Available online at www.sciencedirect.com
ScienceDirect
Fungal Genetics and Biology 44 (2007) 357–367
FUNGAL GENETICS AND BIOLOGY
www.elsevier.com/locate/yfgbi

Background selection at the chitin synthase II (*chs2*) locus in *Paracoccidioides brasiliensis* species complex

Daniel R. Matute [a,*,1], Isaura P. Torres [a], Catalina Salgado-Salazar [a,b], Angela Restrepo [a], Juan G. McEwen [a,b]

[a] *Corporación para Investigaciones Biológicas (CIB), Medellín, Colombia*
[b] *Facultad de Medicina, Universidad de Antioquia, Medellín, Colombia*

Received 8 June 2006; accepted 9 January 2007
Available online 19 January 2007

Abstract

In fungi, chitin synthases have been classified into five classes according to differences in regions of high sequence conservation. The current investigation was initiated to examine the causes for the polymorphism patterns found in a class II chitin synthase gene (*chs2*) of *Paracoccidioides brasiliensis*, in an attempt to determine the evolutionary forces affecting the chitin synthesis metabolic pathway. Neutrality tests were applied to the *chs2* sequences exhibited by *P. brasiliensis* species complex. According to these tests and based on non-synonymous differences, *P. brasiliensis* data rejected the null hypothesis for a pure drift mutational process owing to a large excess of unique polymorphisms. In contrast, the synonymous and intron site differences did not reject the null hypothesis. This pattern appears consistent with weak selection against most amino acid changes, in which the effect of background selection was not detectable at synonymous nor at intron sites.

Keywords: *Paracoccidioides brasiliensis*; Background selection; Positive selection; Chitin synthase; Neutrality tests; Non-synonymous differences

1. Introduction

Chitin, the β-1-4-linked polymer of *N*-acetylglucosamine, is a fibrous cellulose-like polysaccharide that serves as the major structural component of the cell wall in many species of fungi and as the exoskeleton in arthropods (Ruiz-Herrera et al., 2002). In many species of yeasts, chitin is used to maintain the structure of the junction between the mother cell and the bud, whereas in filamentous fungi chitin is usually the major supporting component of the cell wall (Georgopapadakou and Tkacz, 1995; Schmidt et al., 2003).

Paracoccidioides brasiliensis, a thermally dimorphic fungal pathogen, is the etiological agent of paracoccidioidomycosis (PCM), a systemic disorder restricted geographically to Central and South America and one of the most important mycoses in the area, especially for the male rural population. It has been estimated that a total of approximately 10 million individuals have been infected. Furthermore, skin reactivity tests in healthy residents of certain endemic areas have shown the presence of subclinical infections in a large proportion of rural inhabitants (Brummer et al., 1993; Restrepo, 2003). Despite its clinical importance in these geographical areas, the epidemiology, population and evolutionary studies of this fungus have been scarce. Previous data indicate that this fungus consists of at least three distinct species. Two of those genetically isolated groups are phylogenetic species: PS2, composed of Brazilian and Venezuelan isolates and PS3, restricted to Colombia while S1, is a paraphyletic species with a wide habitat range across all the localities where PCM has been reported, except for Colombia (Matute et al., 2006).

In *P. brasiliensis*, chitin represents as much as 43% of the dry weight of the wall of the pathogenic yeast form and

* Corresponding author. Fax: +57 4441 5514.
E-mail address: dmatute@uchicago.edu (D.R. Matute).
[1] Current address: Department of Ecology and Evolution, University of Chicago, 5801 South Ellis, Chicago, IL 60637, USA.

1087-1845/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.fgb.2007.01.006

13% of the hyphal cell wall (San-Blas, 1985). The absence of this polysaccharide from human cells makes its biosynthesis a logical target for the development of antifungal antibiotics.

Chitin synthases have been classified into five classes based on differences in regions of high sequence conservation (Ruiz-Herrera et al., 2002). The functional implications of these various classes of isoenzymes are not yet clear. In *P. brasiliensis*, a gene encoding for a class II chitin synthase (*chs2*) has been reported (Nino-Vega et al., 1998) becoming the most studied gene of the chitin synthase synthesis pathway (Nino-Vega et al., 1998; Nino-Vega et al., 2000). Despite the fact that yeast cells contain more chitin than hyphae, the levels of mRNA for *CHS2* are higher in the latter than in the former. This supports observations in other fungi in which transcription levels do not often correlate with chitin content indicating that post-transcriptional regulation of *chs2* gene expression is important for morphogenesis (Nino-Vega et al., 2000).

The current investigation was aimed at determining the causes of the patterns of polymorphisms found in the *chs2* gene in *P. brasiliensis* and defining the evolutionary forces affecting the chitin synthesis metabolic pathway. The estimation of gene genealogies from DNA sequencing drawn from species or populations provides a powerful approach to quantify mutation, selection, genetic random drift, migration, and recombination.

Several types of selection have been described for molecular sequences. Negative selection is defined as any type of selection where new mutations are selected against (Li, 1997). Background selection is a particular type of negative selection that acts on slightly deleterious mutations and causes a reduction in the amount of genetic variability at linked neutral sites (Charlesworth, 1994; Charlesworth et al., 1993; Hudson, 1994).

Positive selection, on the other hand, is any type of selection where new mutations are advantageous. Positive directional selection occurs when successive amino acid changes make a protein better adapted in a particular biological context, so the changes will tend to be fixed (Storz, 2005; Vallender and Lahn, 2004). Positive diversifying selection is the natural selection strategy by which multiple phenotypes in a population are favored, resulting in an overall increase in genetic diversity within the species (Li, 1997).

The central result of this study indicates that *P. brasiliensis* data rejected the null hypothesis of a pure drift mutation process, owing to a large excess of unique polymorphisms in non-synonymous sites. In contrast, the synonymous and intron site differences did not reject the null hypothesis. Moreover, the positive selection hypothesis was discarded for the coding region by the evaluation of nested likelihood models. This pattern appears to be consistent with weak selection against most amino acid changes, in which the effect of background selection is not detectable at synonymous or intron sites.

