

Research article

Open Access

A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis

David A Fitzpatrick*¹, Mary E Logue¹, Jason E Stajich² and Geraldine Butler¹

Address: ¹School of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland and ²Department of Molecular Genetics and Microbiology, Duke University, Durham, North Carolina 27708, USA

Email: David A Fitzpatrick* - david.a.fitzpatrick@ucd.ie; Mary E Logue - mary.e.logue@ucd.ie; Jason E Stajich - stjajich@berkeley.edu; Geraldine Butler - geraldine.butler@ucd.ie

* Corresponding author

Published: 22 November 2006

Received: 21 July 2006

BMC Evolutionary Biology 2006, 6:99 doi:10.1186/1471-2148-6-99

Accepted: 22 November 2006

This article is available from: <http://www.biomedcentral.com/1471-2148/6/99>

© 2006 Fitzpatrick et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: To date, most fungal phylogenies have been derived from single gene comparisons, or from concatenated alignments of a small number of genes. The increase in fungal genome sequencing presents an opportunity to reconstruct evolutionary events using entire genomes. As a tool for future comparative, phylogenomic and phylogenetic studies, we used both supertrees and concatenated alignments to infer relationships between 42 species of fungi for which complete genome sequences are available.

Results: A dataset of 345,829 genes was extracted from 42 publicly available fungal genomes. Supertree methods were employed to derive phylogenies from 4,805 single gene families. We found that the average consensus supertree method may suffer from long-branch attraction artifacts, while matrix representation with parsimony (MRP) appears to be immune from these. A genome phylogeny was also reconstructed from a concatenated alignment of 153 universally distributed orthologs. Our MRP supertree and concatenated phylogeny are highly congruent. Within the Ascomycota, the sub-phyla Pezizomycotina and Saccharomycotina were resolved. Both phylogenies infer that the Leotiomycetes are the closest sister group to the Sordariomycetes. There is some ambiguity regarding the placement of *Stagonospora nodurum*, the sole member of the class Dothideomycetes present in the dataset.

Within the Saccharomycotina, a monophyletic clade containing organisms that translate CTG as serine instead of leucine is evident. There is also strong support for two groups within the CTG clade, one containing the fully sexual species *Candida lusitanae*, *Candida guilliermondii* and *Debaryomyces hansenii*, and the second group containing *Candida albicans*, *Candida dubliniensis*, *Candida tropicalis*, *Candida parapsilosis* and *Lodderomyces elongisporus*. The second major clade within the Saccharomycotina contains species whose genomes have undergone a whole genome duplication (WGD), and their close relatives. We could not confidently resolve whether *Candida glabrata* or *Saccharomyces castellii* lies at the base of the WGD clade.

Conclusion: We have constructed robust phylogenies for fungi based on whole genome analysis. Overall, our phylogenies provide strong support for the classification of phyla, sub-phyla, classes and orders. We have resolved the relationship of the classes Leotiomycetes and Sordariomycetes, and have identified two classes within the CTG clade of the Saccharomycotina that may correlate with sexual status.

Background

Traditional methods of systematics based on morphology of vegetative cells, sexual states, physiological responses to fermentation and growth tests can assign fungal species to particular genera and families. However, higher-level relationships amongst these groups are less certain and are best elucidated using molecular techniques. Today single-gene phylogenies (especially 18S ribosomal DNA-based ones) have established many of the accepted relationships between fungal organisms. The benefits of the 18S rDNA approach are the vertical transmission of this gene, its ubiquity and the fact that it has slowly evolving sites. However, single-gene analyses are dependent on the gene having an evolutionary history that reflects that of the entire organism, an assumption that is not always true. It has been estimated that there are approximately 1.42 million fungi species yet to be discovered [1,2]. It follows that it is essential that we develop methods to infer a robust phylogeny of known taxonomic groups, so we can provide a framework for future studies.

Between 1990 and 2003, 560 fungal research papers reporting phylogenies were published, of which about 84% were derived using rDNA [3]. Protein coding genes are rarely used in fungal phylogenetics but when used they have the ability to resolve deep level phylogenetic relationships [4]. Phylogeny reconstruction based on a single gene may not be robust, as vital physiological processes and basic adaptive strategies do not always correlate with ribosomal derived trees [5]. Individual genes also contain a limited number of nucleotide sites and therefore limited resolution. An alternative approach to a single gene phylogeny is to combine all available phylogenetic data. There are two commonly used methods to do this: multigene concatenation and supertree analysis.

Multigene concatenation proposes that phylogenetic analysis should always be performed using all available character data, essentially sticking many aligned genes together to give a large alignment. Combining the data increases its informativeness, helps resolve nodes, basal branching and improve phylogenetic accuracy [6]. Gene concatenation has been justified on philosophical grounds, as it attempts to maximise the informativeness and explanatory power of the character data used in the analysis [7]. Numerous genome phylogenies have been derived by concatenation of universally distributed genes [8-13]. One advantage of concatenated phylogenies is that observed branch lengths are comparable across the entire tree, as they are derived from common proteins. This allows an objective, quantitative analysis of the consistency of traditional groupings [8]. However, gene concatenation also has some well-documented problems. For example, erroneous phylogenetic inferences can be made if recombination has occurred within the individual data-

sets used. Phylogenetic inference from sequence data can also be misled by systematic errors (e.g. compositional biases) [14]. These errors can be exacerbated when longer sequences are used, leading to strong support for inferences that in reality may be false.

A supertree analysis on the other hand generates a phylogeny from a set of input trees that possess fully or partially overlapping sets of taxa. Because the input trees need only overlap minimally, each source tree must share at least two taxa with one other source tree; more generally, supertree methods take as input a set of phylogenetic trees and return one or more phylogenetic trees that represent the input trees [15]. Because of the way supertrees summarise taxonomic congruence, they limit the impact of individual genes on the global topology and account for extensive differences in evolutionary rates and substitution patterns among genes in a gene-by-gene manner [16]. Therefore, we can get a phylogeny that is truly representative of the entire genome. Supertree techniques are slowly becoming commonplace in biology [17-22] and will play an important role in ascertaining the tree of life.

This study undertook a phylogenomic approach [23,24] to fungal taxonomy. Using both supertree and concatenated methods, all available fungal genomic data was analysed in an effort to address some long-standing questions regarding ancestry and sister group relationships amongst diverse fungal species.

Results and discussion

Genome data infers a robust fungal phylogeny

Our dataset consisted of 345,829 protein-coding genes from 42 fungal genomes (Table 1). Overall we identified 4,805 putative orthologous gene families (see methods). Maximum likelihood (ML) phylogenies were reconstructed for individual gene families. These 4,805 trees were used as input data for our supertree analysis, constructed using three different methods: matrix representation with parsimony (MRP) [25,26], the average consensus method (AV) [27], and the most similar supertree analysis (MSSA) method [21]. All three methods inferred congruent phylogenies, all supertree results discussed here are based on the MRP and AV phylogenies (Figure 1A&B). The results for the MSSA supertrees can be found in additional material [see additional file 1]. The YAPTP (yet another permutation tail probability randomization) test [21], which tests the null hypothesis that congruence between the input trees is no better than random, was used to assess the degree of congruence between input trees. The distribution of the scores of the 100 optimal supertrees from the YAPTP test is between 84,184 – 84,464, whereas the original unpermuted data received a score of 27,686. These scores suggested that congruence across the input trees is greater than expected by chance (P

> 0.01) [21,22] and we deemed the data suitable for supertree analyses.

Presently there is a heated philosophical debate as to what is the best approach for reconstructing genome phylogenies. Instead of using supertree methods, some prefer to concatenate universally distributed genes. In an attempt to circumvent this argument we decided to use a global congruence [28] approach, where both ideologies are used and the resulting phylogenies are cross-corroborated.

From our analysis, we initially located 227 protein families that were universally distributed between all taxa. Seven of the genomes present in this analysis have undergone a genome duplication. In an effort to minimize the effects of hidden paralogy, we only considered genes that were found in conserved syntenic blocks for selected organisms (see methods). Overall 153 of the 227 gene families met these criteria, and were used for further analysis [see additional file 2]. These gene families were individually aligned and concatenated together to give an alignment of exactly 38,000 amino acids in length. A ML phylogeny was reconstructed (Figure 2) and compared to the supertree derived from 4,805 gene families (Figure 1). In the following discussion we use the phylum, sub-phylum and class taxonomic scheme of the NCBI taxonomy browser [29].

Overall, there is a high degree of congruence between supertree and concatenated alignment phylogenies (Figures 1 & 2). Unsurprisingly all phylogenies inferred 3 strongly supported phyla branches, the Zygomycota, the Basidiomycota and the Ascomycota (Figures 1 & 2).

The Basidiomycota form a well-supported clade. The three members of the Hymenomycetes class form a robust subgroup with 100% bootstrap support (BP). Within the Hymenomycetes there is a clade containing the two members {*Coprinus cinereus* and *Phanerochaete chrysosporium*} of the order Agaricales, separate from *Cryptococcus neoformans*, which belongs to the order Tremellales.

The majority of the species studied in this analysis belong to the Ascomycota phylum. Within the Ascomycota there are two main subphyla, the Pezizomycotina and Saccharomycotina. Both these groups form separate well-supported sub-phyla clades (Figures 1 & 2). *Schizosaccharomyces pombe*, the only member of the Schizosaccharomycetes, sits outside these two sub-phyla clades.

Within the Pezizomycotina a number of well-defined class-clades are observed, namely the Sordariomycetes, the Leotiomycetes and Eurotiomycetes (Figures 1 & 2). The

relationship between these classes has been the subject of debate. Our supertrees and concatenated phylogenies infer that the Leotiomycetes and Sordariomycetes are sister classes. This agrees with the poorly supported rDNA based analysis of Lumbsch *et al* [30] but is in disagreement with Lutzoni *et al* [3], who based on a four gene combined dataset placed the Dothideomycetes as a sister group to the Sordariomycetes. The grouping of Leotiomycetes and Sordariomycetes in both our phylogenies is highly supported (100% BP) and is likely to represent the true relationship. Furthermore, a recent phylogenomic study of 17 Ascomycota genomes by Robbertse *et al* [12] reported similar inferences.

