Positive Directional Selection in the Proline-Rich Antigen (*PRA*) Gene Among the Human Pathogenic Fungi *Coccidioides immitis*, *C. posadasii* and Their Closest Relatives

Hanna Johannesson, Pilar Vidal,[1] *Josep Guarro,*[1] *Roger A. Herr,*[2] *Garry T. Cole,*[2] *and John W. Taylor*

Department of Plant and Microbial Biology, University of California, Berkeley

In this study, we investigate the possibility of selection acting on the proline-rich antigen (*PRA*) gene in natural populations of the two human pathogens, *Coccidioides immitis* and *Coccidioides posadasii*, and three of their close relatives, *Chrysosporium lucknowense*, *Chrysosporium queenslandicum*, and *Uncinocarpus reesii*. We addressed the following questions: Is diversifying selection acting on *PRA* in the pathogenic species as a result of avoidance of the host's immune system, and has adaptation to a pathogenic life style lead to positive directional selection and increased rate of evolution in *PRA* between the species? For these purposes, we amplified and sequenced from 40 individuals belonging to the five species, the entire coding region of the *PRA* gene, as well as partial sequences from the coding region of each of the three housekeeping genes glyderaldehyde-3-phosphate dehydrogenase, glutamine synthetase A, and hexokinase A. We used likelihood-based methods to compare models of different types of selective pressure among codons to analyze the mode of evolution of the genes and found that the *PRA* gene evolves under positive selection, but the investigated parts of the housekeeping genes evolve primarily under purifying selection. We found a very low level of intraspecific variability and no evidence of diversifying selection, suggesting that the increased rate of evolution in the *PRA* gene is not a result of avoidance of the host's immune system. Neither did likelihood-based analyses suggest that selection was stronger on the branch separating pathogenic and nonpathogenic species. Instead, we suggest that positive selection act on *PRA* as a consequence of spore cell–wall morphogenesis unique to each species.

Introduction

The neutral theory of evolution asserts that the majority of changes at the molecular level are fixed by random drift of selectively equivalent mutations (Kimura 1983), but the "arms races" run by hosts and their pathogens offer clear opportunities for selection to play a prominent evolutionary role. In this study, we investigate the possibility of selection acting on the proline-rich antigen (*PRA*) gene in natural populations of the two human pathogenic fungi, *Coccidioides immitis* and *Coccidioides posadasii*, and their three close relatives, *Chrysosporium lucknowense*, *Chrysosporium queenslandicum*, and *Uncinocarpus reesii*. We address the question of diversifying selection acting on *PRA* in the pathogenic species as a result of avoidance of the host's immune system. We also ask if adaptation to a pathogenic life style has led to positive directional selection and an increased rate of evolution in *PRA* among the species.

Coccidioides spp. are the etiological agents of the human respiratory disease known as coccidioidomycosis or San Joaquin Valley fever (Galgiani 1999). During the most recent epidemic of coccidioidomycosis, which struck California in the beginning of the past decade, the number of case reports increased 10-fold (Pappagianis 1994). Two pathogenic species of *Coccidioides*, with nearly identical phenotypes, now are recognized, *C. immitis* and *C. posadasii* (Fisher et al. 2002). Both species have been shown to have a recombining genetic structure (Burt et al. 1996; Fisher et al. 2000). The species are dimorphic, living as hyphal saprobes in the desert soil or as unicellular pathogens that convert into multicellular sporulating spherules in the mammalian host. Death and decay of the host results in the fungus reverting to its saprobic morphology (Maddy and Crecelius 1967; Saubolle 1996) from which new air-dispersed infectious propagules are produced. Thus, direct transmission of the fungus between hosts does not occur (Pappagianis 1988).

In the life cycle of *Coccidioides*, spherules are exposed to the host's immune system; thus, their surface proteins are candidates for anticoccidioidomycosis vaccines. One of the first immunogenic proteins identified was the highly glycosylated proline-rich antigen (PRA), also known as antigen 2 (Ag2) (Cox, Brummer, and Lecara 1977; Cox 1989; Dugger et al. 1991; Galgiani et al. 1992). *PRA* is a member of a gene family of at least eight paralogous genes in *Coccidioides* (Herr, Hung, and Cole 2003). Phylogenetic analyses of the *PRA* sequences used in this study show that alleles from all five species coalesce more recently than any gene duplication, and all coalesce to one of the eight paralogous genes (data not shown). The protein is suggested to have an endoglucanase activity and to be important for spherule cell-wall morphogenesis during the infection process in *Coccidioides* (Zhu et al. 1996*b*). This idea is supported by the finding of an increased expression of the gene during spherule development and maturation (Peng et al. 1999). It is located in the fungal cell wall (Galgiani et al. 1992), most probably attached to the cell-wall matrix, and contains a putative N-terminal signal peptide for export to the cell surface (Peng et al. 2002). Accordingly, the first 111 amino acids (aa) of PRA show a structural similarity to cell-wall proteins

[1] Present address: Unitat de Microbiologia, Facultat de Medicina i Ciènces de la Salut, Universitat Rovira i Virgili, 43201-Reus, Spain.

[2] Present address: Department of Microbiology and Immunology, Medical College of Ohio, Toledo.

Key words: *Coccidioides*, positive selection, proline-rich antigen, vaccine, paml, antigen.

E-mail: hjohanne@nature.berkeley.edu.

Mol. Biol. Evol. 21(6):1134–1145. 2004
DOI:10.1093/molbev/msh124
Advance Access publication March 19, 2004

Table 1
Fungal Material Used in the Study

Species	Strain[a]	Origin[b]
C. immitis	RMSCC 2102	South CA, USA
	RMSCC 2012	Central CA, USA
	RMSCC 2017	Central CA, USA
	RMSCC 2271	Central CA, USA
	RMSCC 2274	Central CA, USA
	RMSCC 2275	Central CA, USA
	RMSCC 2278	Central CA, USA
	RMSCC 3505	Mexico
	RMSCC 5273	Mexico
C. posadasii	RMSCC 2103	South CA, USA
	RMSCC 1038	AZ, USA
	RMSCC 1039	AZ, USA
	RMSCC 1040	AZ, USA
	RMSCC 1049	AZ, USA
	RMSCC 1444	AZ, USA
	RMSCC 3490	Mexico
	RMSCC 3503	Mexico
	RMSCC 2345	Mexico
	RMSCC 2346	Mexico
	RMSCC 2347	Mexico
	RMSCC 2348	Mexico
	RMSCC 2377	South America
	RMSCC 2378	South America
	RMSCC 2379	South America
	RMSCC 3272	South America
C. lucknowense	FMR 6082	Catalonia, Spain
	FMR 6083	Catalonia, Spain
	FMR 6126	Catalonia, Spain
C. queenslandicum	FMR 6074	Catalonia, Spain
	FMR 6075	Catalonia, Spain
	FMR 6076	Catalonia, Spain
	FMR 6078	Catalonia, Spain
	FMR 6079	Catalonia, Spain
	FMR 6080	Catalonia, Spain
U. reesii	UAMH 1704	CA, USA
	UAMH 1955	Argentina
	UAMH 2002	Italy
	UAMH 2050	Italy
	UAMH 2847	Hungary
	UAMH 3918	Italy

[a] RMSCC: Roche Molecular Systems Culture Collection, Alameda, CA, USA; UAMH: University of Alberta Microfungus Collection and Herbarium, Edmonton, AB, Canada; FMR: Facultat Medicina de Reus, Reus, Spain.