2. Materials and methods

2.1. P. brasiliensis isolates and growth conditions

Sixty-five of the 67 *P. brasiliensis* individuals used in this study have been described previously (Matute et al., 2006). The entire sample encompassed individuals from the three described *P. brasiliensis* species: S1 ($n = 39$), PS2 ($n = 6$) and PS3 ($n = 22$) and included six different paracoccidiodomycosis endemic areas (Supplementary Table 1). The samples analyzed have been previously deposited in GenBank, Accession Nos. DQ004114 to DQ004243. The two newly reported sequences have been deposited in GenBank under Accession Nos. DQ640004 and DQ640005. The corresponding cultures were grown as reported previously (Diez et al., 1999). Total DNA was extracted from the yeast culture with protocols using glass beads (Van Burik et al., 1998) or maceration of frozen cells (Morais et al., 2000).

2.2. PCR and sequencing conditions

OLIGO 4.0 (National Biosystems, Plymouth, Minn.) was used to design the oligonucleotide primers (Operon Technologies Inc., Alameda, CA) selecting two regions of the *chs2* locus: the promotor and the exon 1 were amplified with 5′-GCAACTCTACCCCAGAGCAAGC-3′ and 5′-GGGTTCCAGGTCTTGGGCTAC-3′ and the exon 2 to the exon 4 region was amplified with the primers: 5′-CTTAACGGTGCCTTCTTTGCGG-3′ and 5′-GTGAAAGTATTGTTGCCCAGCG-3′. The other amplified regions were the Exon 2 of glucan synthase (5′-TCCGAGTACGAGAGAGTCCGC-3′ and 5′-CCATCACCGATCCTATTTTGCC-3′), exon 3 from β-glucan synthase (5′-TGGCATTCGATGTCTCAGC-3′ and 5′-CTAATAGTCTTCCCCCAAAC-3′) (DQ003855–DQ003983), exons 2–3 from adenyl ribosylation factor (5′-TCTCATGGTTGGCCTCGATGCTGCC-3′ and 5′-GAGCCTCGACGACACGGTCACGATC-3′) (DQ004049–DQ004113).

The cyclic reaction termination method was done with the Big-Dye Terminator Cycle Sequencing Kits (Applied Biosystems, Foster City, CA) in order to determine the nucleotide sequences. Sequence data were collected from both strands and examined with Sequence Navigator v. 1.0.1 (Applied Biosystems). The sequences were aligned manually, and coding regions were assigned by visual inspection.

2.3. Substitution rate estimation

Seventy-five sequences composed of 18S ribosomal RNA gene, internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2 were downloaded from GenBank. The individuals and the accession numbers are registered in Supplementary Table 2. The sequences were aligned using ClustalW (Thompson et al., 1994).

The most likely model of substitutions for this dataset was determined by performing a series of likelihood ratio tests of different models using Modeltest version 3.04 (Posada and Crandall, 1998). Maximum Parsimony and Maximum likelihood (Parmley et al., 1994) estimates of tree topologies were obtained with PAUP* version 4.0b (Swofford, 2002). Trees were constructed from the data enforcing molecular clock and without enforcing. In all ML analyses, base frequencies and the transition/transversion ratio from the data were estimated; to account for rate heterogeneity among sites, we used a gamma distribution (Yang, 1993) estimating the alpha shape parameter from the data. Likelihood values obtained were used to test heterogeneity rates across lineages by using the likelihood ratio (LR) test (Felsenstein, 1988). Significance was assessed by comparing $\Lambda_{\mathrm{TREE}} = -2 * \mathrm{LR}$, where LR is the difference between the −ln likelihood of the tree, with and without enforcing a molecular clock, with a χ^2 distribution (with $n-2$ degrees of freedom, where n is the number of taxa).

The dates for the calibration points used in this study were based on the Taylor et al. (2005) scale of time and Redecker et al. (2000). Ages of clades were minimum ages estimated conservatively for *P. brasiliensis* species complex by the first appearance of fossils clearly referable to one of the 18S ribosomal RNA gene tree included lineages based on morphological synapomorphies.

Because tests for heterogeneity rates among lineages were highly significant, we dated the nodes by using the non-parametric rate smoothing (NPRS) method of Sanderson (1997). The 18S ribosomal RNA gene tree with branch lengths was transformed into ultrametric trees by using the NPRS method implemented in the software TreeEdit version 1.0 (http://evolve.zoo.ox.ac.uk/software/TreeEdit/main.html). To transform relative time to absolute ages, we calibrated the trees by using dates from the fossil record. We reapplied the NPRS procedure to 100 bootstrapped matrices obtained by re-sampling the data using PAUP, in order to compute error estimates for the ages inferred.

All distances were calculated using MEGA (Kumar et al., 1994). For coding regions, rates of nucleotide substitution were estimated by Nei and Gojobori distance measures (Nei and Gojobori, 1986), as well as by fossil estimates. To reduce the variance in absolute rate estimates, information from all three sequences was used to estimate substitution rates. We computed the average rate of *chs2* non-synonymous and synonymous substitution since time T*i* according to the method proposed by Muse and Gaut (1994).

The flanking sequences of the microsatellite 11*B*12*B* (*FS11B12B*: DQ673992–DQ674054) and the flanking sequences of microsatellite *KL* (*FSKL*: DQ674181–DQ674243) from *P. brasiliensis* were obtained from GenBank and used to estimate the effective population size (N*e*).

2.4. Estimates of recombination

Analysis of polymorphisms and estimation of recombination rates in the *chs2* locus, were determined with the DNASP program (Rozas et al., 2003). In those cases where no incompatible sites were observed, the recombination rate was zero. Parsimony analyses of polymorphic sites were also performed. The observation of a consistency index of one (homoplasy index equals zero) for the best tree was synonymous with a lack of recombination. The informative Sites Index was also implemented using the program PIST (http://evolve.zoo.ox.ac.uk/software/PIST/PIST.html) (Worobey, 2001).

2.5. Codon-based likelihood analyses

Several likelihood-based tests were used to search for evidence of positive selection using the CODEML program of the PAML version 3.15 package (Yang, 1997; Yang and Bielawski, 2000). For each model, equilibrium codon frequencies were estimated from the average nucleotide frequencies at each codon position, amino acid distances were assumed to be equal, and the transition/transversion ratio (κ) was estimated from the data. For all other parameters, we used the default settings (Yang and Bielawski, 2000). Given the observed intraspecific variability and the lack of homoplasy found in individual gene trees, we could assume linkage between collinear sites (i.e., no recombination).

To verify which of the models best fit the data, likelihood ratio tests (Λ) were performed by comparing twice the differences in log-likelihood values ($\Lambda = -2 * \mathrm{LR}$) between two models using a χ^2 distribution, with the number of degrees of freedom equal to the difference in the number of parameters between the models.