There is conflict however between our supertrees and concatenated phylogenies regarding the positioning of *Stagonospora nodorum* (the only representative of the Dothideomycetes lineage). The supertrees (Figure 1) place *S. nodorum* beside the Eurotiomycetes (100% BP), and supports the analysis of Lutzoni *et al* [3] who also group the Dothideomycetes and Eurotiomycetes lineages together. Conversely, our concatenated alignment (Figure 2) infers that *S. nodorum* is more closely related to the Sordariomycetes and Leotiomycetes lineages (100% BP). Based on their concatenated alignment Robbertse *et al* [12] have also reported conflicting inferences regarding the phylogenetic position of *S. nodorum* [12]. Their phylogenies reconstructed using neighbor joining and maximum likelihood methods inferred a sister group relationship between *S. nodorum* and Eurotiomycetes in line with our supertree inference. However a phylogeny inferred using maximum parsimony placed *S. nodorum* at the base of the Pezizomycotina [12]. To confidently resolve this incongruence additional Dothideomycetes genomes will be required.

Within the Eurotiomycetes class there is a clade corresponding to the order Onygenales {*Histoplasma capsulatum*, *Coccidioides immitis* and *Uncinocarpus reesii*}. The Onygenales clade is of interest as it contains *Coccidioides immitis*. This organism was initially classified as a protist [31] but further research showed it was fungal, and separate studies placed it in three different divisions of Eumycota [32-34]. Subsequent ribosomal phylogeny studies [35,36] suggested a close phylogenetic relationship between *C. immitis* and *U. reesii* to the exclusion of *H. capsulatum*. Our supertrees and concatenated phylogenies based on whole genome data concur with the placement of *C. immitis* and *U. reesii* as sister taxa.

The Eurotiomycetes branch containing the *Aspergillus* clade is also of interest, as supertree and concatenated phylogenies infer that *A. oryzae* and *A. terreus* are each others closest relatives (Figures 1 & 2) (100% BP respectively). A minor difference between the supertrees and concate-

Table 1: Fungal organisms used in this analysis are listed. Phylum, sub-phylum and classes are shown. *Gene sets were generated in house.

Species	Phylum	Genes	Citation or sequencing group
<i>Candida albicans</i>	Ascomycota	6,662	[86]
<i>Candida dubliniensis</i> *	Ascomycota	6,027	Sanger Centre
<i>Candida tropicalis</i> *	Ascomycota	6,530	Broad-FGI
<i>Candida parapsilosis</i> *	Ascomycota	4,891	Sanger Centre
<i>Candida lusitanae</i> *	Ascomycota	5,941	Broad-FGI
<i>Candida guilliermondii</i> *	Ascomycota	5,235	Broad-FGI
<i>Debaryomyces hansenii</i>	Ascomycota	6,896	[49]
<i>Saccharomyces bayanus</i>	Ascomycota	4,492	[87, 88]
<i>Saccharomyces castellii</i>	Ascomycota	4,677	[87]
<i>Saccharomyces cerevisiae</i>	Ascomycota	5,873	[89]
<i>Saccharomyces kudriavzevii</i>	Ascomycota	3,768	[87]
<i>Saccharomyces mikatae</i>	Ascomycota	4,525	[87, 88]
<i>Saccharomyces paradoxus</i>	Ascomycota	4,788	[88]
<i>Candida glabrata</i>	Ascomycota	5,272	[49]
<i>Kluyveromyces lactis</i>	Ascomycota	5,331	[49]
<i>Saccharomyces kluyveri</i>	Ascomycota	2,968	[87]
<i>Kluyveromyces waltii</i>	Ascomycota	5,214	[90]
<i>Ashbya gossypii</i>	Ascomycota	4,718	[91]
<i>Yarrowia lipolytica</i>	Ascomycota	6,666	[49]
<i>Magnaporthe grisea</i>	Ascomycota	11,109	[92]
<i>Neurospora crassa</i>	Ascomycota	10,620	[93]
<i>Podospora anserina</i> *	Ascomycota	10,443	Broad-FGI
<i>Chaetomium globosum</i>	Ascomycota	11,124	Broad-FGI
<i>Trichoderma reesei</i> *	Ascomycota	13,248	JGI-DOE
<i>Fusarium graminearum</i>	Ascomycota	11,640	Broad-FGI
<i>Fusarium verticillioides</i> *	Ascomycota	12,751	Broad-FGI
<i>Aspergillus oryzae</i>	Ascomycota	12,062	NITE (Japan)
<i>Aspergillus nidulans</i>	Ascomycota	9,541	Broad-FGI
<i>Aspergillus fumigatus</i>	Ascomycota	9,923	TIGR and Sanger Centre
<i>Aspergillus terreus</i> *	Ascomycota	10,285	Microbia
<i>Uncinocarpus reesii</i> *	Ascomycota	6,573	Broad-FGI
<i>Histoplasma capsulatum</i> *	Ascomycota	6,605	Broad-FGI
<i>Coccidioides immitis</i> *	Ascomycota	6,622	Broad-FGI
<i>Sclerotinia sclerotiorum</i>	Ascomycota	14,522	Broad-FGI
<i>Botrytis cinerea</i>	Ascomycota	16,448	Broad-FGI
<i>Stagonospora nodorum</i>	Ascomycota	16,597	Broad-FGI
<i>Schizosaccharomyces pombe</i>	Ascomycota	4,991	[94]
<i>Coprinus cinereus</i> *	Basidiomycota	9,452	Broad-FGI
<i>Phanerochaete chrysosporium</i> *	Basidiomycota	10,216	[95]
<i>Cryptococcus neoformans</i>	Basidiomycota	6,594	[96]
<i>Ustilago maydis</i>	Basidiomycota	6,522	Broad-FGI
<i>Rhizopus oryzae</i>	Zygomycota	17,467	Broad-FGI

*Broad-FGI, MIT/Harvard Broad Institute, funded through the Fungal Genome Initiative; JGI-DOE, Department of Energy Joint Genome Institute, Walnut Creek, California; Microbia, Cambridge, Massachusetts; Sanger Caenter, Wellcome Trust Sanger Center, Hinxton, Cambridge, United Kingdom.

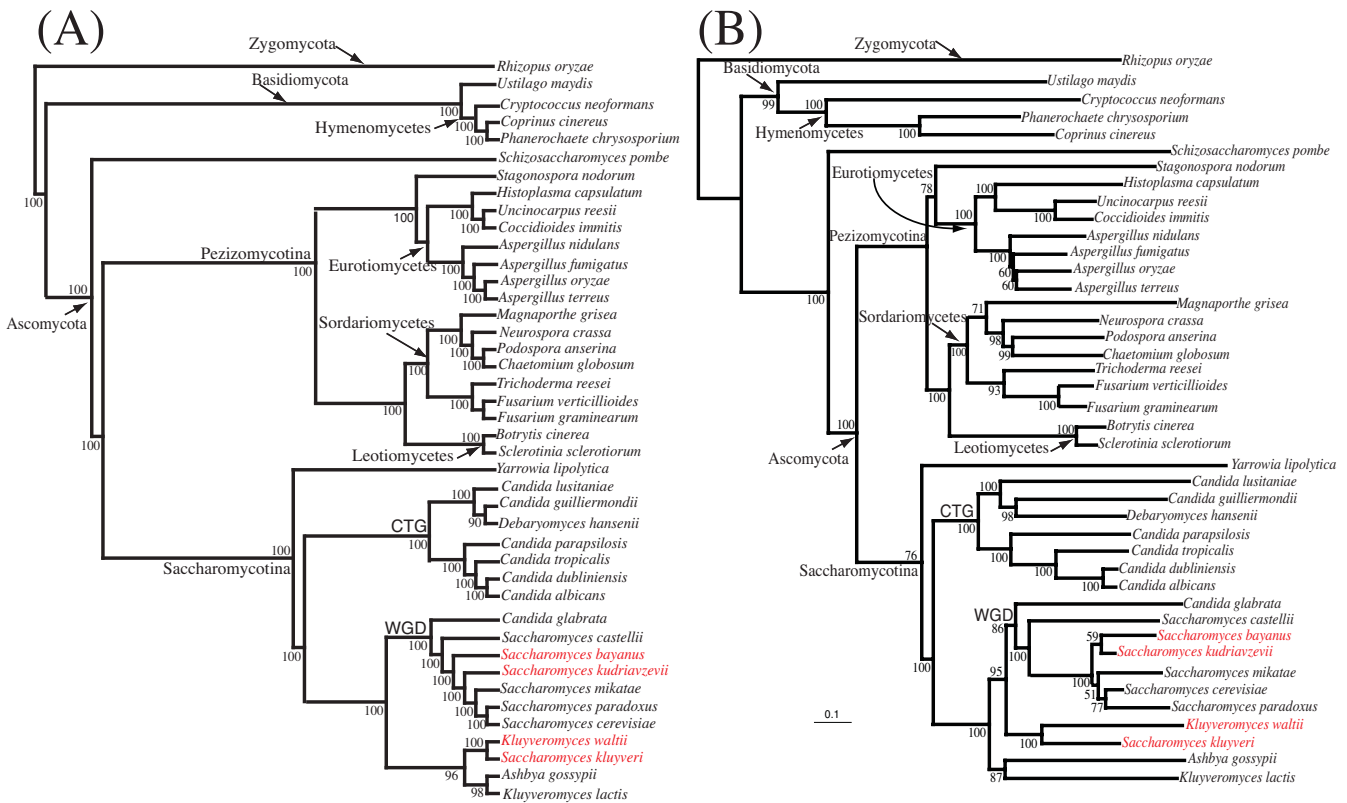


Figure 1
MRP (A) and AV (B) fungal supertrees derived from 4,805 fungal gene families. Bootstrap scores for all nodes are displayed. The AV supertree method makes use of input tree branch lengths. *Rhizopus oryzae* has been selected as an outgroup. The Basidiomycota and Ascomycota phyla form distinct clades. Subphyla and class clades are highlighted. Two clades of special interest include the node that contains the organisms that translate CTG as serine instead of leucine, and the node that contains the genomes that have undergone a genome duplication (WGD). Topological differences between supertree phylogenies are highlighted in red font.

nated phylogenies regards the phylogenetic position of *A. nidulans* and *A. fumigatus*. The concatenated alignment infers that these organisms are sister taxa (100% BP), the supertrees fails to make this inference and instead positions *A. fumigatus* beside the {*A. oryzae*, *A. terreus*} clade with 100% BP.