[b] Origin of the strains of *Coccidioides* spp. follow the classification of populations made by Fisher et al. (2001); all strains of *Chrysosporium* originate from different beaches or rivers in Catalonia (see Vidal and Guarro [2002]).

suggested to be important for cell-wall morphogenesis in *Candida albicans* (Braun et al. 2000). A proline-threonine–rich, tetrapeptide repeat region, a common feature of fungal cell-wall proteins, is found in the central domain. This region is expected to be highly glycosylated and cross-linked to cell-wall polysaccharides. Sequence analyses have suggested several possible functional domains of the protein, including a protein kinase C and two casein kinase II phosphorylation sites (Zhu et al. 1996*b*). It has been demonstrated that people with coccidioidomycosis make both B-cell and T-cell antigenic responses to deglycosylated PRA (Dugger et al. 1991; Galgiani et al. 1992; Magee and Cox 1995; Zhu et al. 1996*a*; Zhu et al. 1997; Peng et al. 2002), although the exact location of the immunogenic regions of the protein is unknown.

Just as infectious disease is thought to be a major selection force that drives and maintains the extraordinary diversity of the major histocompatibility complex (MHC) in humans (reviewed by Hughes and Yeager [1998]), the immune system of vertebrates itself has been proven to exert natural selection on pathogens, favoring avoidance of immune recognition (e.g., Deitsch, Moxon, and Wellems [1997]). The study of sequence divergence of genes coding for antigens in natural populations of a pathogen can be of great practical importance because pathogens with a large sequence diversity of antigens present a major challenge to successful vaccine design (Parmley et al. 1994; Crewther et al. 1996; Araujo, Slifer, and Kim 1997; Renia et al. 1997). Additionally, identification of regions with particularly high rates of nonsynonymous nucleotide substitutions can provide clues to the location of immunogenic regions (Hughes 1992).

Selection at the molecular level is typically detected by comparing the ratio of nonsynonymous (d_N) to synonymous (d_S) substitutions between species (ω). Positive selection is inferred when ω exceeds 1, whereas purifying selection is inferred when ω is less than 1. Positive directional selection is operating when successive amino acid changes make a protein more efficient at performing a particular task, and the changes are preserved in future lineages. On the other hand, 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. Recently, likelihood methods have been developed that allow ω to vary among branches in a phylogeny (Yang 1998; Yang and Nielsen 1998) as well as among codons (Nielsen and Yang 1998; Yang et al. 2000; Yang and Swanson 2002). This approach provides a more sensitive test of positive selection than pairwise, distance-based estimates in that it offers the possibility to detect sites under positive selection within a gene region with elevated proportions of synonymous changes. Furthermore, it can identify regions of the protein with potential functional importance (Bielawski, Dunn, and Yang 2000; Yang and Bielawski 2000).

In this report, we show that the *PRA* gene evolves under a higher selective pressure than genes encoding proteins that are not surface located (i.e., housekeeping genes) and, thus, are unlikely to be involved in host-immune response. We use likelihood-based methods to verify that the selective pressure is consistent with positive selection. We found a very low level of intraspecific variability, suggesting that the increased rate of evolution in the *PRA* gene is not a result of avoidance of the host's immune system. Rather, we suggest it to be a consequence of species-specific, spore morphogenesis.

Materials and Methods

Fungal Material

Forty isolates belonging to five species were used in this study (table 1). The samples of *Coccidioides* and *Uncinocarpus* represent the known geographical distributions of the species. All isolates of *Coccidioides* were previously genotyped and assigned to species by using

microsatellite markers (Fisher et al. 2002). The isolates of *Chrysosporium lucknowense* and *C. queenslandicum* were identified to species by using morphological and ITS sequence characters (Vidal and Guarro 2002). The six isolates of *U. reesii* originate from the same cryptic species (UIII) in the *U. reesii* species complex that was previously discovered using gene genealogies of three protein-coding genes (Koufopanou et al. 2001). *C. lucknowense*, *C. queenslandicum* and *U. reesii* are among the closest known relatives of *Coccidioides*. They are saprobic species, and no production of spherules or endospores, stages that presumably are essential for disease, have been reported from their life cycles. *C. queenslandicum* occasionally has been reported as pathogenic, producing fungal nail infection (onychomycosis [Reboux et al. 1995]) and has been reported to cause a disseminated infection in a garter snake (Vissiennon et al. 1999). Although it has never been reported as a systemic pathogen of mammals and birds, its ability to grow at 38°C indicates a potential for virulence in mammals (Apinis and Rees 1976). *U. reesii* occasionally has been collected from the lungs of rodents but appears to be only a transient and harmless inhabitant of animals (Pan, Sigler, and Cole 1994).

DNA Manipulations

Total genomic DNA was extracted from lyophilized material according to protocols described previously (Lee and Taylor 1990; Burt et al. 1995).

The entire coding region of the *PRA* gene (609 bp), as well as partial sequences from the coding region of each of the three housekeeping genes glyderaldehyde-3-phosphate dehydrogenase (*GAPDH*, 585 bp of a total of 1,011-bp coding sequence), glutamine synthetase A (*glnA*, 342 of 1,035 bp), and hexokinase A (*hxkA*, 474 of 1470 bp) were amplified from all isolates listed in table 1. The previously published sequence of *PRA* from *Coccidioides posadasii* (GenBank accession number AF013256) was used as a template for primer design for that gene. The target region of the housekeeping genes were selected as follows: Sequences of each of the housekeeping genes, published and characterized from the ascomycetous fungi *Ajellomyces capsulatus*, *Aspergillus nidulans*, and *Aspergillus oryzae* for the loci *GAPDH, glnA* and *hxkA*, respectively, were used to search via Blast for homologous genes in *C. posadasii* (http://tigrblast.tigr.org/ufmg/). In the resulting alignments, stretches of coding regions ranging from 300 to 600 bp, flanked by conserved regions of 20 to 30 bp, suitable for primer design, were identified and selected.

PRA was amplified from the *Coccidioides* isolates with the primer pair PRA-F1 (5′-CCGTTAGACGCA-CATACATA-3′) and PRA-R2 (5′-CGTGCTTGTCAG-TTTTGCTG-3′), and the *Uncinocarpus* and *Chrysosporium* isolates with the pair PRA-F1 and PRA-R3 (5′-AATTTACAGGTAGGCAGCGA-3′). The loci *GAPDH, glnA*, and *hxkA* were amplified from isolates of all species using the primers GAPDH-F1 (5′-GCCTAYATGCT-CAAATAYGAC-3′) and GAPDH-R3 (5′-TTGGCGG-TGGGAACACGCAT-3′), GlnA-F2 (5′-GATGTCTAC-CTTCGCCCYGTC-3′) and GlnA-R1 (5′-CAACCTGG-TAYTCCCAYTGAG-3′), and HxkA-F (5′-CTGYGAR-TAYGGTGCCTTTGA-3′) and HxkA-R (5′-GGCCTT-GAARTGGGGATATTT-3′), respectively. The IUPAC ambiguity coding is used for degenerate primers. All primers were designed manually for this study.

Each PCR reaction was performed using the Expand High Fidelity PCR System (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's recommendation, using an Eppendorf thermal cycler. PCR products were purified using the Qiaquick PCR purification kit (QIAGEN, Chatsworth, Calif.) before sequencing. All sequences were determined with an Applied Biosystems 3100 sequencer using the Taq DyeDeoxi Terminator™ cycle system (ABI).