To test variation in selection on different branches, we used six models implemented in PAML (Nielsen and Yang, 1998; Yang, 1997; Yang and Bielawski, 2000). A simple model that assumes a constant ω across all lineages (one-ratio model, M0) was compared with a more general model that assumes an independent ω for each branch in the phylogeny (free-ratio model, M3). The one-ratio model (M0) assumes one ω for all sites. The discrete model (M3) uses a discrete distribution with three site classes, with the proportions (p0, p1, and p2) and the ω ratios (ω0, ω1, and ω2) estimated from the data. Although M3 model is parameter rich and as such, unlikely to produce accurate ω estimates for all branches (Swanson and Yang, 1998), it is nonetheless useful for identifying lineages in which episodes of positive selection might have occurred. Models of variable ω among sites were used to test for the presence of sites under positive selection ($\omega > 1$) and to identify them. Two of the models were used as null hypotheses for neutrality: the neutral model and the beta model. The neutral model (M1) assumes two classes of sites in the protein, the conserved sites at which $\omega = 0$ and the neutral sites that are defined by a $\omega = 1$. The beta model (M7) uses a β distribution of ω over sites: $\beta(p, q)$, which, depending on parameters p and q,

can take various shapes in the 0 to 1 interval. Two more models allow sites with $\omega > 1$ and can be considered as tests of positive selection. The selection model (M2) has an additional class of sites compared to the neutral model, in which ω is a free parameter and, as such, can change among residues. The beta and ω model (M8) added an extra class of sites to the beta model, estimating the proportion of ω from the data. We used the Λ values of each model to make 3 comparisons: to test for independence of ω for each branch, the one-ratio model (M0) was compared with the discrete model (M3), positive selection; the neutral model (M1) was compared with the selection model (M2); and the beta model (M7) was compared with the beta and ω model (M8) using 4, 2, and 2 degrees of freedom, respectively (Yang and Bielawski, 2000).

Finally, we identified particular sites in the genes that were likely to have evolved under positive selection. This was accomplished using an empirical Bayesian approach previously proposed (Nielsen and Yang, 1998). Unknown parameters in Bayes' equation are first estimated from the data using the likelihood function as applied in the discrete model (Nssites 3). Once these parameters had been estimated, Bayes' theorem was used to estimate the posterior probability that a given site came from the class of positively selected sites (Nielsen and Yang, 1998; Yang and Bielawski, 2000; Yang et al., 2005).

2.6. Sequence diversity tests of neutrality

A number of tests have been developed to determine significant departures from neutral evolution of sequenced data. Tajima's statistic test represented by T (Tajima, 1989), Fu's statistic test (Fu, 1996), F, Fu's and Li's tests, D^* and F^*, respectively (Fu and Li, 1993), were calculated. The expected difference between all these estimates should be zero for a neutral drift-mutation process at equilibrium. Tests of neutrality and determination of their associated significance were done by using the programs by Fu (http://www.hgc.sph.uth.tmc.edu/fu/genealogy/test2/welcome.html) (Fu, 1997). The tests for neutrality for *FS11B12B* and *FSKL* loci were conducted with DNASP (Rozas et al., 2003).

3. Results

3.1. Nucleotide sequence polymorphism

The DNA sequence for the *chs2* locus was determined for all 67 samples (Table 1). The region sequenced was 1162 bp in length, including 828 bp of exon sequence (part of exon 1 and 2, all of exon 3, and part of exon 4) and 334 bp of intron sequence (all of introns 2 and 3). There were 21 sites polymorphic for nucleotide substitutions, 16 in exons and 5 in introns (Table 2). Neither insertion-deletion events nor polymorphisms in the 5′UTR region (300pb) were observed in the sequences studied. There was little apparent bias towards transitions, as 12 transitions and 8 transversions were observed. Of the 16 nucleotide substitutions located in the coding regions, 5 of them were synonymous and 11 non-synonymous. The frequency of each polymorphism is listed in Table 2. From the mutations located in the *chs2* exons, 7 of the non-synonymous substitutions were non-conservative with regards to the simple biochemical classification of amino acids (i.e., polar, nonpolar, acidic, basic, aromatic, and cysteine).

The most appropriate evolutionary model for the *chs2* was GTR + I, with a log-likelihood value of −1767.19. The GTR + I + G model had the same likelihood value but the simplest model was preferred because of computational time. The topology obtained by maximum parsimony and maximum likelihood was the same and is represented in Fig. 1. The nucleotide substitutions defined 19 haplotypes (Fig. 1 and Table 1). No more than 5 mutations separated any two haplotypes.

3.2. Substitution rate estimation

From the analysis of the 18S region the tree represented in Fig. 2 was obtained. The use of two calibration points had little impact on the estimates of the age of nodes. The first estimates had been proposed by Taylor et al. (2005) in which the divergence of Basidiomycetous and Ascomycetous fungi was placed at 500 million years ago (mya). In this case, the divergence date for *P. brasiliensis* PS2 and PS3 was 8.04 mya (SD = 1.17). The second date used to calibrate the tree was the Glomalean fungal fossil 600 mya reported by Redecker et al. (2000). When using the latter date, the divergence was 8.37 mya (SD = 0.98). No significant differences were found between the two dates ($t = 1.9591$, $p = 0.0515$).

The GB distance between S1 and PS2 for synonymous substitutions in the *chs2* locus was 1.244 (SE = 0.733) and for non-synonymous 0.795 (SE = 0.318). For synonymous substitutions the GB distance between PS3 and S1 was 0.706 (SE = 0.594) and for the non-synonymous substitutions was 0.578 (SE = 0.203). These calculations made possible the application of the method proposed by Muse and Gaut (1994) for the calculation of the substitution rate for *chs2*, which was equal to 4.27×10^{-8} non-synonymous substitutions per locus per year and 6.22×10^{-8} synonymous substitutions per locus per year.

The Jukes Cantor distance for *FS11B12B* locus between PS2 and S1 was equal to 0.012 (SE = 0.004) and between S1 and PS3 was 0.019 (SE = 0.008). In the *FS56* case the distance between S1 and PS2 equaled 0.017 (SE = 0.007) and between S1 and PS3 was 0.005 (SE = 0.003). The substitution rate for *FS11B12B* equaled 9.259×10^{-9} per locus per year and for *FS56* was 6.57×10^{-10} per locus per year.

