Pisabarro, A. G. & Ramirez, L. (2009). Telomere organization in the ligninolytic basidiomycete Pleurotus ostreatus. *Appl Environ Microbiol.*, Vol. 75, No. 5, pp. 1427-1436.

Pryde, F. E., Gorham, H. C. & Louis, E. J. (1997). Chromosome ends: all the same under their caps. *Curr Opin Genet Dev.*, Vol. 7, No. 6, pp. 822-828.

Pryde, F. E. & Louis, E. J. (1997). Saccharomyces cerevisiae telomeres. A review. *Biochemistry (Mosc).* Vol. 62, No. 11, pp. 1232-1241.

Rachidi, N., Martinez, M. J., Barre, P. & Blondin, B. (2000). Saccharomyces cerevisiae PAU genes are induced by anaerobiosis. *Mol Microbiol.*, Vol. 35, No. 6, pp. 1421-1430.

Rehmeyer, C., Li, W., Kusaba, M., Kim, Y. S., Brown, D., Staben, C., Dean, R. & Farman, M. (2006). Organization of chromosome ends in the rice blast fungus, Magnaporthe oryzae. *Nucleic Acids Res*, Vol. 34, No. 17, pp. 4685-4701.

Rhodes, D., Fairall, L., Simonsson, T., Court, R. & Chapman, L. (2002). Telomere architecture. *EMBO Rep*, Vol. 3, No. 12, pp. 1139-1145.

Riethman, H., Ambrosini, A., Castaneda, C., Finklestein, J., Hu, X. L., Mudunuri, U., Paul, S. & Wei, J. (2004). Mapping and initial analysis of human subtelomeric sequence assemblies. *Genome Res.*, Vol. 14, No. 1, pp. 18-28.

Riethman, H., Ambrosini, A. & Paul, S. (2005). Human subtelomere structure and variation. *Chromosome Res*, Vol. 13, No. 5, pp. 505-515.

Riethman, H. C., Xiang, Z., Paul, S., Morse, E., Hu, X. L., Flint, J., Chi, H. C., Grady, D. L. & Moyzis, R. K. (2001). Integration of telomere sequences with the draft human genome sequence. *Nature.*, Vol. 409, No. 6822, pp. 948-951.

Ruiz-Duenas, F. J. & Martinez, A. T. (2009). Microbial degradation of lignin: how a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microb Biotechnol.*, Vol. 2, No. 2, pp. 164-177. doi: 110.1111/j.1751-7915.2008.00078.x.

Samper, E., Goytisolo, F. A., Slijepcevic, P., van Buul, P. P. & Blasco, M. A. (2000). Mammalian Ku86 protein prevents telomeric fusions independently of the length of TTAGGG repeats and the G-strand overhang. *EMBO Rep*, Vol. 1, No. 3, pp. 244-252.

Sanchez-Alonso, P. & Guzman, P. (1998). Organization of chromosome ends in Ustilago maydis. RecQ-like helicase motifs at telomeric regions. *Genetics.*, Vol. 148, No. 3, pp. 1043-1054.

Sanchez-Alonso, P. & Guzman, P. (2008). Predicted elements of telomere organization and function in *Ustilago maydis*. *Fungal Genet Biol.*, Vol. 45 Suppl 1, No., pp. S54-62.

Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J. & Pluckthun, A. (2001). In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc Natl Acad Sci U S A*, Vol. 98, No. 15, pp. 8572-8577.

Schoeftner, S. & Blasco, M. A. (2008). Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol*, Vol. 10, No. 2, pp. 228-236.

Schoeftner, S. & Blasco, M. A. (2009). Chromatin regulation and non-coding RNAs at mammalian telomeres. *Semin Cell Dev Biol*, Vol. 21, No. 2, pp. 186-193.

Selker, E. U. (1990). Premeiotic instability of repeated sequences in Neurospora crassa. *Annu Rev Genet*, Vol. 24, No., pp. 579-613.