Phylogeny Reconstruction

We assumed that the true evolutionary history of each of the genes under study is the same as the evolutionary history of the species, and accordingly, we used the aligned sequences from the coding regions of the loci to infer a phylogeny for the included species to use as an input topology for the likelihood analyses of positive selection. Because of the low intraspecific variability, only one isolate per species was included. All analyses were carried out using the maximum-parsimony (MP) and maximum-likelihood (ML) analyses in PAUP* version 4.0b (Swofford 2001). By performing a series of likelihood ratio tests of different models using Modeltest version 3.04 (Posada and Crandall 1998), we found that the most likely model of substitutions for this data set were a general time-reversible model of substitution (data not shown). No rooting of the trees was performed.

To verify that the trees inferred from data sets for the genes were not in significant conflict, the partition homogeneity test in PAUP* 4.0b was used between the data sets in all possible pairwise combinations, using 500 replicates and the heuristic general search option. This test randomly shuffles phylogenetically informative sites among the two paired loci, and if the data sets are compatible, shuffling of sites between the loci should not produce summed tree lengths significantly greater than that produced by the observed data (Farris et al. 1995; Huelsenbeck, Bull, and Cunningham 1996).

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 3.13d package (Yang 1997; Yang et al. 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 use the default settings provided by Yang et al. (2000). Given the low observed intraspecific variability, as well as the clear species limits of the included species (see above), we assumed linkage between collinear sites (i.e., no recombination within each data set). To verify which of the models best fits the data, likelihood

ratio tests (LRTs) were performed by comparing twice the differences in log-likelihood values ($-2\ln\Delta$) 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.

Positive selection may act at discrete points during the evolution of a lineage, rather than constantly across an entire phylogeny; therefore, we examined whether ω varies across all lineages for each gene (Goldman and Yang 1994; Yang 1998; Yang, Swanson, and Vacquier 2000). For this test, a simple model that assumes a constant ω across all lineages (one-ratio model, M0) is compared with a more general model that assumes an independent ω for each branch in the phylogeny (free-ratio model, M1). The free-ratio model was used to estimate the ω value for each branch in the phylogeny. Although this model is parameter rich and unlikely to produce accurate ω estimates for all branches (Yang, Swanson, and Vacquier 2000), it is nonetheless useful for identifying lineages where episodes of positive selection might have occurred. To examine whether any particular lineage in the given phylogeny has a different ω than the other lineages, two-ratio models (M2), which allow a different ω for each branch from the background ω_0, were compared with the one-ratio model (M0) for each gene.

Models of variable ω among sites were used to test for the presence of sites under selection ($\omega > 1$) and to identify them. We used six models outlined by Nielsen and Yang (1998) and implemented in PAML (Yang 1997; Yang et al. 2000). The one-ratio model (Nssites 0) assumes one ω for all sites. Two of the models assume neutrality. The neutral model (Nssites 1) assumes two classes of sites in the protein, the conserved sites at which $\omega = 0$ and the neutral sites in which $\omega = 1$. The beta model (Nssites 7) uses a β distribution of ω over sites: β (p,q), which, depending on parameters p and q, can take various shapes in the interval (0,1). Three models allow for sites with ω greater than 1 and can be considered tests of positive selection. The selection model (Nssites 2) adds a third class of sites to the neutral model, in which ω is a free parameter. The discrete model (Nssites 3) uses a general discrete distribution with three site classes, with the proportions (p_0, p_1, and p_2) and the ω ratios (ω_0, ω_1, and ω_2) estimated from the data. The beta&ω model (Nssites 8) adds an extra class of sites to the beta model, with the proportion of ω estimated from the data, thus allowing for sites with ω greater than 1. We used LRTs to make 3 comparisons: the one-ratio model (Nssites 0) was compared with the discrete model (Nssites 3), the neutral model (Nssites 1) was compared with the selection model (Nssites 2), and the beta model (Nssites 7) was compared with the beta&ω model (Nssites 8) using 4, 2, and 2 degrees of freedom, respectively (Yang et al. 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 outlined by 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 have been estimated, Bayes' theorem is 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).

Results

Sequence Variability in PRA and Housekeeping Genes

A very low level of intraspecific variability was found within the coding regions of the investigated loci. In *PRA*, we found 1, 2, and 3 intraspecific substitutions in *Coccidioides immitis*, *Coccidioide posadasii*, and *Chrysosporium queenslandicum*, respectively. We found one substitution in *GAPDH* within *C. queenslandicum*, one substitution in *glnA* within *C. immitis*, and three substitutions in *hxkA* within *C. posadasii*. All substitutions but one (positioned in the *PRA* gene in *C. queenslandicum*) were synonymous. The sequences upon which the analyses were made are submitted to GenBank under the accession numbers AY536445 to AY536464.

Variability among species was substantial in the coding part of all four investigated gene loci. As shown in figure 1, the region of *PRA* rich in tetrapeptide repeats of TXX'P, where X is Ala, Glu, or His, and X' is Ala, Glu, or Gln, differed in both the number of tetrapeptide repeats and in the nonrepetitive sequence interspersed among the repeats for the five species. The number of repeats ranged from six for *C. queenslandicum* to nine for *C. immitis* and *C. posadasii*. The variability between species, exclusive of the repetitive, ambiguously aligned part of the *PRA* gene, is shown in table 2. For all four loci, the level of polymorphic nucleotide sites ranged from 21.7% to 29.1%. The proportion of polymorphic codons was significantly lower for *PRA* than it was for the *GAPDH, glnA,* or *hxkA* loci (Fisher's exact test, $P < 0.05$, 0.005, and 0.001, respectively). In contrast, the proportion of polymorphic codons with nonsynonymous substitutions was significantly higher for *PRA* than for any of the other loci (Fisher's exact test, $P < 0.001$); more than half of the codon polymorphisms were caused by nonsynonymous replacements in the *PRA*. Among the housekeeping genes, the proportion of polymorphic codons and the ratio of nonsynonymous to synonymous substitutions varied considerably. The *GAPDH* locus showed a significantly lower proportion of polymorphic codons but a higher ratio of nonsynonymous to synonymous substitutions than the *glnA* and *hxkA* genes. Compared with the *hxkA* locus, the *glnA* locus was significantly less polymorphic but exhibited a higher proportion of nonsynonymous to synonymous substitutions (Fisher's exact test, $P < 0.05$) (table 2).

Phylogenetic Analyses

The partition homogeneity test in PAUP* 4.0b revealed that shuffling of informative sites between the data sets did not produce summed tree lengths significantly greater than that produced by the observed data for any of the pairs of housekeeping loci ($P < 0.001$), indicating that there is no significant conflict between the three data sets. One single topology resulted from MP analysis of the combined data sets of the three housekeeping genes (fig. 2).