A number of subclass clades are evident in the Sordariomycetes clade. For example *Fusarium graminearum*, *Fusarium verticillioides* and *Trichoderma reesei* belong to the subclass Hypocreomycetidae. Similarly *Neurospora crassa*, *Chaetomium globosum* and *Podospora anserina* all belong to the subclass Sordariomycetidae. The inferred phylogenetic relationships amongst the Sordariomycetidae organisms concurs with previous phylogenetic studies [37].

Relationships within the Saccharomycotina lineage

Overall the MRP and AV supertree topologies (Figure 1A&B) are very similar. A noticeable difference occurs in the branch directly adjacent to the WGD clade. The MRP tree (and the concatenated phylogeny (Figure 2)) places

the grouping of {*K. waltii*, *S. kluyveri*} and {*K. lactis*, *A. gossypii*} as sister branches, while the AV supertree infers that {*K. waltii*, *S. kluyveri*} are closer to the WGD clade than to the {*K. lactis*, *A. gossypii*} clade. Recently Jeffroy *et al* [38] constructed a multigene phylogeny (using 13 of the 42 species included in our analysis) that is congruent with our MRP supertree for these species. They state that *K. lactis* and *A. gossypii* are evolving faster than *S. kluyveri* and *K. waltii* and are therefore likely to be "attracted" to long branches. The AV method makes use of branch length information from individual gene trees, and we suspect the inferred AV supertree phylogeny amongst the {*K. lactis*, *A. gossypii*} and {*S. kluyveri*, *K. waltii*} clades may be suffering from long-branch attraction artifacts. As additional taxa can help break long branches, it is likely that stochastic errors will be eradicated with the addition of extra genome data when it becomes available, thus eliminating erroneous inferences.

The sister group relationships amongst the *Saccharomyces sensu stricto* species also differs between our supertree phy-

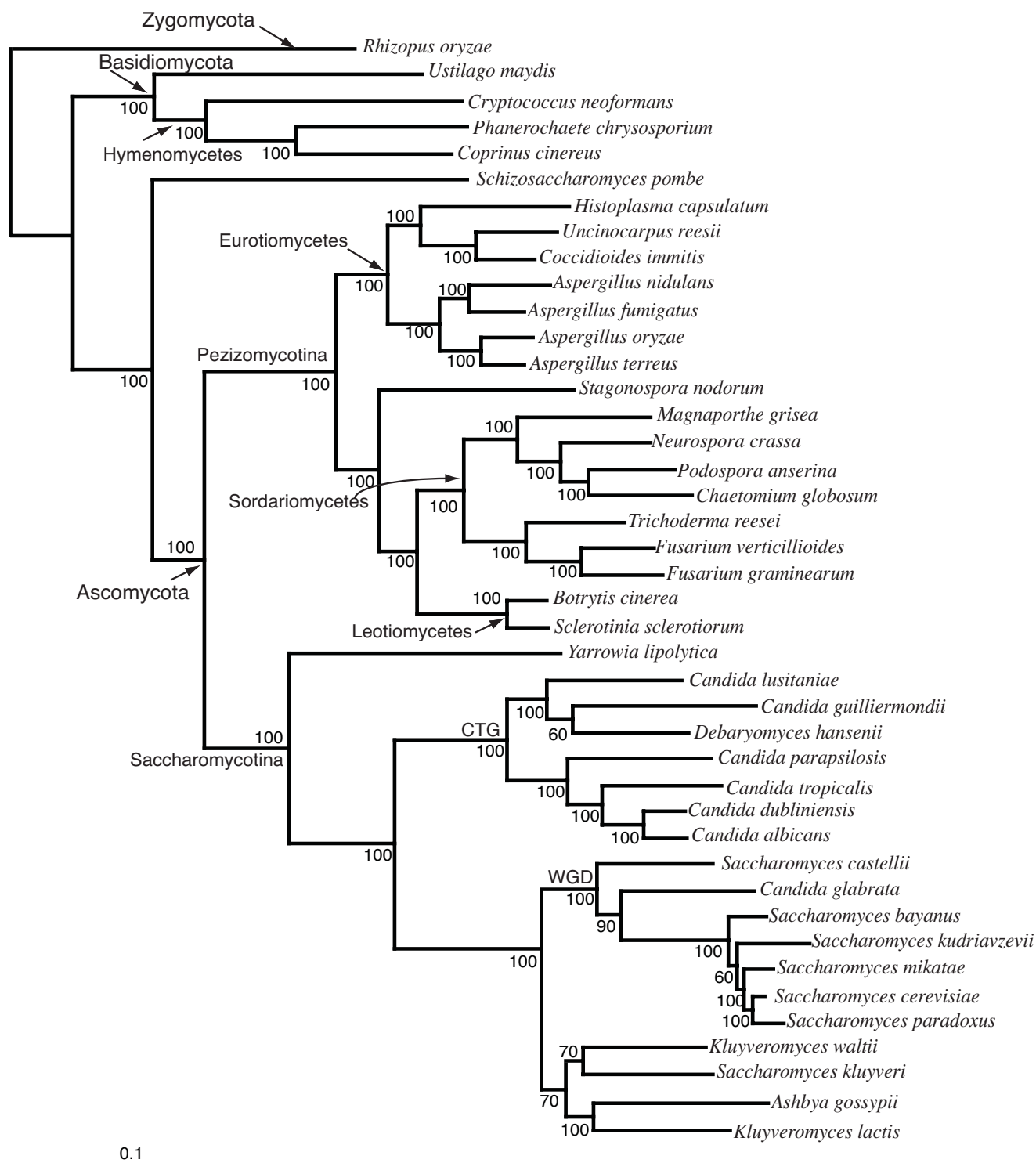


Figure 2
Maximum likelihood phylogeny reconstructed using a concatenated alignment of 153 universally distributed fungal genes. The concatenated alignment contains 42 taxa and exactly 38,000 amino acid positions. The optimum model according to ModelGenerator [85] was found to be WAG+I+G. The number of rate categories was 4 (alpha = 0.83) and the proportion of invariable sites was approximated at 0.20. Bootstrap scores for all nodes are displayed. *S. castellii* is found at the base of the WGD node.

logenies (Figure 1A&B). For example, the MRP phylogeny places *S. bayanus* at the base of the *Saccharomyces sensu stricto* node and infers a ladderised topology amongst the *Saccharomyces sensu stricto* species. The MRP inferences (Figure 1A) match those proposed by our multigene phylogeny (Figure 2) and are identical to that proposed by Jeffroy *et al.* Alternatively, the AV supertree infers that *S. bayanus* and *S. kudriavzevii* are sister taxa (Figure 1B). There is also an interesting difference regarding the relative position of *Candida glabrata* and *Saccharomyces castellii*, the supertrees and the multigene phylogeny constructed by Jeffroy *et al* [38] place *C. glabrata* at the base of the clade containing the organisms that have undergone a WGD (Figure 1A). Alternatively, our concatenated alignment infers a phylogeny with *S. castellii* at the base of the WGD clade (Figure 2), in agreement with syntenic studies [39].

It is possible that the differences between the phylogenies inferred by the MRP and AV supertrees for the *Saccharomyces sensu stricto* group are due the inclusion of paralogous sequences from the WGD species. We therefore constructed a supertree based exclusively on the species that have undergone the WGD, using 1,368 putative orthologous gene families (see methods). ML phylogenies were reconstructed for all gene families. The WGD-specific supertree (Figure 3) concurs with the MRP fungal supertree (Figure 1A) and the phylogeny of Jeffroy *et al*, suggesting this topology is correct.

The placement of *C. glabrata* as the most basal WGD genome is in disagreement with the tree inferred from the concatenated alignment (Figure 2). We therefore investigated the influence of fast evolving sites. Using a gamma distribution, we placed fast-evolving sites for each gene family into one of 8 categories, where site class 8 was the most heterogeneous, and class 1 were stationary. We systematically removed the fastest evolving sites one at a time, and rebuilt ML phylogenies based on these reduced alignments. Supertrees were once again reconstructed for these new phylogeny sets. When the two fastest classes of sites were removed, (reducing the combined length of all 1,368 genes by ~18% and ~30%), the resultant supertrees group *S. castellii* and *C. glabrata* as a monophyletic group and fail to differentiate which is closer to the outgroup [see additional file 3]. When we additionally remove the third fastest evolving site class (reducing the combined length by ~38%), the final supertree [see additional file 3] again infers *C. glabrata* at the base of the WGD clade (Figure 3). In an effort to account for compositional biases we also recoded the underlying amino acid alignments into the six Dayhoff groups and inferred individual gene phylogenies using the Bayesian criterion [see additional file 4]. The resultant supertree is identical to that shown in Figure

3, and again places *C. glabrata* at the base of the WGD clade.