Sherr, C. J. & McCormick, F. (2002). The RB and p53 pathways in cancer. *Cancer Cell*, Vol. 2, No. 2, pp. 103-112.

Shore, D. & Bianchi, A. (2009). Telomere length regulation: coupling DNA end processing to feedback regulation of telomerase. *Embo J*, Vol. 28, No. 16, pp. 2309-2322.

Slijepcevic, P., Xiao, Y., Dominguez, I. & Natarajan, A. T. (1996). Spontaneous and radiation-induced chromosomal breakage at interstitial telomeric sites. *Chromosoma*, Vol. 104, No. 8, pp. 596-604.

Smogorzewska, A., Karlseder, J., Holtgreve-Grez, H., Jauch, A. & de Lange, T. (2002). DNA ligase IV-dependent NHEJ of deprotected mammalian telomeres in G1 and G2. *Curr Biol*, Vol. 12, No. 19, pp. 1635-1644.

Soler, D., Genesca, A., Arnedo, G., Egozcue, J. & Tusell, L. (2005). Telomere dysfunction drives chromosomal instability in human mammary epithelial cells. *Genes Chromosomes Cancer*, Vol. 44, No. 4, pp. 339-350.

Starr, J. M., McGurn, B., Harris, S. E., Whalley, L. J., Deary, I. J. & Shiels, P. G. (2007). Association between telomere length and heart disease in a narrow age cohort of older people. *Exp Gerontol*, Vol. 42, No. 6, pp. 571-573.

Takai, H., Smogorzewska, A. & de Lange, T. (2003). DNA damage foci at dysfunctional telomeres. *Curr Biol*, Vol. 13, No. 17, pp. 1549-1556.

Tang, H., Lyons, E., Pedersen, B., Schnable, J. C., Paterson, A. H. & Freeling, M. (2011). Screening synteny blocks in pairwise genome comparisons through integer programming. *BMC Genomics*, Vol. 12, No., pp. 102.

Teng, S. C., Chang, J., McCowan, B. & Zakian, V. A. (2000). Telomerase-independent lengthening of yeast telomeres occurs by an abrupt Rad50p-dependent, Rif-inhibited recombinational process. *Mol Cell*, Vol. 6, No. 4, pp. 947-952.

Teng, S. C. & Zakian, V. A. (1999). Telomere-telomere recombination is an efficient bypass pathway for telomere maintenance in Saccharomyces cerevisiae. *Mol Cell Biol*, Vol. 19, No. 12, pp. 8083-8093.

Teunissen, A. W. & Steensma, H. Y. (1995). Review: the dominant flocculation genes of Saccharomyces cerevisiae constitute a new subtelomeric gene family. *Yeast.*, Vol. 11, No. 11, pp. 1001-1013.

Tien, M. (1987). Properties of ligninase from Phanerochaete chrysosporium and their possible applications. *Crit Rev Microbiol*, Vol. 15, No. 2, pp. 141-168.

Uchida, W., Matsunaga, S., Sugiyama, R. & Kawano, S. (2002). Interstitial telomere-like repeats in the Arabidopsis thaliana genome. *Genes Genet Syst*, Vol. 77, No. 1, pp. 63-67.

van Steensel, B., Smogorzewska, A. & de Lange, T. (1998). TRF2 protects human telomeres from end-to-end fusions. *Cell*, Vol. 92, No. 3, pp. 401-413.

Vega, L. R., Mateyak, M. K. & Zakian, V. A. (2003). Getting to the end: telomerase access in yeast and humans. *Nat Rev Mol Cell Biol*, Vol. 4, No. 12, pp. 948-959.

Verdun, R. E. & Karlseder, J. (2007). Replication and protection of telomeres. *Nature*, Vol. 447, No. 7147, pp. 924-931.