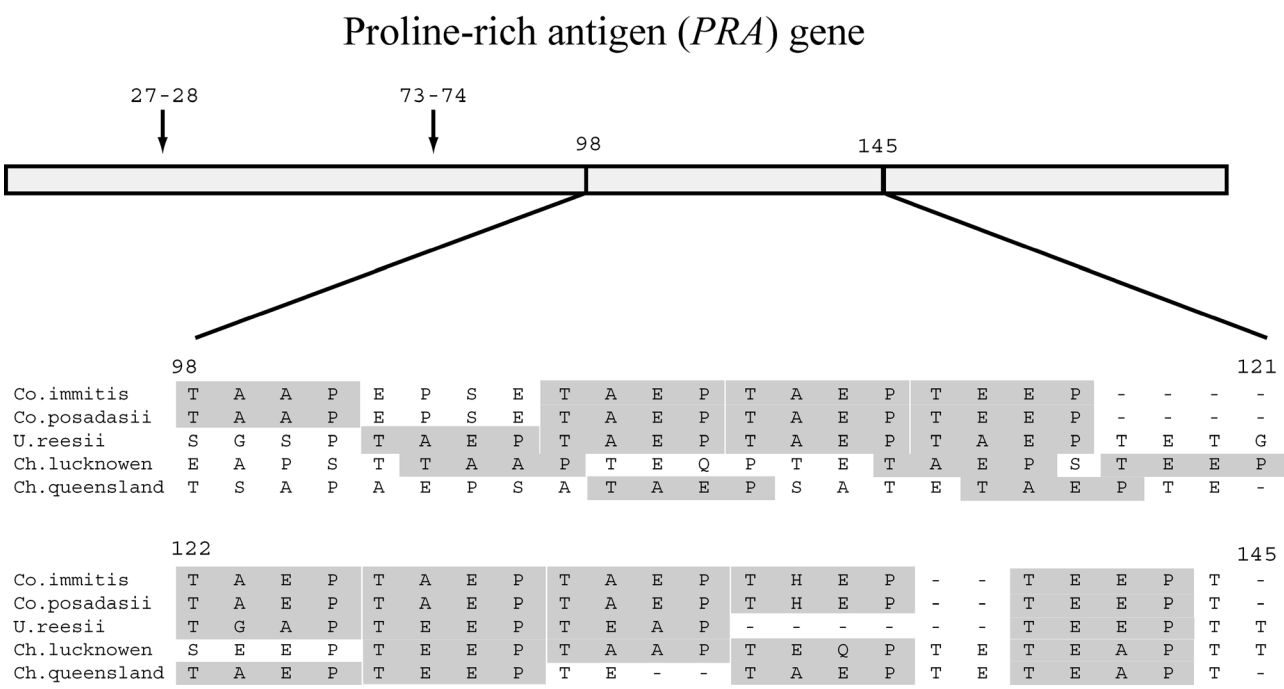

FIG. 1.—Amino-acid sequences of the proline-threonine rich tetrapeptide region in *PRA* of the investigated species, ranging from amino acid 98 to amino acid 145 in *Chrysosporium lucknowense*. Intraspecific variability of the entire gene was less than 0.5% for all species. Tetrapeptides are put in shaded boxes. The positions of the two introns are indicated by arrows.

An identical topology was recovered under ML analyses under the best-fit likelihood model obtained from the program Modeltest 3.04 or when using uneven weighting of the characters based on codon position in the MP analysis. This phylogenetic relationship of the species is in accordance with the phylogeny that was inferred from rDNA internal transcribed spacer sequences (Vidal et al. 2000). We used this unrooted tree topology as the basis for all likelihood-based tests of selection performed in this study.

The partition homogeneity test revealed that the *PRA* data set was in significant conflict with the tree constructed from the *hxkA* data set, and the *PRA* topology differed from the housekeeping topology in that the position of *U. reesii* and *C. queenslandicum* were switched. For comparative purposes, we also used the topology obtained from the *PRA* data set in a second set of tests of selection.

Evolutionary Pattern Among Lineages and Sites

We found no evidence of a variation of ω among lineages, but we did find a substantial variation of ω among sites in the data set. The estimated values of ω for each lineage obtained by the free-ratio model (M1) range from 0.20 to 0.38 for the *PRA* gene, 0.01 to 0.08 for *GAPDH*, 0 to 0.09 for *glnA*, and 0 to 0.22 for *hxkA*. The free-ratio model (M1) was compared with a model that assumes a constant ω across all lineages (M0) by performing LRTs with 6 degrees of freedom, and the model assuming a constant rate of ω across all lineages (M0) could not be rejected for any of the genes. Furthermore, no significant differences in ω across branches for each gene were detected when LRTs were used to compare the one-ratio model (M0) with a two-ratio model (M2) (which allows a different ω for each particular branch compared with the background ω_0) (data not shown).

Log-likelihood values and parameter estimates under models of variable ω among sites are listed in table 3. Using the one-ratio model (Nssites 0), the averaged value of ω for the *PRA* gene was 0.259, significantly higher than for any of the housekeeping genes (0.046, 0.023, and 0.020 for *GAPDH, glnA*, and *hxkA*, respectively; *t*-test, $P < 0.001$). The values of parameters under the discrete model (Nssites 3) for the different genes indicate that 67.6% of the sites in the nonrepetitive part of *PRA* gene are under purifying selection ($\omega = 0$), whereas 6.8% belong to a site class with $\omega = 5.49$, indicative of positive selection. The additional 25.5% of the sites belong to a class with ω of 0.76, indicating relaxed selective constraints or neutrality. In contrast, in the housekeeping genes, the vast majority of sites have evolved under purifying selection ($\omega = 0$ or is very close to 0), judging by the parameters of the discrete model (Nssites 3). An exception to this is a small fraction ($p_2 = 2.6\%$) of sites in the *GAPDH* gene that belong to a site class with ω_2 estimated to be 1.854, indicative of positive selection. Furthermore, one amino acid site was found to have a posterior probability greater than 95% of having ω greater than 1 in the *GAPDH* gene. This site is not positioned in any of the two active domains of the protein (Rossman et al. 1975; Templeton et al. 1992); neither is it a NAD+ binding site nor a catalysis site. This small fraction of sites with a higher ω is most probably the reason why the discrete model (Nssites 3) fits the GAPDH data set significantly better than the one-ratio

Table 2
Variability of Coding Regions of Each Gene Among All Five Species

	Locus			
	PRA (465 bp, 155 aa)[a]	GAPDH (585 bp, 195 aa)	*glnA* (342 bp, 114 aa)	hxkA (474 bp, 158 aa)
Pol. nucleotide sites	116 (24.9)	127 (21.7)	82 (23.9)	138 (29.1)
Polymorphic codons	73 (47.0)	104 (53.3)	71 (62.2)	120 (75.9)
Codons with nonsynonymous substitutions	39	24	7	14
Codons with synonymous substitutions	34	80	64	106

NOTE.—Percentage of total number of nucleotides and codons respectively, are shown within parentheses.
[a] Not including unalignable part shown in figure 1.

model (Nssites 0), as opposed to the glnA and hxkA data sets (table 4).

All models that allow for sites with ω greater than 1 (discrete, selection, and beta&ω models), that is, the models of positive selection, fit the PRA data significantly better than the corresponding neutral models (one-ratio, neutral and beta models) (table 4). The posterior probabilities that the codons of the nonrepetitive part of *PRA* belong to one of the three estimated classes with different selective pressures obtained from the discrete model are shown in figure 3. Three sites with a posterior probability greater than 95% of having an ω greater than 1 were identified using the Bayesian approach outlined by Nielsen and Yang (1998) (table 3). If the threshold for positive selection is reduced to a posterior probability greater than 0.5 that a site belongs to a class with ω greater than 1 (e.g., Miller [2003]), then 10 such sites are found scattered in the sequence (fig. 3). Eight of the sites were found in the region shown to be responsible for protective immunity in mice (Zhu et al. 1997; Peng et al. 2002). One site was the arginine (R) residue of the TGR target site for protein kinase C phosporylation, a site suggested by Zhu et al. (1996*b*) to be present in the *PRA* sequence of *Coccidioides immitis*, based on the report by Woodgett, Gould, Hunter (1986). The TGR site is intact for the two *Coccidioides* species and *Chrysosporium queenslandicum*, but arginine is replaced by alanine (A) in *U. reesii* and histidine (H) in *Chrysosporium lucknowense*. As a consequence, the protein kinase C phosporylation site is disrupted in the two latter species.