To analyse the degree of conflicting phylogenetic signal within the concatenated alignment, a phylogenetic network was constructed (Figure 4). Numerous alternative splits are present (491 in total). A bootstrap analysis was performed on the phylogenetic network [see additional file 5]. It is interesting to note that we never observe a split that excludes either *C. glabrata* or *S. castellii* from the remaining WGD organisms. This conflicts with the concatenated phylogeny (Figure 2), which strongly infers that *C. glabrata* sits beside the remaining WGD organisms to the exclusion of *S. castellii*. It is possible that a systematic bias [40] may be influencing our supertrees, as synteny information clearly shows that *S. castellii* diverges from the *Saccharomyces sensu stricto* lineage before *S. castellii*, [39]. Therefore topologies that place *C. glabrata* as an outgroup to the *Saccharomyces sensu stricto* lineage and *S. castellii* are unreliable [39] and need closer scrutiny. These incongruences suggest that genome data for additional basal WGD species is required to confidently resolve inferences at the base of the WGD clade.

Phylogenetic relationships amongst *Candida* species

Both super tree (Figure 1) and superalignment (Figure 2) topologies inferred a robust monophyletic clade containing organisms which translate CTG as serine instead of leucine [41-44]. This codon reassignment has been proposed to have occurred ~170 million years ago [45]. Further inspection showed that there are two distinct CTG sub-clades, the first contains {*Candida lusitanae*, *Candida guilliermondii*, *Debaromyces hansenii*} and the second containing {*Candida tropicalis*, *Candida albicans*, *Candida dubliniensis*, *Candida parapsilosis*} (Figure 1). *C. lusitanae* and *C. guilliermondii* are haploid yeasts, and are apparently fully sexual [46-48]. *D. hansenii* is homothallic, with a fused mating locus [49,50]. In contrast, members of the second clade have at best a cryptic sexual cycle and have never been observed to undergo meiosis [51-55]. We decided to investigate this clade in further detail, and performed specific supertree, spectral and network analyses. Trace sequence data for *Lodderomyces elongisporus*, once proposed as the sexual form (teleomorph) of *C. parapsilosis* were also included [56,57].

We located 2,146 putative orthologous gene families from our CTG database (see methods). ML phylogenies were reconstructed for all gene families, and a supertree based on these trees was reconstructed. The resultant CTG specific supertree placed *L. elongisporus* within the asexual clade (Figure 5A) with high BP support (100%), in agreement with other phylogenetic studies [58,59]. A CTG specific phylogenetic network was also constructed and infers that *L. elongisporus* groups beside *C. parapsilosis*, although

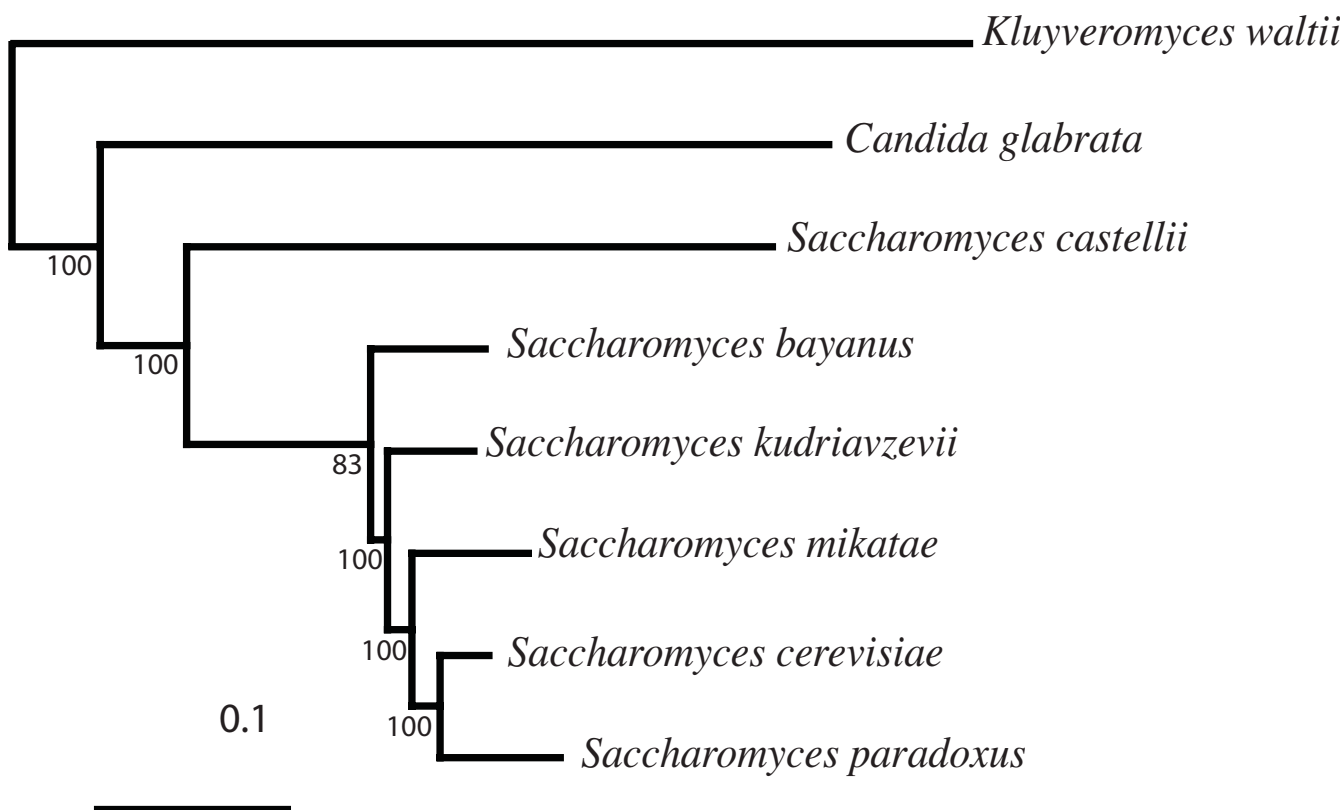


Figure 3

Average consensus supertree of WGD-specific clade inferred from 1,368 underlying phylogenies. MRP and MSSA supertrees are identical. Bootstrap scores are shown at all nodes. Bayesian analysis of recoded protein alignments and further supertree analysis yielded identical results.

there is a degree of conflict with this inference illustrated by a number of alternative splits (Figure 5B). Interestingly there is no conflict for the grouping of *C. albicans* and *C. dubliniensis* illustrating their high genotypic similarity [60]. These results raise interesting questions regarding the sexual status of the *Candida* species. It is possible that the "asexual" species are in fact fully sexual. *C. albicans* and *C. dubliniensis* have been observed to mate [53], and in addition the *C. albicans* genome contains most of the requirements for meiosis [61]. In contrast the evidence that *L. elongisporus* reproduces sexually is sketchy, and is based on the appearance of asci, with one (or sometimes two) spores [62]. It is clear that further analysis is required, which will be greatly aided when the fully annotated genome sequences of *L. elongisporus* and *C. parapsilosis* become available.

Our CTG specific supertree also suggests that *D. hansenii* and *C. guilliermondii* are sister taxa, as they are grouped together with high support (100% BP) to the exclusion of *C. lusitaniae*. Other studies [58,63] have placed *C. lusita-*

niae in a clade beside *C. guilliermondii*, and inferred a closer relationship between the two compared with *Debaryomyces* species. We found 1,208 gene families present in all CTG taxa; these were concatenated together to give a nucleotide alignment of 1,291,068 sites or 860,712 sites when third codon positions are removed. A phylogenetic network based on this nucleotide alignment (Figure 5B) corroborated the CTG-specific supertree regarding the grouping of *C. guilliermondii* and *D. hansenii* as sister taxa to the exclusion of *C. lusitaniae*. Subsequent spectral analyses (Figure 5C) reinforce our CTG specific supertree and network inferences. For example, split A (Figure 5C) shows the relatively high degree of support for the grouping of three sexual species {*C. lusitaniae*, *C. guilliermondii* and *D. hansenii*} as sister taxa. Split C groups *C. guilliermondii* and *D. hansenii* together, in agreement with our CTG supertree and network. However, there is nearly equal character support for the grouping of *C. lusitaniae* and *D. hansenii* (0.00609 vs. 0.00501) illustrated by split E (Figure 5C). Therefore, based on whole genome comparisons there is only marginal evidence for the grouping