Paracoccidioides brasiliensis is a haploid organism (Almeida et al., 2007) in which $\theta = 2 * \mu * Ne$ (Rannala, 1996; Rannala and Hartigan, 1996), where θ is the estimator of nucleotidic diversity, μ is the mutation rate and Ne is the effective population size of the population. The Ne was established from θ calculated according to the

Table 1
Distribution of nucleotide polymorphisms in the *chs2* gene loci of *P. brasiliensis* species complex

Position	27	30	72	122	143	144	333	1963	1972	1976	2133	2240	2286	2289	2361	2369	2411	2424	2469	2473
	E	E	E	E	E	E	E	E	E	E	I	E	E	E	I	I	I	E	E	E
Amino acids																				
Consensus	D	S	L	Q	S	L	E	F	A	Y		D	Y	T				M	T	V
Substitutions	E	P	F	.	.	V	Q	L	.	D		.	H	P				R	N	.
Nucleic acids																				
B1	T	T	C	A	T	C	G	T	T	T	C	T	T	A	G	A	A	T	C	C
B2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
B4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
B5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
B6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	G	.	.
B7	.	.	.	G	A	.	.	.	.	.	.	C	.	C	.	.	T	G	.	.
B8	.	.	.	.	.	.	.	.	.	.	.	C	C	.	.	.	.	.	.	.
B9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B11	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
B12	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B13	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
B14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B15	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
B16	.	.	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.
B17	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B18	.	.	.	.	.	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.
B19	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B21	.	.	.	.	.	G	.	.	.	.	.	C	.	.	.	G	.	.	.	.
B22	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B23	.	.	.	G	A	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
B24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B25	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
B26	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B27	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.
C1	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C2	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C3	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C4	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C5	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C6	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	G
C7	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C8	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C9	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C10	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C11	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C12	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C13	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C14	.	.	.	.	.	.	.	.	G	.	T	C	.	.	.	.	.	.	.	.
C15	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C16	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C17	.	.	.	.	.	.	.	.	.	G	T	C	.	.	.	.	.	.	.	.
C18	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C19	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C20	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C21	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
C22	G	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.
V1	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
V2	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
V3	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
V4	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
V5	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
V6	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
U1	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.

(*continued on next page*)

Table 1 (*continued*)

Position	27	30	72	122	143	144	333	1963	1972	1976	2133	2240	2286	2289	2361	2369	2411	2424	2469	2473
PE1	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	.	.
A1	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	.	.
A2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A6	.	.	.	.	.	G	.	.	.	.	.	C	.	.	.	G	.	.	.	.
A7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
P1	.	.	.	.	.	G	.	.	.	.	.	C	.	.	.	G	.	.	.	.
P2	.	.	.	.	.	G	.	.	.	.	.	C	.	.	.	G	.	.	.	.

The numbers indicate positions in the reference sequences; dots indicate nucleotides or amino acids identical to the consensus sequence. N, Non-synonymous; S, Synonymous; R, transversion; T, transition.

Table 2
Polymorphism characteristics in the sample of *chs2* from *P. brasiliensis* species complex

	Position	Frequency	Change	R or T	Base	Amino acid	Class	Base	Amino acid	Class
Exon regions										
a	27	1/67	N	R	T	Asp	Acidic	G	Glu	Acidic
b	30	3/67	N	T	T	Ser	Polar	C	Pro	Iminoacid
c	72	5/67	N	T	C	Leu	Nonpolar	T	Phe	Aromatic
d	122	2/67	S	T	A	Gln	—	G	Gln	—
e	143	2/67	S	R	T	Ser	—	A	Ser	—
f	144	4/67	N	R	C	Leu	Nonpolar	G	Val	Nonpolar
g	333	1/67	N	R	G	Glu	Acidic	C	Gln	Polar
h	1963	1/67	N	R	T	Phe	Aromatic	G	Leu	Nonpolar
i	1972	1/67	S	R	T	Ala	—	G	Ala	—
j	1976	1/67	N	R	T	Tyr	Aromatic	G	Asp	Acidic
l	2240	21/67	S	T	C	Asp	—	T	Asp	—
m	2286	1/67	N	T	T	Tyr	Aromatic	C	His	Aromatic
n	2289	1/67	N	R	A	Thr	Polar	C	Pro	Iminoacid
r	2424	4/67	N	R	T	Met	Nonpolar	G	Arg	Basic
s	2469	1/67	N	R	C	Thr	Polar	A	Asn	Polar
t	2473	1/67	S	R	C	Val	Nonpolar	G	Val	Nonpolar
Intron regions										
k	2133	21/67		T	C	—		T	—	
o	2361	1/67		T	G	—		A	—	
p	2369	6/67		T	A	—		G	—	
q	2411	2/67		R	A	—		T	—	

R, transitions; T, transversitions.

Tajima parameters for each one of the species and the estimated mutation rate. In our calculations μ was obtained from absolute time, rather than from the number of generations. The generation time of *P. brasiliensis* has not been measured in nature; therefore, to convert mutation rates from number of mutations per locus per year to number of mutations per locus per generation we used a constant, *G*, that represents the number of generations per year. The N*e* estimates for each species were also calculated with the loci that showed no selective pressures and a neutral evolutionary mode. The estimated effective size of each species, proportional to *G*, is presented in Table 3.

3.3. Estimates of recombination

One implication of the low polymorphism reported for the *chs2* is that the apparent rate of recombination is much reduced in comparison with organisms in which the degree of polymorphism is higher and recombination is detectable. The estimate of the recombination parameter in *chs2* (Rm: Minimal number of recombination events) for each one of the species and for the species complex was equal to 0, showing the absence of recombination in *chs2*. Similarly, the IST results showed no significant deviation for any of the species from the null hypothesis of clonality, because in the three species equaled zero. The only strain that presented recombination was B17 (Fig. 1).