Based on the LRT statistics, the selection model fits all the housekeeping data sets significantly better than the neutral model (table 4). Unlike the *PRA* gene, however, the housekeeping genes appear to be under purifying selection. In support of this interpretation, the free parameter (ω_2) of the selection models (Nssites 2) is estimated to be less than 1 for all three genes (table 3). The selection model has a limitation that could mask sites under positive selection. When a gene has a high proportion of slightly deleterious mutations ($0 < \omega < 1$), the free class in the selection model is forced to account for these and any positively selected mutations are then incorporated into the class of neutral sites ($\omega = 1$) (Yang et al. 2000). This situation might apply to the *GAPDH* locus, which has a very small fraction of sites with an ω ratio greater than 1 as judged by the discrete model (Nssites 3). It is unlikely to apply for the *glnA* and *HxkA* loci, however, where the proportion of the sites assumed to be neutral ($\omega = 1$) under the selection model is 0.000 and 0.001, respectively (table 3). Furthermore, no sites belong to a class with ω greater than 1 for any of the other models for the *GlnA* and *HxkA* loci, again indicating an absence of sites under positive selection for these two loci.

The results did not change significantly when using the topology obtained from the *PRA* data set in the tests of selection acting on *PRA*. The model assuming a constant rate of ω across all lineages (M0) could not be rejected for this topology, and although we found minor changes in parameter estimates of the different models of ω over codons, the main results were the same; all models of selection fit the data significantly better than the corresponding neutral models, and the identified sites were found to be the same as in table 3. Thus, the difference in the input topology did not seem to affect the result in this study, which is in accordance with what has been shown previously (Ford 2001; Yang and Nielsen 2002).

Discussion

We did not find any evidence of diversifying selection acting on the *PRA* gene. A very low level of intraspecific variability was found in both the antigenic and housekeeping genes, even in the domains of PRA that have been shown to contain both linear and conformational B-cell reactive epitopes and that account for all the protective immunity in mice obtained by vaccination (Zhu et al. 1997; Jiang et al. 2002; Peng et al. 2002). The observed low levels of intraspecific variability are in accordance with the results obtained from a previous study of the diversity of a part of the *PRA* gene in a few individuals from each of the *Coccidioides* species (Peng et al. 1999) as well as from a study of other gene loci in *Coccidioides*, including a T-cell reactive site of a dioxygenase gene (Koufopanou et al. 2001). Thus, in contrast to proteins of other species shown to be critical in host-pathogen interactions (Shpaer and Mullins 1993; Hughes and Hughes 1995; Endo, Ikeo, and Gojobori 1996; Deitsch, Moxon, and Wellems 1997), the increased selective pressure on the *PRA* gene cannot be attributed to diversifying selection with the aim to escape recognition by the host's immune system. Based on the data presented

here, immunization with PRA epitopes from one isolate is expected to demonstrate protection broadly across the entire species, which supports the effort to use PRA in vaccine development to prevent coccidioidomycosis (Kirkland et al. 1998; Pappagianis 2001).

The observed lack of diversifying selection acting on antigens of *Coccidioides*, as compared with its presence in other human pathogens, may be explained by the difference in their ecology. Whereas *Coccidioides* is a dimorphic pathogen that has to go through its hyphal saprobic phase to produce new infection propagules, many other pathogens are obligate parasites with no part of their life cycle outside the host. For that reason, the obligate parasites can be assumed to be under a much higher evolutionary pressure to overcome the host's immune system. This explanation is supported by the absence of reports showing evidence of diversifying selection in antigens within other dimorphic fungal pathogens. Instead, they have been found to evolve neutrally (e.g., the H antigen in *Histoplasma capsulatum* [Kasuga et al. 2003]) or to be conserved, like the T-cell epitope of antigen gp43 in *Paracoccidioides brasiliensis* (Morais et al. 2000).

We did find evidence of positive selection acting on the *PRA* gene. When analyzing the entire gene sequence of *PRA* and each housekeeping locus, except for the unalignable proline-threonine rich tetrapeptide repeat region of the *PRA* gene, we found that the *PRA* gene is under a higher selective pressure than the housekeeping genes. As outlined above, the nonrepetitive regions of *PRA* show a significantly higher proportion of polymorphic codons with replacement substitutions than any of the housekeeping genes, and the overall value of ω of this part of *PRA* is significantly higher than for any of the housekeeping genes. Furthermore, several sites in the *PRA* gene had a ω value significantly greater than 1, and the gene fits models of positive selection significantly better than the corresponding neutral models.

Typically, genes evolve under purifying selection (Endo, Ikeo, and Gojobori 1996) and variation in ω among sites over time can be attributed to differential selective constraints among sites. The evidence for a higher selective pressure on the nonrepetitive part of the *PRA* gene, compared with the housekeeping genes, must result from relaxed selective constraints, positive selection, or both, associated with individual sites. Using the codon-site search approach (as implemented in PAML) to explore our data, we found that the vast majority of the sites in the housekeeping genes evolve under purifying selection. In contrast, the sites of the nonrepetitive part of *PRA* were found divided into three classes, one class with purifying selection ($\omega = 0$), one indicative of relaxed selective constraints ($\omega = 0.76$), and one indicating positive selection ($\omega = 5.49$).

The expression of *PRA* can be considered phase specific because it is up-regulated during the spherule phase in *Coccidioides* (Galgiani et al. 1992; Peng et al. 1999). The finding of a larger fraction of sites under relaxed selective constraints in a gene that has a limited temporal expression in the life cycle (*PRA*) is analogous to the finding that genes with limited tissue expression are under fewer functional constraints than ubiquitously expressed genes (Hastings 1996; Duret and Mouchiroud 2000). Signals of ω between 0 and 1 can also be a muted signal for positive selection, especially when estimated over a long period, because positive selection can be episodic and followed by purifying selection (Zhang, Rosenberg, and Nei 1998; Schaner et al. 2001).

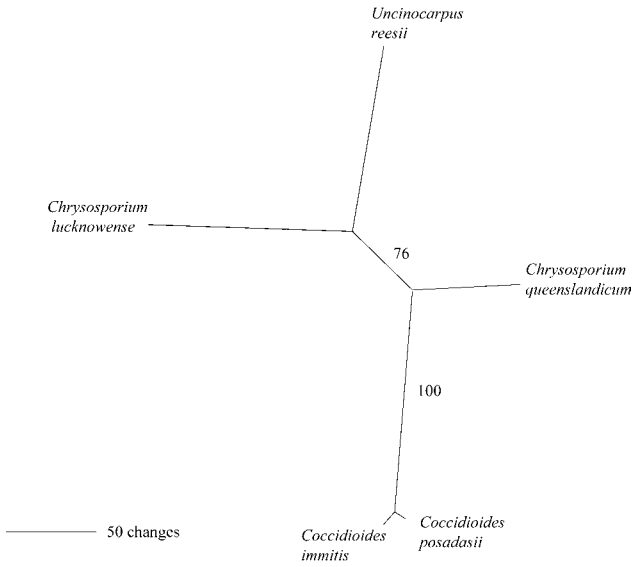

FIG. 2.—Unrooted phylogram of the species included in the study, based on the combined data set from the housekeeping genes. Branches are labeled a to e, and bootstrap values above 75% are indicated by the branches.