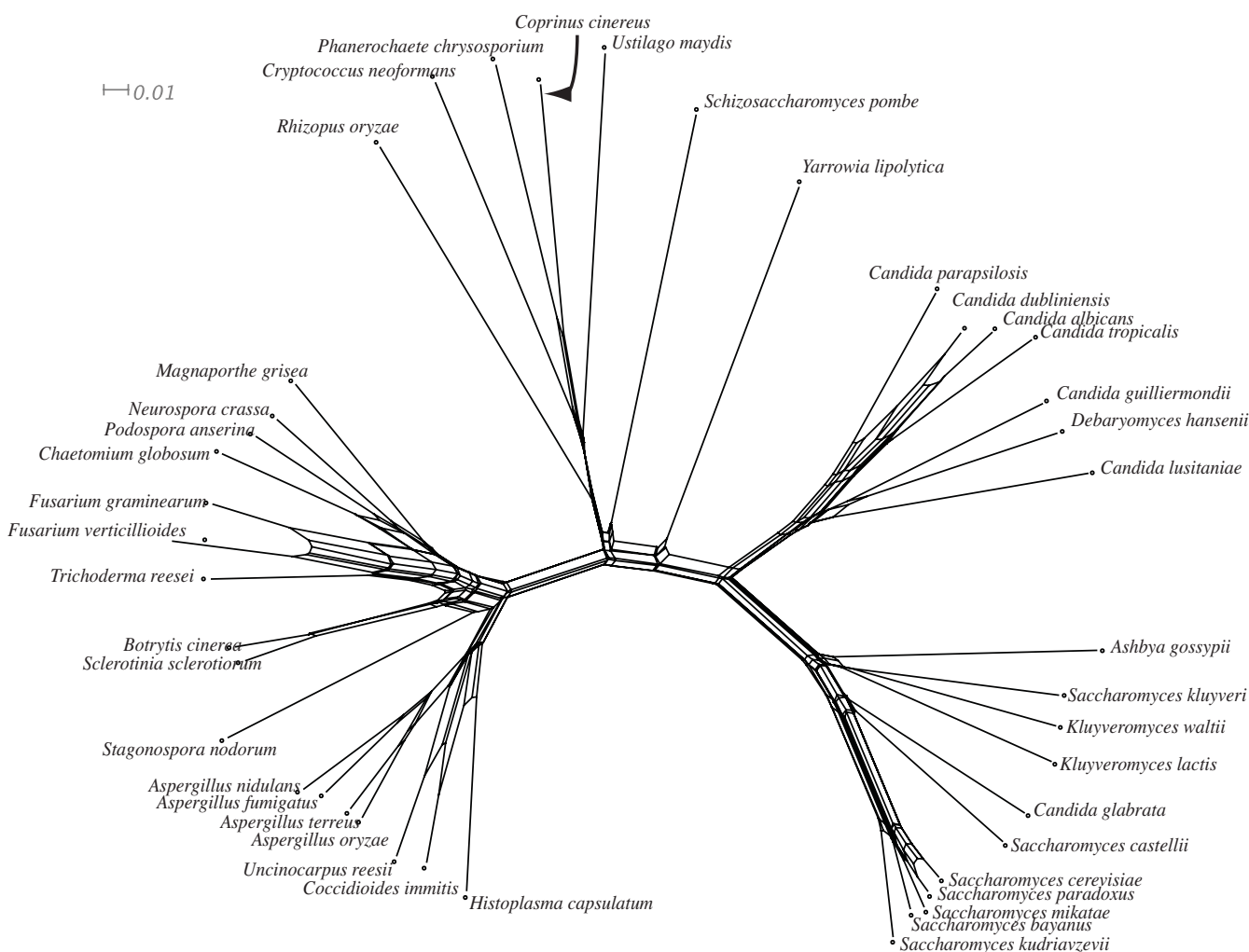


Figure 4
Phylogenetic network reconstructed using a concatenated alignment of 153 universally distributed fungal genes. The NeighborNet method was used to infer splits within the alignment. For display purposes bootstrap scores are not shown [see additional file 5].

of *C. guilliermondii* with *D. hansenii* to the exclusion of *C. lusitaniae*.

Conclusion

In this study we set out to reconstruct a fungal phylogeny from whole genome sequences. Two alternative strategies were chosen (supertrees and concatenated methods), and overall we observed a high degree of congruence between both approaches. We recovered robust fungal, phyla, sub-phyla and class clades. Overall our inferences agreed with previous phylogenetic studies based on single genes and morphological characteristics.

The phylogenomic approach undertaken in this study is novel in fungal phylogenetics as it circumvents problems associated with single gene phylogenies and selection of

robust phylogenetic markers. Our results suggest that it may be possible to piece together the tree of life using whole genomes. This is of interest as we expect the number of available genomes to increase substantially in tandem with new sequencing strategies [64], which continue to decrease the costs associated with sequencing. However, our study also shows that certain nodes of the tree (such as the WGD clade) are difficult to resolve even with genome scale data.

Methods

Sequence data

The fungal database used in this analysis consisted of 42 genomes (Table 1). Of these 28 are complete and gene datasets are available. Gene annotation for genomes with no annotations was performed using two separate

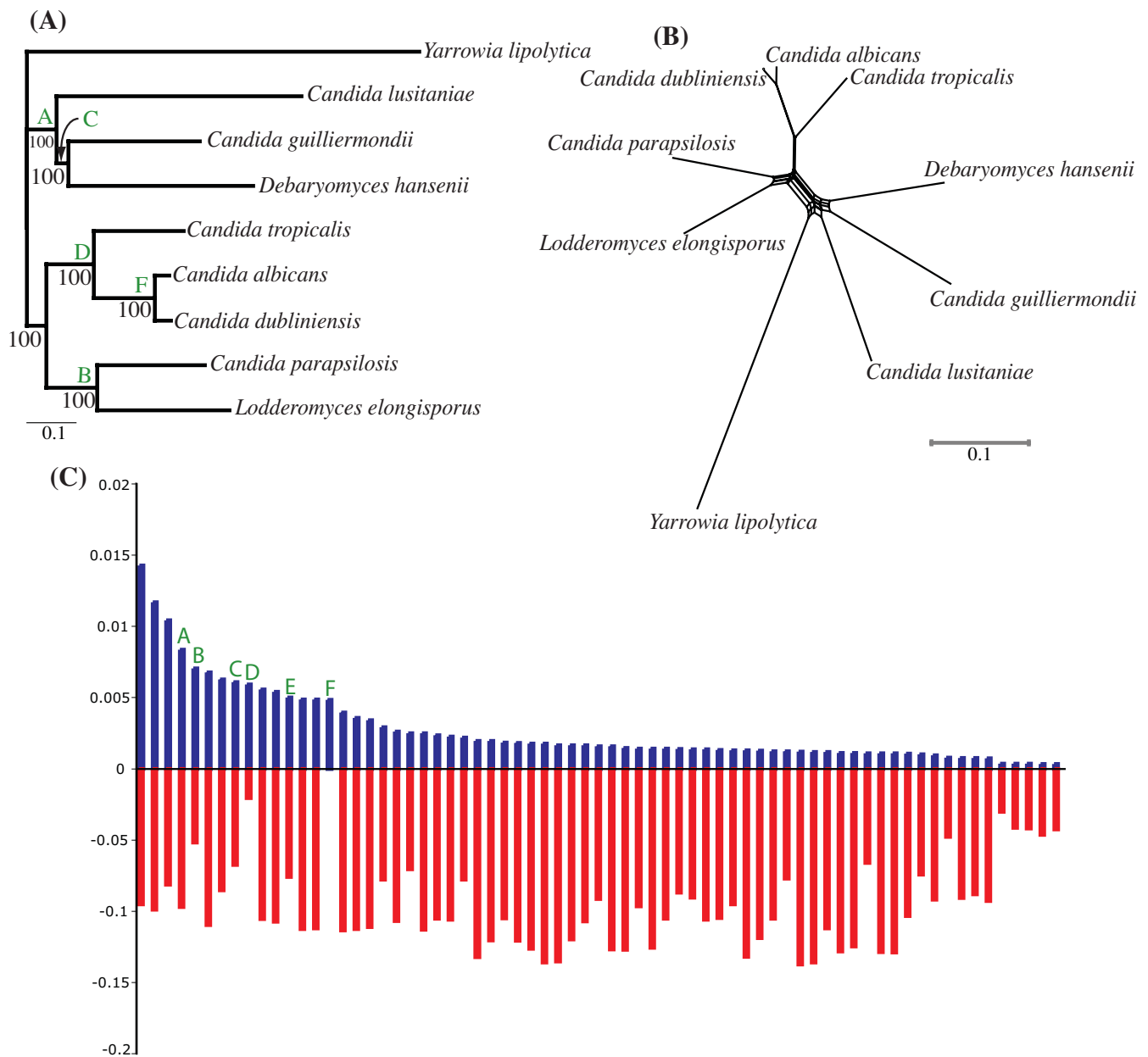


Figure 5

Average consensus supertree of CTG specific clade (A). *Y. lipolytica* was chosen as an outgroup. Bootstrap scores are shown at all nodes. (B) A phylogenetic network of 1,208 concatenated genes was inferred with the NeighborNet method. The topologies of CTG-clade specific supertree and network are congruent. (C) Spectral analysis of the concatenated alignment). Bars above the x-axis represent frequency of support for each split. Bars below the x-axis represent the sum of all corresponding conflicts. Letters above columns represent particular splits in the data, and where applicable these have also been mapped onto the supertree.

approaches. The first involved a reciprocal best BLAST [65] search with a cutoff E-value of 10^{-7} of *Candida albicans* protein coding genes against unannotated *Candida* genomes (Table 1). Top BLAST hits longer than 300 nucleotides were retained as putative open reading frames. The second approach involved a pipeline of analysis that combined several different gene prediction programs including *ab initio* programs SNAP [66], Genezilla [67], and AUGUSTUS [67] with gene models from Exonerate [68] and Genewise [69] based on alignments of proteins and Expressed sequence tags. The lines of evidence were merged into a single gene prediction using a combiner GLEAN (AJ Mackey, Q Liu, FCN Pereira, DS Roos, unpublished data). These annotations are freely available for download [70].

Reconstruction of individual gene trees

Fungal homologous sequences were identified using the BLASTP algorithm [65] with a cutoff E-value of 10^{-7} by randomly selecting a sequence from the database, finding its homologs, and removing the entire family from the database. Another randomly selected sequence from within the reduced database was then used as the new starting point for the next search. This procedure was repeated until all sequences had been removed from the database. Gene families with more than one representative from any species were not considered for further analysis. Those remaining families with a minimum of four sequences, and longer than 100 amino acids in length were selected for phylogenetic analysis. In total 5,316 protein families met these criteria. Individual protein families were aligned using ClustalW 1.81 [71] with the default settings. The average length of each protein alignment was 697 sites. Due to the large number of protein families it was not possible to manually curate all alignments. We therefore used only conserved alignment blocks, located using Gblocks version 0.91 b [72]. This filtering stage reduced the average length of our alignments to 214 sites. Permutation tail probability tests (PTP) [73,74] were performed on each alignment to test for the presence of evolutionary signal better than random ($P < 0.01$). We found that 511 alignments failed the PTP test; therefore 4,805 were used for phylogenetic reconstruction analysis. Using MultiPhyl [75] appropriate protein substitution models were selected and used to reconstruct ML phylogenies for each individual gene family. Bootstrap resampling was carried out 100 times on each alignment and the results were summarised with the majority-rule consensus method with a threshold of 70%. These phylogenies were used as input data in our supertree analysis. To account for possible compositional biases within our data, neighbor joining [76] phylogenies were also reconstructed based on distances derived from the LogDet transformation [77].