3.4. Positive selection tests: Codon-based likelihood analyses

Log-likelihood values and parameter estimates under each model are listed in Table 4. Selection models did not provide a better fit to the data; comparisons of M2 neutral model yielded Λ values of 0.0 ($df = 2$, $P < 0.0001$). Likewise,

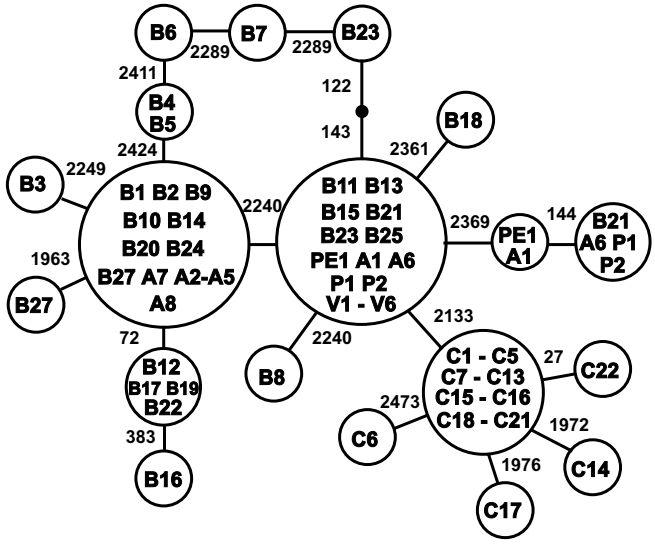

Fig. 1. Minimum spanning tree showing *chs2* haplotype relationships in *P. brasiliensis* species complex. Letters in adjacent branches represent specific polymorphisms. The black dots represent non-sampled haplotypes.

tests between beta (M7) and beta and ω (M8) models did not support positive selection ($\Lambda = 0.0$, $P < 0.0001$). The free-ratio model (M3) was compared with a model that assumes a constant ω across all lineages (M0) by performing a Λ test with 6 degrees of freedom, and the model assuming a constant rate of ω across all lineages (M0) could not be rejected for *chs2* ($\Lambda = 18.64$, $P < 0.0001$). The ω value (0.798) shows a slight non-significant departure from neutral hypothesis towards weak negative selection. No evidence of ω variation among lineages or among sites was found.

3.5. Sequence diversity tests for neutrality

Neutrality tests were applied to the *chs2* sequences of the *P. brasiliensis* species complex and their associated significance levels determined by Fu's program (Fu, 1997) as summarized in Table 5. Partitioning the data into groups (e.g., exon, intron, synonymous, non-synonymous) allowed comparison of the evolutionary dynamics between groups. When the entire data set was taken into account, the tests always gave significant results. For the *chs2* exon sequence data, all tests revealed statistically significant values, while none of them were significant for the intron sequence data. Further partitioning of the data into synonymous and non-synonymous sites showed statistically significant values for the latter sites but not for former. The distribution of synonymous and intron polymorphisms thus conform to neutral expectations. Taken together, it appears that the excess of singleton amino acid replacements at the non-synonymous sites is responsible for the significant statistical values at higher levels of site classification.

The other 4 loci were not affected by selection and all of them were evolving under the neutral expectations (Table 4).

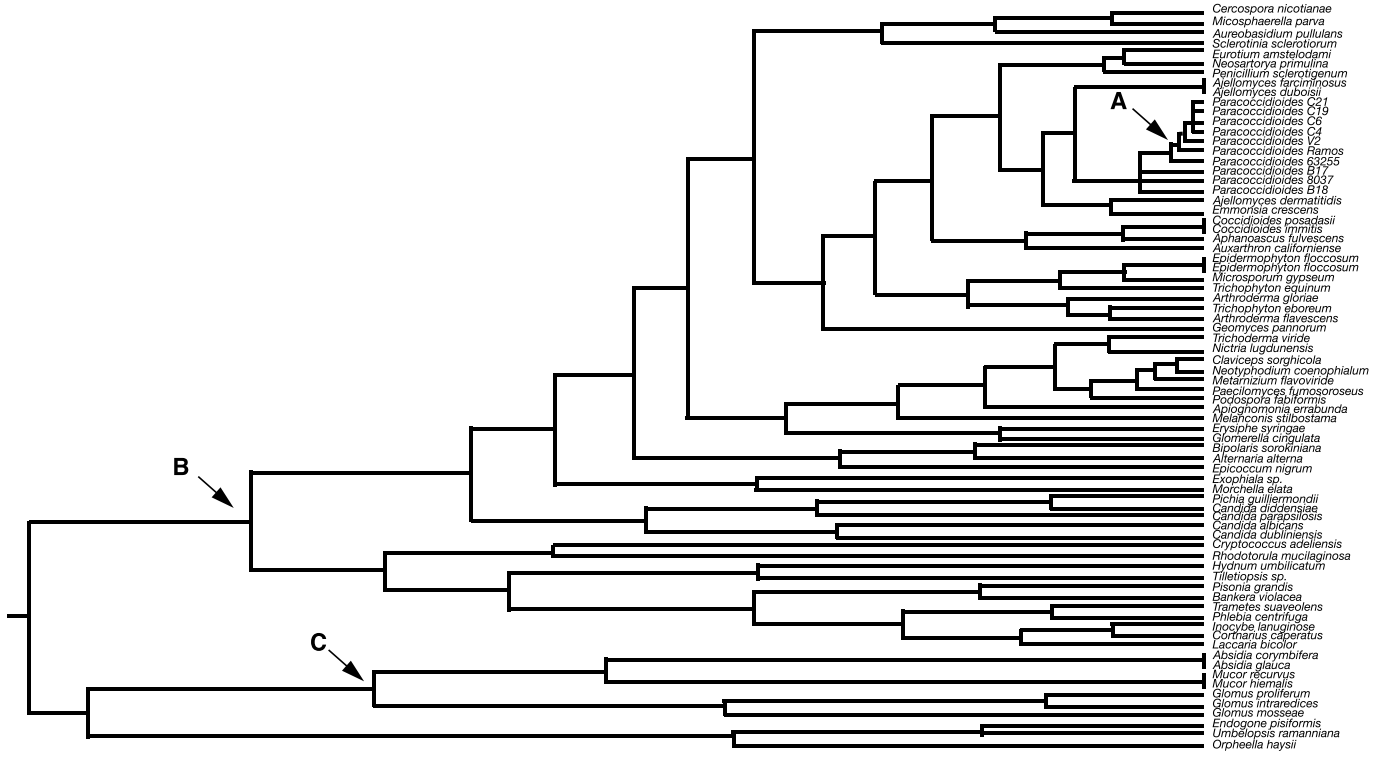

Fig. 2. Maximum likelihood tree of the 18S ribosomal unit of fungi, rooted at Zygomecetes and *Glomus* sp. sequences, based on nucleotide sequences. Sequences were obtained from the Gen Bank and their numbers are specified in Supplementary Table 1. The dates for the calibration points on the time scale of Taylor et al. (arrow A) and Redecker et al. (arrow B) are marked with an arrow. Ages of clades are estimated conservatively for *P. brasiliensis* species complex (arrow C).