Adaptation to a pathogenic life style of the species of *Coccidioides* does not seem to have been associated with accelerated positive selection of *PRA*. This gene is suggested to have an endoglucanase activity and to be involved in morphogenesis of spherules (Zhu et al. 1996*b*). In *Coccidioides*, the conversion to spherules occurs after infecting the host, so the regulation of spherule growth becomes highly relevant to the pathogenesis of the fungus. However, the tests of differential values of ω across branches in the phylogeny did not support an elevated ω on the branch separating the pathogenic *Coccidioides* species from their nonpathogenic relatives.

It is possible that the rapid evolution of *PRA* is driven by positive selection solely as a consequence of the role of PRA in spore morphogenesis in all species included in this study. However, other mechanisms could explain the rapid divergence of the *PRA* gene. As mentioned above, *PRA* apparently is a member of a gene family with up to eight members. Consequently, the acquisition of new functions between these paralogous genes in the ancestor of the fungi studied here could be the explanation to the observed high selective pressure acting on *PRA*. The adaptive evolution of novel protein function is thought to result from a period of relaxed purifying selection immediately after gene duplication, in which mutations that provide the duplicated gene with an advantageous altered function may be positively selected (Ohno 1970; Ohta 1993; Lynch and Conery 2000). However, investigating the possibility of the elevated rate of ω in *PRA* as a consequence of acquisition of new functions after gene duplication is the scope of future research.

Table 3
Likelihood Values, Parameter Estimates, and Sites Under Positive Selection As Inferred Under Six Models of ω over Codons, and Applied to Each of the Four Loci

Locus	Model	lnλ	Parameter Estimate	d_N/d_S	Positively Selected Sites
PRA	One-ratio (Nssites 0)	−1223.40	$\omega = 0.259$	0.259	None
	Neutral (Nssites 1)	−1179.10	$p_0 = 0.702$, $\omega_0 = 0.000$ $p_1 = 0.298$, $\omega_1 = 1.000$	0.298	Not allowed
	Selection (Nssites 2)	−1172.77	$p_0 = 0.694$, $\omega_0 = 0.000$ $p_1 = 0.246$, $\omega_1 = 1.000$ $p_2 = 0.060$, $\omega_2 = 6.120$	0.613	42*, 120*
	Free-ratio (Nssites 3)	−1172.40	$p_0 = 0.676$, $\omega_0 = 0.000$ $p_1 = 0.255$, $\omega_1 = 0.760$ $p_2 = 0.068$, $\omega_2 = 5.49$	0.571	42**, 92*, 120*
	Beta (Nssites 7)	−1179.11	$p = 0.002$, $q = 0.006$	0.300	Not allowed
	Beta&ω (Nssites 8)	−1172.74	$p_0 = 0.939$, $p = 0.011$ $q = 0.0327$ $p_1 = 0.0613$ $\omega = 5.943$	0.602	42**, 120*
GAPDH	One-ratio (Nssites 0)	−1379.00	$\omega = 0.046$	0.046	Not allowed
	Neutral (Nssites 1)	−1378.36	$p_0 = 0.854$, $\omega_0 = 0.000$ $p_1 = 0.146$, $\omega_1 = 1.000$	0.146	Not allowed
	Selection (Nssites 2)	−1366.74	$p_0 = 0.546$, $\omega_0 = 0.000$ $p_1 = 0.044$, $\omega_1 = 1.000$ $p_2 = 0.411$, $\omega_2 = 0.066$	0.071	None
	Free-ratio (Nssites 3)	−1366.43	$p_0 = 0.656$, $\omega_0 = 0.000$ $p_1 = 0.318$, $\omega_1 = 0.110$ $p_2 = 0.026$, $\omega_2 = 1.854$	0.083	92*
	Beta (Nssites 7)	−1368.15	$p = 0.092$, $q = 1.175$	0.067	Not allowed
	Beta&ω (Nssites 8)	−1366.45	$p_0 = 0.975$, $p = 0.258$ $q = 6.138$ $p_1 = 0.025$, $\omega = 1.890$	0.084	92*
glnA	One-ratio (Nssites 0)	−807.97	$\omega = 0.023$	0.023	None
	Neutral (Nssites 1)	−808.31	$p_0 = 0.924$, $\omega_0 = 0.000$ $p_1 = 0.076$, $\omega_1 = 1.000$	0.076	Not allowed
	Selection (Nssites 2)	−801.58	$p_0 = 0.885$, $\omega_0 = 0.000$ $p_1 = 0.000$, $\omega_1 = 1.000$ $p_2 = 0.115$, $\omega_2 = 0.223$	0.026	None
	Free-ratio (Nssites 3)	−801.58	$p_0 = 0.465$, $\omega_0 = 0.000$ $p_1 = 0.420$, $\omega_1 = 0.000$ $p_2 = 0.115$, $\omega_2 = 0.223$	0.026	None
	Beta (Nssites 7)	−801.63	$p = 0.011$, $q = 0.265$	0.024	Not allowed
	Beta & w (Nssites 8)	−801.63	$p_0 = 1.000$, p=0.010 $q = 0.233$ $p_1 = 0.000$, $\omega = 0.408$	0.024	None
hxkA	One-ratio (Nssites 0)	−1232.90	$\omega = 0.020$	0.020	None
	Neutral (Nssites 1)	−1258.05	$p_0 = 0.895$, $\omega_0 = 0.000$ $p_1 = 0.105$, $\omega_1 = 1.000$	0.105	Not allowed
	Selection (Nssites 2)	−1230.95	$p_0 = 0.707$, $\omega_0 = 0.000$ $p_1 = 0.001$, $\omega_1 = 1.000$ $p_2 = 0.292$, $\omega_2 = 0.070$	0.021	None
	Free-ratio (Nssites 3)	−1230.95	$p_0 = 0.621$, $\omega_0 = 0.000$ $p_1 = 0.095$, $\omega_1 = 0.000$ $p_2 = 0.285$, $\omega_2 = 0.073$	0.021	None
	Beta (Nssites 7)	−1231.01	$p = 0.226$, $q = 9.482$	0.021	Not allowed
	Beta&ω (Nssites 8)	−1231.01	$p_0 = 1.000$, $p = 0.226$ $q = 9.488$ $p_1 = 0.000$, $\omega = 4.030$	0.021	None

NOTE.—* Posterior probability > 95% of having $\omega > 1$; ** posterior probability > 99% of having $\omega > 1$.