We were concerned that our strategy for locating orthologous gene families was too liberal. Therefore, we also utilised a second stricter database search strategy that located 809 gene families [see additional file 1 & additional file 6].

Supertree reconstruction

In total 4,805 input trees were used as source data for this supertree analysis. Using the supertree software package CLANN 3.0.3b1 [78] three supertree methods were used to reconstruct fungal phylogenies, the average consensus method (AV) [27], the most similar supertree analysis (MSSA) method [21], and matrix representation with parsimony (MRP) [25,26]. Using CLANN 3.0.3b1, 100 bootstrap resamplings were also carried out on the input data. We tested for the presence of signal within our data using the YAPTP test.

Multigene analysis

All proteins from the genome sequences were compared with FASTP [79] to find orthologous genes via a best bi-directional strategy. The ortholog sets for each pair of species were combined with single-linkage clustering to form multi-gene clusters of orthologs. In order to identify a set of single-copy genes across all organisms, only those clusters with exactly one member per species were considered for further analysis, we located 227 protein families that contain all fungal taxa. To help identify ohnologs and possible paralogs (with reference to the genomes that have undergone a genome duplication) we used the yeast genome browser [80,81] to filter out genes that have no syntenic evidence. Overall 153 gene families were used for further analysis [see additional file 2]. Individual gene families were aligned, manually edited and concatenated together to yield an alignment with 38,000 amino acid sites. A ML phylogeny was reconstructed for this alignment using the MultiPhyml software. Branch supports were determined via bootstrapping. In an attempt to visualise the degree of phylogenetic conflict within this concatenated alignment a phylogenetic network was generated using the NeighborNet method [82].

Investigation of specific clades

CTG clade

The genomes of *C. albicans*, *C. dubliniensis*, *C. tropicalis*, *C. parapsilosis*, *D. hansenii*, *C. guilliermondii*, *C. lusitaniae* and the outgroup *Y. lipolytica* were combined to give a CTG specific database. Data for *L. elongisporus* was retrieved from the NCBI trace database and coding genes were predicted using a reciprocal best BLASTP search against *C. albicans*. In total 2,146 gene families were longer than 100 amino acids in length, with evolutionary signal, were retained for supertree analysis. ML phylogenies were reconstructed for all gene families as described above, and representative supertrees were reconstructed. A concatenated alignment based on 1,208 genes containing all CTG

taxa was created. Alternative splits in the concatenated alignment were found using the NeighborNet method [82], and represented as a phylogenetic network with the SplitsTree software [83]. Using the software package Spectrum [84] we also performed a spectral analysis on this nucleotide alignment.

WGD clade

The WGD clade includes the genomes of *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S. bayanus*, *S. castellii* and *C. glabrata*. *K. waltii* was selected as an outgroup. For a gene family to be retained, every gene within that family must locate every other family member (and nothing else) in a reciprocal BLASTP search (cutoff E-value of 10^{-7}), be in single copy and contain a minimum of 4 taxa. We found 1,368 single gene families that met our criteria for supertree analysis. ML phylogenies were reconstructed for individual gene families as explained earlier. Phylogeny sets were also generated using Bayesian and distance based methods; [see additional file 4].

Authors' contributions

DAF and GB were involved in the design phase. MEL & JES predicted genes in unannotated genomes. DAF & JES sourced putative orthologs. DAF performed all phylogenetic analyses. DAF and GB drafted the manuscript. All authors read and approved the final manuscript.

Additional material

Additional File 1

MSSA supertree derived from 4,805 fungal gene families. Bootstrap scores for all nodes are displayed. *Rhizopus oryzae* has been selected as an outgroup. The Basidiomycota and Ascomycota phyla form distinct clades. Subphyla and class clades are highlighted.

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-6-99-S1.eps]

Additional File 2

Descriptions of the 153 universally distributed genes.

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-6-99-S2.doc]

Additional File 3

Average consensus supertrees for WGD specific clade. For each of the 1,368 underlying gene families, fast evolving sites were categorised into 8 classes. Different site classes were systematically removed and phylogenies were reconstructed based on reduced alignments. (A) Fastest evolving sites (class 8) were removed. (B) The two fastest evolving site classes (classes 7 and 8) were removed. (C) The three fastest evolving site classes (classes 6, 7 and 8) were removed. Supertrees A and B group *S. castellii* and *C. glabrata* together, supertree C places *C. glabrata* at the base of the WGD clade.

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-6-99-S3.eps]

Additional File 4

Additional Methods and Results.

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-6-99-S4.doc]

Additional File 5

Bootstrap scores for phylogenetic Network.

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-6-99-S5.doc]

Additional File 6

Supertrees (AV (A), MRP (B) and MSSA (C)) derived from the strict gene family dataset that contains 809 genes. Bootstrap scores are shown at selected nodes. Overall there is agreement with supertrees derived from the larger (liberal) dataset.

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-6-99-S6.eps]

Acknowledgements

The authors wish to acknowledge the Wellcome Trust Sanger Institute and BROAD institute of MIT & Harvard for releasing data ahead of publication. We thank Dr Chris Creevey for providing software and insight into the location of orthologous gene families. Special thanks to NUI Maynooth and Thomas Keane for allowing us run data on their distributed phylogenetics platform. We would like to acknowledge the financial support of the Irish Research Council for Science, Engineering and Technology (IRCSET) and Science Foundation Ireland (SFI). We wish to acknowledge the SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support. J.E.S. was supported by an NSF graduate research fellowship.

References

- Hawksworth DL: **The fungal dimension of biodiversity: magnitude, significance, and conservation.** *Mycol Res* 1991, **95**:641–655.
- Hawksworth DL: **The magnitude of fungal diversity: the 1.5 million species estimate revisited.** *Mycol Res* 2001, **109**:1422–1432.
- Lutzoni F, Kauff F, Cox CJ, McLaughlin D, Celio G, Dentinger B, Padamsee M, Hibbett D, James TY, Baloch E, Grube M, Reeb V, Hofstetter V, Schoch C, Arnold AE, Miadlikowska J, Spatafora J, Johnson D, Hambleton S, Crockett M, Shoemaker R, Sung G, Lucking R, Lumbsch T, O'Donnell K, Binder M, Diederich P, Ertz D, Gueidan C, Hansen K, Harris R, Hosaka K, Lim Y, Matheny B, Nishida H, Pfister D, Rogers J, Rossmann A, Schmitt I, Sipman H, Stone J, Sugiyama J, Yahr R, Vilgalys R: **Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits.** *Am J Bot* 2004, **91(10)**:1446–1480.
- Liu YJ, Whelen S, Hall BD: **Phylogenetic relationships among ascomycetes: evidence from an RNA polymerase II subunit.** *Mol Biol Evol* 1999, **16(12)**:1799–1808.
- Boucher Y, Douady CJ, Papke RT, Walsh DA, Boudreau ME, Nesbo CL, Case RJ, Doolittle WF: **Lateral gene transfer and the origins of prokaryotic groups.** In *Annu Rev Genet Volume 37*. United States ; 2003:283–328.
- Barrett M, Donoghue MJ, Sober E: **Against consensus.** *Systematic Zoology* 1991, **40**:486–493.
- Kluge AG: **A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes).** *Systematic Biology* 1989, **38**:7–25.

8. Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P: **Toward automatic reconstruction of a highly resolved tree of life.** *Science* 2006, **311(5765)**:1283-1287.
9. Brown JR, Douady CJ, Italia MJ, Marshall WE, Stanhope MJ: **Universal trees based on large combined protein sequence data sets.** In *Nat Genet Volume 28*. Issue 3 United States ; 2001:281-285.
10. Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF: **A kingdom-level phylogeny of eukaryotes based on combined protein data.** *Science* 2000, **290(5493)**:972-977.
11. Rokas A, Williams BL, King N, Carroll SB: **Genome-scale approaches to resolving incongruence in molecular phylogenies.** In *Nature Volume 425*. Issue 6960 England ; 2003:798-804.
12. Robbertse B, Reeves JB, Schoch CL, Spatafora JW: **A phylogenomic analysis of the Ascomycota.** *Fungal Genet Biol* 2006.
13. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, Lumbsch HT, Rauhut A, Reeb V, Arnold AE, Amtoft A, Stajich JE, Hosaka K, Sung GH, Johnson D, O'Rourke B, Crockett M, Binder M, Curtis JM, Slot JC, Wang Z, Wilson AW, Schussler A, Longcore JE, O'Donnell K, Mozley-Stanridge S, Porter D, Letcher PM, Powell MJ, Taylor JW, White MM, Griffith GW, Davies DR, Humber RA, Morton JB, Sugiyama J, Rossman AY, Rogers JD, Pfister DH, Hewitt D, Hansen K, Hambleton S, Shoemaker RA, Kohlmeyer J, Volkman-Kohlmeier B, Spotts RA, Serdani M, Crous PW, Hughes KW, Matsuura K, Langer E, Langer G, Untereiner WA, Lucking R, Budel B, Geiser DM, Aptroot A, Diederich P, Schmitt I, Schultz M, Yahr R, Hibbett DS, Lutzoni F, McLaughlin DJ, Spatafora JW, Vilgalys R: **Reconstructing the early evolution of Fungi using a six-gene phylogeny.** *Nature* 2006, **443(7113)**:818-822.
14. Gadagkar SR, Rosenberg MS, Kumar S: **Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree.** *J Exp Zool B Mol Dev Evol* 2005, **304(1)**:64-74.
15. Wilkinson M, Cotton JA, Creevey C, Eulenstein O, Harris SR, Lapointe FJ, Levasseur C, McInerney JO, Pisani D, Thorley JL: **The shape of supertrees to come: tree shape related properties of fourteen supertree methods.** *Syst Biol* 2005, **54(3)**:419-431.
16. Bull JJ, Huelsenbeck JP, Cunningham CW, Swofford DL, Waddell PJ: **Partitioning and Combining Data in Phylogenetic Analysis.** *Systematic Biology* 1993, **42(3)**:384-397.
17. Daubin V, Gouy M, Perriere G: **A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history.** *Genome Res* 2002, **12(7)**:1080-1090.
18. Jones KE, Purvis A, MacLarnon A, Bininda-Emonds OR, Simmons NB: **A phylogenetic supertree of the bats (Mammalia: Chiroptera).** In *Biol Rev Camb Philos Soc Volume 77*. Issue 2 England ; 2002:223-259.
19. Ruta M, Jeffery JE, Coates MI: **A supertree of early tetrapods.** In *Proc R Soc Lond B Biol Sci Volume 270*. Issue 1532 England ; 2003:2507-2516.
20. Pisani D, Yates AM, Langer MC, Benton MJ: **A genus-level supertree of the Dinosauria.** In *Proc R Soc Lond B Biol Sci Volume 269*. Issue 1494 England ; 2002:915-921.
21. Creevey CJ, Fitzpatrick DA, Philip GK, Kinsella RJ, O'Connell M J, Pentony MM, Travers SA, Wilkinson M, McInerney JO: **Does a tree-like phylogeny only exist at the tips in the prokaryotes?** *Proc R Soc Lond B Biol Sci* 2004, **271(1557)**:2551-2558.
22. Fitzpatrick DA, Creevey CJ, McInerney JO: **Genome phylogenies indicate a meaningful alpha-proteobacterial phylogeny and support a grouping of the mitochondria with the Rickettsiales.** *Mol Biol Evol* 2006, **23(1)**:74-85.
23. Delsuc F, Brinkmann H, Philippe H: **Phylogenomics and the reconstruction of the tree of life.** *Nat Rev Genet* 2005, **6(5)**:361-375.
24. Eisen JA: **Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis.** *Genome Res* 1998, **8(3)**:163-167.
25. Baum BR: **Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees.** *Taxon* 1992, **41**:3-10.
26. Ragan MA: **Matrix representation in reconstructing phylogenetic relationships among the eukaryotes.** *Biosystems* 1992, **28(1-3)**:47-55.
27. Lapointe FJ, Cucumel G: **The average consensus procedure: combination of weighted trees containing identical or overlapping sets of taxa.** In *Syst Biol Volume 46*. Issue 2 England ; 1997:306-312.
28. Levasseur C, Lapointe FJ: **War and peace in phylogenetics: a rejoinder on total evidence and consensus.** *Syst Biol* 2001, **50(6)**:881-891.
29. **Taxonomy Browser** <http://130.14.29.110/Taxonomy/>.
30. Lumbsch HT, Schmitt I, Lindemuth R, Miller A, Mangold A, Fernandez F, Huhndorf S: **Performance of four ribosomal DNA regions to infer higher-level phylogenetic relationships of inoperculate euascomycetes (Leotiomyceta).** *Mol Phylogenet Evol* 2005, **34(3)**:512-524.
31. Rixford E, Gilchrist C: **Two cases of protozoan (coccidoidal) infection of the skin and other organs.** *Johns Hopkins Hospital Report* 1896, **1**:209-268.
32. Ophuls MD: **Further observations on a pathogenic mould formerly described as a protozoan (Coccidioides immitis, Coccidioides pyrogenes).** *J Exp Med* 1905, **6**:443-485.
33. Ciferri R, Redaelli P: **Morfologia, biologia e posizione sistematica di Coccidioides immitis stiles e delle sue varietà, con notizie sul granuloma coccidioido.** *R Accad Ital* 1936, **7**:399-474.
34. Baker EE, Mrak M, Smith CE: **The morphology, taxonomy, and distribution of Coccidioides immitis rixford and gilchrist 1896.** *Farlowia* 1943, **1**:199-244.
35. Bowman BH, White TJ, Taylor JW: **Human pathogenic fungi and their close nonpathogenic relatives.** *Mol Phylogenet Evol* 1996, **6(1)**:89-96.
36. Pan S, Sigler L, Cole GT: **Evidence for a phylogenetic connection between Coccidioides immitis and Uncinocarpus reesii (Onygenaceae).** *Microbiology* 1994, **140 (Pt 6)**:1481-1494.
37. Berbee ML: **The phylogeny of plant and animal pathogens in the Ascomycota.** *Physiological and Molecular Plant Pathology* 2001, **59(4)**:165-187.
38. Jeffroy O, Brinkmann H, Delsuc F, Philippe H: **Phylogenomics: the beginning of incongruence?** *Trends Genet* 2006, **22(4)**:225-231.
39. Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH: **Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts.** *Nature* 2006, **440(7082)**:341-345.
40. Phillips MJ, Delsuc F, Penny D: **Genome-scale phylogeny and the detection of systematic biases.** *Mol Biol Evol* 2004, **21(7)**:1455-1458.
41. Kawaguchi Y, Honda H, Taniguchi-Morimura J, Iwasaki S: **The codon CUG is read as serine in an asporogenic yeast Candida cylindracea.** *Nature* 1989, **341(6238)**:164-166.
42. Ohama T, Suzuki T, Mori M, Osawa S, Ueda T, Watanabe K, Nakase T: **Non-universal decoding of the leucine codon CUG in several Candida species.** *Nucleic Acids Res* 1993, **21(17)**:4039-4045.
43. Santos MA, Tuite MF: **The CUG codon is decoded in vivo as serine and not leucine in Candida albicans.** *Nucleic Acids Res* 1995, **23(9)**:1481-1486.
44. Sugita T, Nakase T: **Nonuniversal usage of the leucine CUG codon in yeasts: Investigation of basidiomycetous yeast.** *J Gen Appl Microbiol* 1999, **45(4)**:193-197.
45. Massey SE, Moura G, Beltrao P, Almeida R, Garey JR, Tuite MF, Santos MA: **Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in Candida spp.** *Genome Res* 2003, **13(4)**:544-557.
46. Rodrigues de Miranda L: **Clavisporea, a new yeast genus of the Saccharomycetales.** *Antonie Van Leeuwenhoek* 1979, **45(3)**:479-483.
47. Wickerham LJ, Burton KA: **A clarification of the relationship of Candida guilliermondii to other yeasts by a study of their mating types.** *J Bacteriol* 1954, **68(5)**:594-597.
48. Young LY, Lorenz MC, Heitman J: **A STE12 homolog is required for mating but dispensable for filamentation in Candida lusitanae.** *Genetics* 2000, **155(1)**:17-29.
49. Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neugeglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisrarme A, Boyer J, Cattolico L, Confaniolieri F, De Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud JM, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard GF, Straub ML, Suleau A, Swennen D, Tekaiia F, Wesolowski-Louvel M, Westhof E, Wirth B, Zenioui-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron

- B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P, Souciet JL: **Genome evolution in yeasts.** *Nature* 2004, **430(6995)**:35-44.
50. Fabre E, Muller H, Therizols P, Lafontaine I, Dujon B, Fairhead C: **Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing, and subtelomeres.** *Mol Biol Evol* 2005, **22(4)**:856-873.
51. Hull CM, Johnson AD: **Identification of a mating type-like locus in the asexual pathogenic yeast *Candida albicans*.** *Science* 1999, **285(5431)**:1271-1275.
52. Hull CM, Raisner RM, Johnson AD: **Evidence for mating of the "asexual" yeast *Candida albicans* in a mammalian host.** *Science* 2000, **289(5477)**:307-310.
53. Pujol C, Daniels KJ, Lockhart SR, Srikantha T, Radke JB, Geiger J, Soll DR: **The closely related species *Candida albicans* and *Candida dubliniensis* can mate.** *Eukaryot Cell* 2004, **3(4)**:1015-1027.
54. Magee BB, Magee PT: **Induction of mating in *Candida albicans* by construction of MTL α and MTL α strains.** *Science* 2000, **289(5477)**:310-313.
55. Logue ME, Wong S, Wolfe KH, Butler G: **A genome sequence survey shows that the pathogenic yeast *Candida parapsilosis* has a defective MTL α allele at its mating type locus.** *Eukaryot Cell* 2005, **4(6)**:1009-1017.
56. Hamajima K, Nishikawa A, Shinoda T, Fukazawa Y: **Deoxyribonucleic acid base composition and its homology between two forms of *Candida parapsilosis* and *Lodderomyces elongisporus*.** *J Gen Appl Microbiol* 1987, **33**:299-302.
57. Nakase T, Komagata K, Fukazawa Y: **A comparative taxonomic study on two forms of *Candida parapsilosis*.** *J Gen Appl Microbiol* 1979:375-386.
58. Diezmann S, Cox CJ, Schonian G, Vilgalys RJ, Mitchell TG: **Phylogeny and evolution of medical species of *Candida* and related taxa: a multigenic analysis.** *J Clin Microbiol* 2004, **42(12)**:5624-5635.
59. James SA, Collins MD, Roberts IN: **The genetic relationship of *Lodderomyces elongisporus* to other ascomycete yeast species as revealed by small-subunit rRNA gene sequences.** *Lett Appl Microbiol* 1994, **19(5)**:308-311.
60. Sullivan DJ, Westerneng TJ, Haynes KA, Bennett DE, Coleman DC: ***Candida dubliniensis* sp. nov.: phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals.** *Microbiology* 1995, **141 (Pt 7)**:1507-1521.
61. Tzung KW, Williams RM, Scherer S, Federspiel N, Jones T, Hansen N, Bivolarevic V, Huizar E, Kemp C, Surzycki R, Tamse R, Davis RW, Agabian N: **Genomic evidence for a complete sexual cycle in *Candida albicans*.** *Proc Natl Acad Sci U S A* 2001, **98(6)**:3249-3253.
62. van der Walt JP: ***Lodderomyces*, a new genus of the Saccharomycetaceae.** *Antonie Van Leeuwenhoek* 1966, **32(1)**:1-5.
63. Daniel HM, Sorrell TC, Meyer W: **Partial sequence analysis of the actin gene and its potential for studying the phylogeny of *Candida* species and their teleomorphs.** *Int J Syst Evol Microbiol* 2001, **51(Pt 4)**:1593-1606.
64. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437(7057)**:376-380.
65. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** In *Nucleic Acids Res Volume 25*. Issue 17 ENGLAND ; 1997:3389-3402.
66. Korf I: **Gene finding in novel genomes.** *BMC Bioinformatics* 2004, **5**:59.
67. Majoros WH, Pertea M, Salzberg SL: **TigrScan and Glimmer-HMM: two open source ab initio eukaryotic gene-finders.** *Bioinformatics* 2004, **20(16)**:2878-2879.
68. Slater GS, Birney E: **Automated generation of heuristics for biological sequence comparison.** *BMC Bioinformatics* 2005, **6(1)**:31.
69. Birney E, Clamp M, Durbin R: **GeneWise and Genomewise.** *Genome Res* 2004, **14(5)**:988-995.
70. Annotations: <http://fungal.genome.duke.edu>.
71. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** In *Nucleic Acids Res Volume 22*. Issue 22 ENGLAND ; 1994:4673-4680.
72. Castresana J: **Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis.** *Mol Biol Evol* 2000, **17(4)**:540-552.
73. Archie JW: **A randomization test for phylogenetic information in systematic data.** *Systematic Zoology* 1989, **38**:251-278.
74. Faith DP, Cranston PS: **Could a cladogram this short have arisen by chance alone? On permutation tests for cladistic structure.** *Cladistics* 1991, **7**:1-28.
75. MultiPhyl www.cs.may.ie/distributed/multiphyl.php.
76. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4(4)**:406-425.
77. Lockhart PJ, Steel MA, Hendy MD, Penny D: **Recovering Evolutionary Trees under a More Realistic Model of Sequence.** *Mol Biol Evol* 1994, **11(4)**:605-612.
78. Creevey CJ, McLnerney JO: **Clann: investigating phylogenetic information through supertree analyses.** *Bioinformatics* 2005, **21(3)**:390-392.
79. Pearson WR, Lipman DJ: **Improved tools for biological sequence comparison.** *Proc Natl Acad Sci U S A* 1988, **85(8)**:2444-2448.
80. Byrne KP, Wolfe KH: **Visualizing syntenic relationships among the hemiascomycetes with the Yeast Gene Order Browser.** *Nucleic Acids Res* 2006, **34(Database issue)**:D452-5.
81. Byrne KP, Wolfe KH: **The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species.** *Genome Res* 2005, **15(10)**:1456-1461.
82. Bryant D, Moulton V: **Neighbor-net: an agglomerative method for the construction of phylogenetic networks.** *Mol Biol Evol* 2004, **21(2)**:255-265.
83. Huson DH: **SplitsTree: analyzing and visualizing evolutionary data.** *Bioinformatics* 1998, **14(1)**:68-73.
84. Charleston MA: **Spectrum: spectral analysis of phylogenetic data.** *Bioinformatics* 1998, **14(1)**:98-99.
85. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McLnerney JO: **Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified.** *BMC Evol Biol* 2006, **6**:29.
86. Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB, Newport G, Thorstenson YR, Agabian N, Magee PT, Davis RW, Scherer S: **The diploid genome sequence of *Candida albicans*.** *Proc Natl Acad Sci U S A* 2004, **101(19)**:7329-7334.
87. Cliften P, Sudarsanam P, Desikan A, Fulton L, Fulton B, Majors J, Waterston R, Cohen BA, Johnston M: **Finding functional features in Saccharomyces genomes by phylogenetic footprinting.** *Science* 2003, **301(5629)**:71-76.
88. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES: **Sequencing and comparison of yeast species to identify genes and regulatory elements.** *Nature* 2003, **423(6937)**:241-254.
89. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG: **Life with 6000 genes.** *Science* 1996, **274(5287)**:546, 563-7.
90. Kellis M, Birren BW, Lander ES: **Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*.** *Nature* 2004, **428(6983)**:617-624.
91. Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, Mohr C, Pohlmann R, Luedi P, Choi S, Wing RA, Flavier A, Gaffney TD, Philippsen P: **The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome.** *Science* 2004, **304(5668)**:304-307.
92. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, Read ND, Lee YH, Carbone I, Brown D, Oh YY, Donofrio N, Jeong JS, Soanes DM, Djonovic S, Kolomiets E, Rehmeier C, Li W, Harding M, Kim S, Lebrun MH, Bohnert C, Coughlan S, Butler J, Calvo S, Ma LJ, Nicol R, Purcell S, Nusbaum C, Galagan JE, Birren BW: **The genome sequence of the rice blast fungus *Magnaporthe grisea*.** *Nature* 2005, **434(7036)**:980-986.
93. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, Rehman B, Elkins T, Engels R,

- Wang S, Nielsen CB, Butler J, Endrizzi M, Qui D, Ianakiev P, Bell-Pedersen D, Nelson MA, Werner-Washburne M, Selitrennikoff CP, Kinsey JA, Braun EL, Zelter A, Schulte U, Kothe GO, 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, Kamysysselis M, Mauceli E, Bielke C, Rudd S, Frishman D, Krystofova S, Rasmussen C, Metznerberg RL, Perkins DD, Kroken S, Cogoni C, Macino G, Catcheside D, Li W, Pratt RJ, Osmani SA, DeSouza CP, Glass L, Orbach MJ, Berglund JA, Voelker R, Yarden O, Plamann M, Seiler S, Dunlap J, Radford A, Aramayo R, Natvig DO, Alex LA, Mannhaupt G, Ebbole DJ, Freitag M, Paulsen I, Sachs MS, Lander ES, Nusbaum C, Birren B: **The genome sequence of the filamentous fungus *Neurospora crassa***. *Nature* 2003, **422(6934)**:859-868.
94. Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D, Bowman S, Brooks K, Brown D, Brown S, Chillingworth T, Churcher C, Collins M, Connor R, Cronin A, Davis P, Feltwell T, Fraser A, Gentles S, Goble A, Hamlin N, Harris D, Hidalgo J, Hodgson G, Holroyd S, Hornsby T, Howarth S, Huckle EJ, Hunt S, Jagels K, James K, Jones L, Jones M, Leather S, McDonald S, McLean J, Mooney P, Moule S, Mungall K, Murphy L, Niblett D, Odell C, Oliver K, O'Neil S, Pearson D, Quail MA, Rabbinowitsch E, Rutherford K, Rutter S, Saunders D, Seeger K, Sharp S, Skelton J, Simmonds M, Squares R, Squares S, Stevens K, Taylor K, Taylor RG, Tivey A, Walsh S, Warren T, Whitehead S, Woodward J, Volckaert G, Aert R, Robben J, Grymonprez B, Weltjens I, Vanstreels E, Rieger M, Schafer M, Muller-Auer S, Gabel C, Fuchs M, Dusterhoft A, Fritz C, Holzer E, Moestl D, Hilbert H, Borzym K, Langer I, Beck A, Lehrach H, Reinhardt R, Pohl TM, Eger P, Zimmermann W, Wedler H, Wambutt R, Purnelle B, Goffeau A, Cadieu E, Dreano S, Gloux S, Lelaure V, Mottier S, Galibert F, Aves SJ, Xiang Z, Hunt C, Moore K, Hurst SM, Lucas M, Rochet M, Gaillardin C, Tallada VA, Garzon A, Thode G, Daga RR, Cruzado L, Jimenez J, Sanchez M, del Rey F, Benito J, Dominguez A, Revuelta JL, Moreno S, Armstrong J, Forsburg SL, Cerutti L, Lowe T, McCombie WR, Paulsen I, Potashkin J, Shpakovski GV, Ussery D, Barrell BG, Nurse P: **The genome sequence of *Schizosaccharomyces pombe***. *Nature* 2002, **415(6874)**:871-880.
95. Martinez D, Larrondo LF, Putnam N, Gelpke MD, Huang K, Chapman J, Helfenbein KG, Ramaiya P, Detter JC, Larimer F, Coutinho PM, Henrissat B, Berka R, Cullen D, Rokhsar D: **Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78**. *Nat Biotechnol* 2004, **22(6)**:695-700.
96. Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, Vamathevan J, Miranda M, Anderson IJ, Fraser JA, Allen JE, Bosdet IE, Brent MR, Chiu R, Doering TL, Donlin MJ, D'Souza CA, Fox DS, Grinberg V, Fu J, Fukushima M, Haas BJ, Huang JC, Janbon G, Jones SJ, Koo HL, Krzywinski MI, Kwon-Chung JK, Lengeler KB, Maiti R, Marra MA, Marra RE, Mathewson CA, Mitchell TG, Pertea M, Riggs FR, Salzberg SL, Schein JE, Shvartsbeyn A, Shin H, Shumway M, Specht CA, Suh BB, Tenney A, Utterback TR, Wickes BL, Wortman JR, Wye NH, Kronstad JW, Lodge JK, Heitman J, Davis RW, Fraser CM, Hyman RW: **The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans***. *Science* 2005, **307(5713)**:1321-1324.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