Walter, M. F., Jang, C., Kasravi, B., Donath, J., Mechler, B. M., Mason, J. M. & Biessmann, H. (1995). DNA organization and polymorphism of a wild-type Drosophila telomere region. *Chromosoma.*, Vol. 104, No. 4, pp. 229-241.

Weber, B., Collins, C., Robbins, C., Magenis, R. E., Delaney, A. D., Gray, J. W. & Hayden, M. R. (1990). Characterization and organization of DNA sequences adjacent to the human telomere associated repeat (TTAGGG)n. *Nucleic Acids Res*, Vol. 18, No. 11, pp. 3353-3361.

Welchen, E. & Gonzalez, D. H. (2005). Differential expression of the Arabidopsis cytochrome c genes Cytc-1 and Cytc-2. Evidence for the involvement of TCP-domain protein-binding elements in anther- and meristem-specific expression of the Cytc-1 gene. *Plant Physiol*, Vol. 139, No. 1, pp. 88-100.

Wellinger, R. J., Ethier, K., Labrecque, P. & Zakian, V. A. (1996). Evidence for a new step in telomere maintenance. *Cell*, Vol. 85, No. 3, pp. 423-433.

Wellinger, R. J., Wolf, A. J. & Zakian, V. A. (1993). Saccharomyces telomeres acquire single-strand TG1-3 tails late in S phase. *Cell*, Vol. 72, No. 1, pp. 51-60.

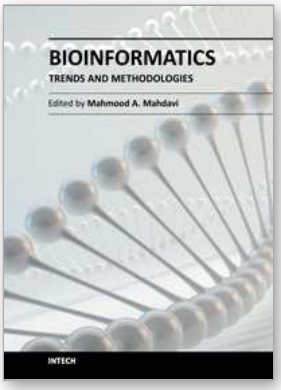
Wilkie, A. O., Higgs, D. R., Rack, K. A., Buckle, V. J., Spurr, N. K., Fischel-Ghodsian, N., Ceccherini, I., Brown, W. R. & Harris, P. C. (1991). Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell.*, Vol. 64, No. 3, pp. 595-606.

Wu, C., Kim, Y. S., Smith, K. M., Li, W., Hood, H. M., Staben, C., Selker, E. U., Sachs, M. S. & Farman, M. L. (2009). Characterization of chromosome ends in the filamentous fungus Neurospora crassa. *Genetics.*, Vol. 181, No. 3, pp. 1129-1145.

Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. (1991). Inhibition of telomerase by G-quartet DNA structures. *Nature*, Vol. 350, No. 6320, pp. 718-720.

Zakian, V. A. (1995). Telomeres: beginning to understand the end. *Science*, Vol. 270, No. 5242, pp. 1601-1607.

Zhao, Y., Sfeir, A. J., Zou, Y., Buseman, C. M., Chow, T. T., Shay, J. W. & Wright, W. E. (2009). Telomere extension occurs at most chromosome ends and is uncoupled from fill-in in human cancer cells. *Cell.*, Vol. 138, No. 3, pp. 463-475.



Bioinformatics - Trends and Methodologies
Edited by Dr. Mahmood A. Mahdavi

ISBN 978-953-307-282-1
Hard cover, 722 pages
Publisher InTech
Published online 02, November, 2011
Published in print edition November, 2011

Bioinformatics - Trends and Methodologies is a collection of different views on most recent topics and basic concepts in bioinformatics. This book suits young researchers who seek basic fundamentals of bioinformatic skills such as data mining, data integration, sequence analysis and gene expression analysis as well as scientists who are interested in current research in computational biology and bioinformatics including next generation sequencing, transcriptional analysis and drug design. Because of the rapid development of new technologies in molecular biology, new bioinformatic techniques emerge accordingly to keep the pace of in silico development of life science. This book focuses partly on such new techniques and their applications in biomedical science. These techniques maybe useful in identification of some diseases and cellular disorders and narrow down the number of experiments required for medical diagnostic.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821