Table 3
Estimates of effective population size (N*e*) and recombination using synonymous and non-synonymous sites from the studied regions of *chs2*

	N*e*/G		Rm	IST
	Synonymous	Non-synonymous		
P. brasiliensis sensu lato	43442	43442	0	0
S1	9405	36182	0	0
PS2	41237	16510	0	0
PS3	21623	10304	0	0
Σ	115707	106438	0	0

G, number of generations per year.

4. Discussion

The central result of this study concerns the *chs2* data whereby the null hypothesis of a pure drift mutation process was rejected on the bases of the available statistical tests employed. Moreover, the substitution rate for this gene was calculated from previous 18S ribosomal DNA sequences.

Many sources of error, as well as bias, can affect molecular-based estimates of divergence times. Obviously, an incorrect topology will yield erroneous estimates although the magnitude of the problem depends on the extent of the topological error. Likewise, inaccurate calibration will bias the resulting estimates for other divergences. Equally serious, however, are the well known heterogeneous rates of evolution among lineages, and a failure to recognize such heterogeneity can compromise resulting estimates of divergence times (Bromham and Penny, 2003). A further potential source of error is the method used to estimate divergence dates. As the phylogenetic tree that we obtained clearly portrayed a violation of the molecular clock, an alternative method designed to accommodate rate inconstancy was used; in this way we managed the problem

Table 4
Likelihood values, parameter estimates, and sites under positive selection as inferred under the six models proposed to calculate ω over codons, and applied to each of the three studied loci

Locus	Model	LR	Parameter estimate	d*N*/d*S*	Positively selected sites ($p > 0.90$)
chs2	One-ratio (M0)	−543.328036	0.798	0.798	None
	Neutral (M1)	−543.33861	p0 = 1.00000 ωo = 0.00000 p1 = 0.00000 ω1 = 0.79797		Not allowed
	Selection (M2)	−543.32813	p0 = 0.00000 ω0 = 0.79797 p1 = 0.00000 ω1 = 1.00000 p2 = 0.00000 ω2 = 1.00000		
	Free-ratio (M3)	−543.33145	p0 = 0.03490 ω0 = 0.30405 p1 = 1.00000 ω1 = 0.79797 p2 = 0.00000 ω2 = 1.09596		None
	Beta (M7)	−543.328141	p = 99.00000 q = 25.06039	0.7982	Not allowed
	Beta & w (M8)	−543.328665	p0 = 1.00000 p = 98.99999 q = 21.60530 (p1 = 0.00000) ω = 1.00000		None
Glucan synthase	One-ratio (M0)	−847.265924	0.1132	0.1132	None
	Neutral (M1)	−847.265924	p0 = 1.00000 ωo = 0.11324 p1 = 0.00000 ω1 = 1.00000	0.1132	Not allowed
	Selection (M2)	−847.265924	p0 = 1.00000 ωo = 0.11324 p1 = 0.00000 ω1 = 1.00000 p2 = 0.00000 ω2 = 3.00000	0.1132	
	Free-ratio (M3)	−847.265924	p0 = 0.03490 ωo = 0.11330 p1 = 0.91236 ω1 = 0.11324 p2 = 0.05274 ω2 = 0.11328	0.1132	None
	Beta (M7)	−847.266062	p = 12.68894 q = 99.00000	0.1133	Not allowed
	Beta & w (M8)	−847.266081	p0 = 1.00000 p = 12.73125 q = 99.00000 (p1 = 0.00000) ω = 1.00000		
arf	One-ratio (M0)	−543.328036	0.798	0.798	None
	Neutral (M1)	−543.328036	p0 = 1.00000 ωo = 0.00000 p1 = 0.00000 ω1 = 0.79797		Not allowed
	Selection (M2)	−543.328036	p0 = 0.00000 ω0 = 0.79797 p1 = 0.00000 ω1 = 1.00000 p2 = 0.00000 ω2 = 1.00000		
	Free-ratio (M3)	−543.328036	p0 = 0.03490 ωo = 0.30405 p1 = 1.00000 ω1 = 0.79797 p2 = 0.00000 ω2 = 1.09596		None
	Beta (M7)	−543.328141	p = 99.00000 q = 25.06039	0.7982	Not allowed
	Beta + w (M8)	−543.328665	p0 = 1.00000 p = 98.99999 q = 21.60530 (p1 = 0.00000) ω = 1.00000		None

Table 5
Estimates of nucleotide diversity (θ), and test statistics across the studied regions of *chs2*

Region	Length, bp	θ	π	T	*D*	*F*	*D**	*F**
chs2								
Overall	1162	0.00405	0.00159	−1.55*	−2.462*	−2.41*	−2.25*	−2.261*
Introns	334	0.00559	0.00468	−0.615	0.048	−0.192	0.062	−0.177
Exons	828	0.00405	0.00159	−1.762*	−3.023*	−2.909*	−2.803*	−2.752*
Synonymous	204.3	0.00513	0.00319	−0.896	−0.968	−1.062	−0.921	−1.023
Non-synonymous	620.7	0.00371	0.00107	−1.882*	−2.967*	−2.925*	−2.785*	−2.789*
Glucan synthase								
Overall	1152	0.002	0.00131	−1.223	−0.888	−1143	−0.816	−1.081
Introns	84	0.00534	0.00716	0.288	0.704	0.643	0.702	0.64
Exons	1068	0.00197	0.00125	−1.509*	−1.279	−1.539	−1.193	−1.467
Synonymous	266.61	0.00633	0.00332	−1.307	−1.02	−1.264	−0.956	−1.209
Non-synonymous	801.39	0.00053	0.00008	−1.239	−0.977	−1.181	−0.949	−1.157
arf								
Overall	407	0.00311	0.00244	−1.402*	−0.679	−1.038	−0.638	−0.999
Introns	158	0.00541	0.00509	−1.259	−0.187	−0.599	−0.168	−0.579
Exons	249	0.00742	0.00534	−0.934	−0.977	−1.076	−0.949	−1.052
Synonymous	66.12	0	0	—	—	—	—	—
Non-synonymous	182.88	0.00152	0.00052	−0.934	−0.977	−1.076	−0.949	−1.052

The asterisk marks the tests that are significant at a $P = 0.05$ level.

created by interspersed long and short branches as revealed by the topology, thus allowing us to estimate the divergence date of the species of *P. brasiliensis*.