The central repeat region of *PRA* appears to evolve by quick fixation of the mutations arising in the different species. It is the most divergent part of the *PRA* gene because the five investigated species vary with respect to repeat number, repeat sequence, and the sequence intervening the repeats. Proline-rich regions frequently occur as multiple, tandem repeats and are widely distributed among prokaryotes and eukaryotes (Williamson 1994). Several lines of evidence have suggested that proline-rich regions of cell surface and secreted proteins play important roles in protein structure, as well as in substrate binding (Beguin 1990; Perfect et al. 1998; Staab et al. 1999; Kay, Williamson, and Sudol 2000), by bringing proteins together in such a way that subsequent interactions are

Table 4
Likelihood Ratio Statistics of Different Models of ω Among Codons

Locus	Comparison	df	2Δl
PRA	One-ratio (Nssites 0) versus discrete (Nssites 3)	4	102, $P < 0.001$
	Neutral (Nssites 1) versus selection (Nssites 2)	2	12.7, $P < 0.01$
	Beta (Nssites 7) versus beta&ω (Nssites 8)	2	12.7, $P < 0.01$
GAPDH	One-ratio (Nssites 0) versus discrete (Nssites 3)	4	25.1, $P < 0.001$
	Neutral (Nssites 1) versus selection (Nssites 2)	2	23.2, $P < 0.001$
	Beta (Nssites 7) versus beta&ω (Nssites 8)	2	3.4
glnA	One-ratio (Nssites 0) versus discrete (Nssites 3)	4	12.8
	Neutral (Nssites 1) versus selection (Nssites 2)	2	13.5, $P < 0.01$
	Beta (Nssites 7) versus beta&ω (Nssites 8)	2	0
hxkA	One-ratio (Nssites 0) versus discrete (Nssites 3)	4	3.9
	Neutral (Nssites 1) versus selection (Nssites 2)	2	54.2, $P < 0.001$
	Beta (Nssites 7) versus beta&ω (Nssites 8)	2	0

NOTE.—When statistically significant, the model that fits the data best is underlined.

more probable, rather than providing a structurally defined complex (Williamson 1994; Kay, Williamson, and Sudol 2000). The role of the proline-rich tetrapeptide region in PRA remains to be determined. The extremely high disparity observed between interspecific and intraspecific variability in this region is evidence that this region is not under balancing selection, but suggests that it is of importance for species-specific properties of the protein (e.g., in substrate binding).

A substantial heterogeneity in mode of evolution was found both among and within the genes investigated in this study. The biochemical properties of proteins suggest that the selection pressure should vary both among genes and among amino acid sites within gene, and the analysis of all genes studied here strongly supports these assertions. We found a substantial heterogeneity in the selective pressure acting on the genes, ranging from *PRA* with the highest rate of ω, to *GAPDH* with a small fraction of sites evolving under positive selection and the rest apparently under purifying selection, finally to *hxkA*, in which all sites under

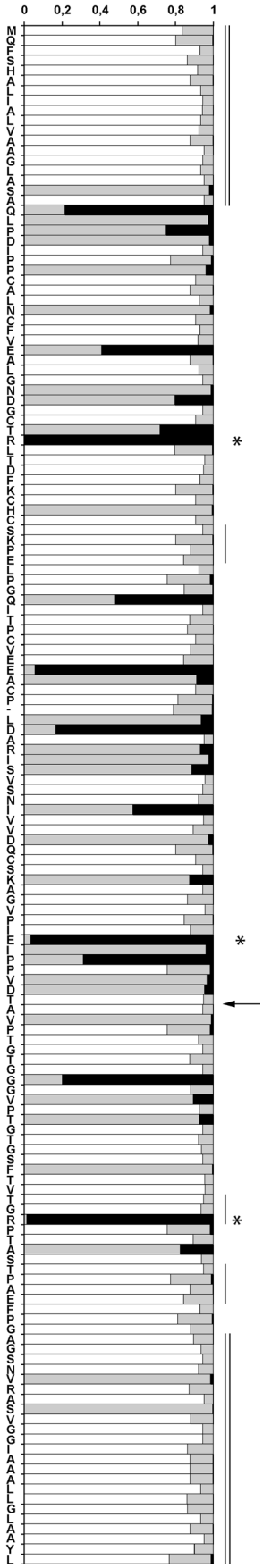

FIG. 3.—Posterior probabilities that sites in the nonrepetitive part of the *PRA* gene belong to site classes with different selective pressures (ω of 5.49 [black], 0.76 [gray], and 0.00 [white bars]) under the free-ratio model. The *PRA* amino acid sequence of *Coccidioides immitis* is shown to the left. Sites with a posterior probability less than 95% to have ω greater than 1 are indicated by an asterisk (*), and the position of the excluded ambiguously aligned repeat part is indicated by an arrow. Double-lined parts are characteristic of signal peptides (Zhu et al., 1996*b*). A protein kinase C phosporylation site (TGR) and two casein kinase II phosporylation sites (SKPE and TPAE) are indicated by single lines.

the free-ratio model (Nssites 3) were found in classes with ω less than 0.08, indicating purifying selection. This diversity of patterns of molecular evolution between genes is in accordance with what has been shown for mammals (Bernardi 1993; Wolfe and Sharp 1993), where differences have been shown to be gene specific (Mouchiroud, Gautier, and Bernardi 1995). Variability in selection also was substantial among sites within the genes. In the *PRA* gene, a high level of heterogeneity was found, whereas the sites of the housekeeping gene fragments evolved more uniformly, as shown by the fact that the model of one-ratio of ω over sites (Nssites 0) could not be rejected for either *glnA* or *hxkA*. Whereas the central repetitive part of *PRA* apparently has evolved very fast, the N and C signal peptides seem to be conserved regions. Signal sequences have emerged as information-rich peptides; based on their structure, they specify different modes of targeting and membrane insertion and even perform functions after being cleaved from the parent protein (reviewed by Martoglio and Dobberstein [1998]). In these species, both the N-terminal and C-terminal signal sequences seem to be functionally constrained and evolve under purifying selection. In the translated, nonrepetitive parts, sites with more or less selective constraints are scattered in the primary sequence, and the sites evolving under purifying selection are not clustered together in the same domain. However, positively selected sites that are scattered in the primary sequence still can be clustered in the crystal structure of the protein, as shown for the major histocompatibility complex (MHC) class I alleles from human populations (Yang and Swanson 2002). One of the sites showing evidence of positive selection in *PRA* is the protein kinase C phosphorylation site in *Coccidioides immitis* that appears to be disrupted in *U. reesii* and *Chrysosporium lucknowense*. Protein phosphorylation is a major mechanism through which hormones and other extracellular agents influence intracellular events such as the regulating the activity of various proteins (Cohen 1982), and our data indicate that selective pressure might act to alter this type of post-translational modification of *PRA* among the species included in this study. The other positively selected sites do not belong to any known active site of the protein. However, changes in inactive sites of a protein can still have a great effect on the protein function (Chen, Greer, and Dean 1995; Jermann et al. 1995), possibly by forcing the main chain to adopt to another conformation, the effects of which may be transmitted to the active site.

Linking molecules and ecology is a fundamental challenge in the study of adaptation, and it requires the integration of several approaches (Golding and Dean 1998). The data presented here highlight the importance of an unbiased analysis of heterogeneous selective pressure and how it can be combined with gene structure and function, as well as ecology of the organism, to understand the evolution of a particular gene of interest.

Acknowledgments

Thanks are due to Elizabeth Turner and Rachel J. Whitaker for useful comments on an earlier draft of the manuscript. Comments received by Scott Edwards and two anonymous reviewers were extremely helpful in improving the original submission. Financial support from Fulbright Commission and Carl Tryggers Stiftelse för Vetenskaplig Forskning to H.J., and from the National Institutes of Health (AI37232) to J.W.T. is gratefully acknowledged.

Literature Cited

Apinis, A. E., and R. G. Rees. 1976. Undescribed keratinophilic fungus from Southern Queensland. Trans. Br. Mycol. Soc. **67**:522–524.

Araujo, F., T. Slifer, and S. Kim. 1997. Chronic infection with *Toxoplasma gondii* does not prevent acute disease or colonization of the brain with tissue cysts following re-infection with different strains of the parasite. J. Parasitol. **83**:521–522.

Beguin, P. 1990. Molecular-biology of cellulose degradation. Annu. Rev. Microbiol. **44**:219–248.

Bernardi, G. 1993. The vertebrate genome: isochores and evolution. Mol. Biol. Evol. **10**:186–204.