As recently described, *P. brasiliensis* exhibits recombination (Matute et al., 2006). Nevertheless, the intraspecific polymorphism degree was shown to be small. In cases like this when polymorphism is low, the apparent recombination rate should also be greatly reduced in comparison with species with higher polymorphism levels. Although recombination may occur, the resulting recombinant products are usually identical to the pre-recombination states. In keeping with this expectation, the Rm and the IST estimates for each species as well as for the species complex, equaled zero (Table 4).

Several likelihood-based tests were used to search for evidence of positive selection using the CODEML program of the PAML package. Codon-based substitution models have been widely used to identify amino acid sites under positive selection in comparative analysis of protein-coding DNA sequences. Statistical distributions have been used to model the variation in ω among sites; some sites can have a ω value above 1 and, as such, be under positive selection, while the rest of the sequence may be under purifying selection with ω below 1. If models M2 and/or M3 describe the data significantly better than models M1 and M0, and if at least one of the estimated values of ω is >1, the studied genetic region is under positive selection (Nielsen and Yang, 1998; Yang and Bielawski, 2000). The sites of *chs2* appeared to have evolved in a uniform way, as shown by the fact that the model of one-ratio of ω over sites (M0) could not be rejected.

Models M7 (β-model) and M8 (β + ω-model) are similar because they model the distribution of ω among amino acid positions as a β-distribution and estimate the parameters of that distribution. M8 differs from M7 in including an additional category of sites that are not part of the β-distribution. A test of whether M8 fits the data significantly better than M7 thus constitutes a formal test to determine whether positive selection has acted or is acting. M8 has been described as the most accurate of the models (Anisimova et al., 2002) and in this case predicted that none of the amino acid sites had a $\omega > 1$, with significant Bayesian posterior probabilities (Table 4). Results from maximum-likelihood models of codon evolution indicate that the replacement mutations within the *chs2* gene are not caused by positive selection. Neutrality tests were applied to evaluate an alternate hypothesis as demographic events, selective sweep and background selection.

The changes in the population size and a selective sweep are only distinguishable when theta estimates based on genes from different regions of the genome are compared (Cummings and Clegg, 1998). For this reason, theta was also calculated for two other coding genes as well as for three additional non-coding genes. A bottleneck would affect all loci, whereas the impact of selective sweep would be confined to the region associated with the locus that had been subject to it.

Recovery from a population bottleneck or a recent demographic expansion are both expected to lead to a transient excess of rare variants, but this would be true for all sites (synonymous and non-synonymous and the other 4 loci) and consequently, this explanation does not appear consistent with the observed data (Ramakrishnan et al., 2005). Similarly, a selective sweep is not consistent with the distribution of synonymous and intron polymorphism because the recovery from a selective sweep would have the same consequences as a recovery from a bottleneck for the region associated with a single locus (Nielsen et al., 2005).

Having rejected the population size changes and the selective sweep hypotheses, attention should be given to the background selection hypothesis, in which the distribution of neutral sites could be skewed towards an excess of rare

alleles owing to their linkage association with negatively selected mutations (Charlesworth, 1994).

Several features of the data obtained appear consistent with the background selection hypothesis of the *chs2* locus. (i) Background selection is expected to be amplified in species in which there is random-assortative mating and in regions in which no genetic recombination is occurring (Charlesworth, 1994; Charlesworth et al., 1993).

There is no evidence for recombination within each one of the species for this gene. (ii) The amino acid replacements in the sample are all unique, as would be expected in the presence of weak negative selection. However, to ensure that selection against deleterious mutations is the most important reason of the observed distribution of non-synonymous changes, the following additional facts deserve consideration: (i) there are nearly twice as many non-synonymous polymorphisms as synonymous polymorphisms in the sample (seven versus four); (ii) there appears to be no restriction on the kinds of amino acid changes accepted; (iii) the frequency of deleterious genes in haploid species should be very low; (iv) the N*e* values calculated from the polymorphism in *chs2* are lower than the N*e* values calculated from regions that evolve neutrally, and (v) the estimates of effective population size are relatively large for the neutral genes, so small selective values should be effective. These observations point to very weak negative selection on the amino acid replacements at non-synonymous sites. The mere fact that 7 non-synonymous changes were observed in a relatively large sample (67 different isolates) implies either very weak selection or some force favoring rare variants.

The distribution of synonymous and intron sites was not perturbed toward a significant excess of rare alleles as would be predicted by the background selection hypothesis. *T* value is negative for synonymous and intron sites, but it is too weak to lead to a significant perturbation in the distribution. As the observed data pertain to a limited region of the chromosome rather than to whole chromosomes (Charlesworth et al., 1993), it is possible that there may be a background effect. This result would appear to suggest rather weak selection at the whole chromosome level, considering that different chromosomal regions may be affected by different patterns of selection. Because of the very reduced levels of polymorphism within each one of the *P. brasiliensis* species and the low recombination levels observed in this region, the observed data could be more likely explained by the selective forces that affect loci over most of the chromosome. Consequently, the factors that have influenced the sequence diversity at *chs2* may be determined by other kind of factors that are not observable. The dynamics at *chs2* may be influenced by selection at other loci on the same chromosome, but at a considerable distance from *chs2*. This explanation would help to account for the large number of amino acid replacements detected in the sample, since selection at this locus may be moderated by selection operating at other chromosomal loci and not directly over it.

For future studies of fungal nucleotide sequence diversity it will be important to focus on other genetic loci and ask whether the processes affecting different loci are heterogeneous within genomes and lineages. With data from more individuals and more loci, we hope to acquire better knowledge about the forces that have shaped the molecular diversity in *P. brasiliensis* species complex.

Acknowledgments

We wish to thank John W. Taylor and E. Rodewald for their particularly insightful comments on an earlier version of the manuscript. This work was supported by Fogarty International Research Collaboration Award grant R03TW01308 (to J.G.M.). Additional funds were obtained from the Comite de Investigaciones de la Universidad de Antioquia (Sostenibilidad 2005–2006). D.R.M. thanks Carlos "El Pibe" Valderrama for the unending inspiration.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fgb.2007.01.006.

References

Almeida, A.J., Matute, D.R., Carmona, J.A., Martins, M., Torres, I., McEwen, J.G., et al. Genome size and ploidy of *Paracoccidioides brasiliensis* reveals a haploid DNA content: Flow cytometry and GP43 sequence analysis. Fungal Genet. Biol. 2007, 44, 25–31.

Anisimova, M., Bielawski, J.P., Yang, Z., 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. Mol. Biol. Evol. 19, 950–958.

Bromham, L., Penny, D., 2003. The modern molecular clock. Nat. Rev. Genet. 4, 216–224.

Brummer, E., Castaneda, E., Restrepo, A., 1993. Paracoccidioidomycosis: an update. Clin. Microbiol. Rev. 6, 89–117.

Charlesworth, B., 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. 63, 213–227.

Charlesworth, B., Morgan, M.T., Charlesworth, D., 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134, 1289–1303.

Cummings, M.P., Clegg, M.T., 1998. Nucleotide sequence diversity at the alcohol dehydrogenase 1 locus in wild barley (*Hordeum vulgare* ssp. *spontaneum*): an evaluation of the background selection hypothesis. Proc. Natl. Acad. Sci. USA 95, 5637–5642.

Diez, S., Garcia, E.A., Pino, P.A., Botero, S., Corredor, G.G., Peralta, L.A., et al., 1999. PCR with *Paracoccidioides brasiliensis* specific primers: potential use in ecological studies. Rev. Inst. Med. Trop. Sao Paulo 41, 351–358.

Felsenstein, J., 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. 22, 521–565.

Fu, Y.X., 1996. New statistical tests of neutrality for DNA samples from a population. Genetics 143, 557–570.

Fu, Y.X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147, 915–925.

Fu, Y.X., Li, W.H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693–709.

Georgopapadakou, N.H., Tkacz, J.S., 1995. The fungal cell wall as a drug target. Trends Microbiol. 3, 98–104.

Hudson, R.R., 1994. How can the low levels of DNA sequence variation in regions of the drosophila genome with low recombination rates be explained? Proc. Natl. Acad. Sci. USA 91, 6815–6818.

Kumar, S., Tamura, K., Nei, M., 1994. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. Comput. Appl. Biosci. 10, 189–191.

Li, W.H., 1997. Molecular Evolution. Sinauer Associates Inc., Sunderland, MA.

Matute, D.R., McEwen, J.G., Puccia, R., Montes, B.A., San-Blas, G., Bagagli, E., et al., 2006. Cryptic speciation and recombination in the fungus *Paracoccidioides brasiliensis* as revealed by gene genealogies. Mol. Biol. Evol. 23, 65–73.

Morais, F.V., Barros, T.F., Fukada, M.K., Cisalpino, P.S., Puccia, R., 2000. Polymorphism in the gene coding for the immunodominant antigen gp43 from the pathogenic fungus *Paracoccidioides brasiliensis*. J. Clin. Microbiol. 38, 3960–3966.

Muse, S.V., Gaut, B.S., 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. 11, 715–724.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148, 929–936.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G., Bustamante, C., 2005. Genomic scans for selective sweeps using SNP data. Genome Res. 15, 1566–1575.

Nino-Vega, G.A., Buurman, E.T., Gooday, G.W., San-Blas, G., Gow, N.A., 1998. Molecular cloning and sequencing of a chitin synthase gene (CHS2) of *Paracoccidioides brasiliensis*. Yeast 14, 181–187.

Nino-Vega, G.A., Munro, C.A., San-Blas, G., Gooday, G.W., Gow, N.A., 2000. Differential expression of chitin synthase genes during temperature-induced dimorphic transitions in *Paracoccidioides brasiliensis*. Med. Mycol. 38, 31–39.

Parmley, S.F., Gross, U., Sucharczuk, A., Windeck, T., Sgarlato, G.D., Remington, J.S., 1994. Two alleles of the gene encoding surface antigen P22 in 25 strains of *Toxoplasma gondii*. J. Parasitol. 80, 293–301.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Ramakrishnan, U., Hadly, E.A., Mountain, J.L., 2005. Detecting past population bottlenecks using temporal genetic data. Mol. Ecol. 14, 2915–2922.

Rannala, B., 1996. The sampling theory of neutral alleles in an island population of fluctuating size. Theor. Popul. Biol. 50, 91–104.

Rannala, B., Hartigan, J.A., 1996. Estimating gene flow in island populations. Genet. Res. 67, 147–158.

Redecker, D., Kodner, R., Graham, L.E., 2000. Glomalean fungi from the Ordovician. Science 289, 1920–1921.

Restrepo, A., 2003. Paracoccidiodomycosis. In: Dismukes, W.E., Pappas, P.G., Sobel, J.D. (Eds.), Clinical Mycology. Oxford University Press, Oxford, New York, NY.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19, 2496–2497.

Ruiz-Herrera, J., Gonzalez-Prieto, J.M., Ruiz-Medrano, R., 2002. Evolution and phylogenetic relationships of chitin synthases from yeasts and fungi. FEMS Yeast Res. 1, 247–256.

San-Blas, G., 1985. *Paracoccidioides brasiliensis*: cell wall glucans, pathogenicity, and dimorphism. Curr. Top. Med. Mycol. 1, 235–257.

Sanderson, M.J., 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14, 1218–1231.

Schmidt, M., Varma, A., Drgon, T., Bowers, B., Cabib, E., 2003. Septins, under Cla4p regulation, and the chitin ring are required for neck integrity in budding yeast. Mol. Biol. Cell. 14, 2128–2141.

Storz, J.F., 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol. Ecol. 14, 671–688.

Swanson, H.I., Yang, J.H., 1998. The aryl hydrocarbon receptor interacts with transcription factor IIB. Mol. Pharmacol. 54, 671–677.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (* and other methods). In: Sinauer Associates, Sunderland, MA.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Taylor, T.N., Hass, H., Kerp, H., Krings, M., Hanlin, R.T., 2005. Perithecial ascomycetes from the 400 million year old Rhynie chert: an example of ancestral polymorphism. Mycologia 97, 269–285.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acid. Res. 22, 4673–4680.

Vallender, E.J., Lahn, B.T., 2004. Positive selection on the human genome. Hum. Mol. Genet. 13 Spec No 2, R245–R254.

Van Burik, J.A., Schreckhise, R.W., White, T.C., Bowden, R.A., Myerson, D., 1998. Comparison of six extraction techniques for isolation of DNA from filamentous fungi. Med. Mycol. 36, 299–303.

Worobey, M., 2001. A novel approach to detecting and measuring recombination: new insights into evolution in viruses, bacteria, and mitochondria. Mol. Biol. Evol. 18, 1425–1434.

Yang, Z., 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. 10, 1396–1401.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.

Yang, Z., Bielawski, J.P., 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evolut. 15, 496–503.

Yang, Z., Wong, W.S., Nielsen, R., 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22, 1107–1118.