Bielawski, J. P., K. A. Dunn, and Z. H. Yang. 2000. Rates of nucleotide substitution and mammalian nuclear gene evolution: approximate and maximum-likelihood methods lead to different conclusions. Genetics **156**:1299–1308.

Braun, B. R., W. S. Head, M. X. Wang, and A. D. Johnson. 2000. Identification and characterization of TUP1-regulated genes in *Candida albicans*. Genetics **156**:31–44.

Burt, A., D. A. Carter, G. L. Koenig, T. J. White, and J. T. Taylor. 1995. A safe method of extracting DNA from *Coccidioides immitis*. Fungal Genet. Newslett. **42**:23.

———. 1996. Molecular markers reveal cryptic sex in the human pathogen *Coccidioides immitis*. Proc. Natl. Acad. Sci. USA **93**:770–773.

Chen, R. D., A. Greer, and A. M. Dean. 1995. A highly active decarboxylating dehydrogenase with rationally inverted coenzyme specificity. Proc. Natl. Acad. Sci. USA **92**:11666–11670.

Cohen, P. 1982. The role of protein phosphorylation in neural and hormonal control of cellular activity. Nature **296**:613–620.

Cox, R. A. 1989. Antigenic structure of *Coccidioides immitis*. Pp. 133–170 *in* E. Kurstak, G. Marquis, P. Auger, D. R. L., and S. Montplaisir, eds. Immunology of fungal diseases. Marcel Dekker, New York.

Cox, R. A., E. Brummer, and G. Lecara. 1977. Invitro lymphocyte responses of coccidioidin skin test-positive and test-negative persons to coccidioidin, spherulin, and a *Coccidioides* cell-wall antigen. Infect. Immun. **15**:751–755.

Crewther, P. E., M. Matthew, R. H. Flegg, and R. F. Anders. 1996. Protective immune responses to apical membrane antigen 1 of *Plasmodium chabaudi* involve recognition of strain-specific epitopes. Infect. Immun. **64**:3310–3317.

Deitsch, K. W., E. R. Moxon, and T. E. Wellems. 1997. Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. Microbiol. Mol. Biol. Rev. **61**:281–293.

Dugger, K. O., J. N. Galgiani, N. M. Ampel, S. H. Sun, D. M. Magee, J. Harrison, and J. H. Law. 1991. An immunoreactive apoglycoprotein purified from *Coccidioides immitis*. Infect. Immun. **59**:2245–2251.

Duret, L., and D. Mouchiroud. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Mol. Biol. Evol. **17**:68–74.

Endo, T., K. Ikeo, and T. Gojobori. 1996. Large-scale search for genes on which positive selection may operate. Mol. Biol. Evol. **13**:685–690.

Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult. 1995. Testing significance of incongruence. Cladistics **10**:315–319.

Fisher, M. C., G. L. Koenig, T. J. White, G. San-Blas, R. Negroni, I. G. Alvarez, B. Wanke, and J. W. Taylor. 2001. Biogeographic range expansion into South America by *Coccidioides immitis* mirrors New World patterns of human migration. **98**:4558–4562.

Fisher, M. C., G. L. Koenig, T. J. White, and J. W. Taylor. 2000. Pathogenic clones versus environmentally driven population increase: analysis of an epidemic of the human fungal pathogen *Coccidioides immitis*. J. Clin. Microbiol. **38**:807–813.

———. 2002. Molecular and phenotypic description of *Coccidioides posadasii* sp nov., previously recognized as the non-California population of *Coccidioides immitis*. Mycologia **94**:73–84.

Ford, M. J. 2001. Molecular evolution of transferrin: evidence for positive selection in salmonids. Mol. Biol. Evol. **18**:639–647.

Galgiani, J. N. 1999. Coccidioidomycosis: A regional disease of national importance—Rethinking approaches for control. Ann. Intern. Med. **130**:293–300.

Galgiani, J. N., S. H. Sun, K. O. Dugger, N. M. Ampel, G. G. Grace, J. Harrison, and M. A. Wieden. 1992. An arthroconidial-spherule antigen of *Coccidioides immitis*—Differential expression during in vitro fungal development and evidence for humoral response in humans after infection or vaccination. Infect. Immun. **60**:2627–2635.

Golding, G. B., and A. M. Dean. 1998. The structural basis of molecular adaptation. Mol. Biol. Evol. **15**:355–369.

Goldman, N., and Z. H. Yang. 1994. Codon-based model of nucleotide substitution for protein-coding DNA-sequences. Mol. Biol. Evol. **11**:725–736.

Hastings, K. E. M. 1996. Strong evolutionary conservation of broadly expressed protein isoforms in the troponin I gene family and other vertebrate gene families. J. Mol. Evol. **42**:631–640.

Herr, R. A., C.-Y. Hung, and G. T. Cole. 2003. A family of proline-rich antigens of *Coccidioides posadasii*. *Abstracts of the general meeting of the American Society for microbiology*. F-106–212.

Huelsenbeck, J. P., J. J. Bull, and E. Cunningham. 1996. Combining data in phylogenetic analysis. Trends Ecol. Evol. **11**:152–158.

Hughes, A. L. 1992. Positive selection and interallelic recombination at the merozoite surface antigen-1 (Msa-1) locus of *Plasmodium falciparum*. Mol. Biol. Evol. **9**:381–393.

Hughes, M. K., and A. L. Hughes. 1995. Natural selection on *Plasmodium* surface proteins. Mol. Biochem. Parasitol. **71**:99–113.

Hughes, A. L., and M. Yeager. 1998. Natural selection at major histocompatibility complex loci of vertebrates. Annu. Rev. Genet. **32**:415–435.

Jermann, T. M., J. G. Opitz, J. Stackhouse, and S. A. Benner. 1995. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. Nature **374**:57–59.

Jiang, C., D. M. Magee, F. D. Ivey, and R. A. Cox. 2002. Role of signal sequence in vaccine-induced protection against experimental coccidioidomycosis. Infect. Immun. **70**:3539–3545.

Kasuga, T., T. J. White, and G. Koenig et al. (18 co-authors). 2003. Phylogeography of the fungal pathogen *Histoplasma capsulatum*. Mol. Ecol. **12**:3383–3401.

Kay, B. K., M. P. Williamson, and M. Sudol. 2000. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. FASEB J. **14**:231–241.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, UK.

Kirkland, T. N., F. Finley, K. I. Orsborn, and J. N. Galgiani. 1998. Evaluation of the proline-rich antigen of *Coccidioides immitis* as a vaccine candidate in mice. Infect. Immun. **66**:3519–3522.

Koufopanou, V., A. Burt, T. Szaro, and J. W. Taylor. 2001. Gene genealogies, cryptic species, and molecular evolution in the human pathogen *Coccidioides immitis* and relatives (Ascomycota, Onygenales). Mol. Biol. Evol. **18**:1246–1258.

Lee, S. B., and J. T. Taylor. 1990. Isolation of DNA from fungal mycelia and single cells. Pp. 282–287 *in* A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, eds. PCR protocols: a guide to methods and applications. Academic Press, San Diego, Calif.

Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. Science **290**:1151–1155.

Maddy, K. T., and T. Crecelius. 1967. Pp. 309–312 *in* D. Ajello, ed. Coccidioidomycosis. University of Arizona Press, Tucson, Ariz.

Magee, D. M., and R. A. Cox. 1995. Roles of gamma interferon and interleukin-4 in genetically determined resistance to *Coccidioides immitis*. Infect. Immun. **63**:3514–3519.

Martoglio, B., and B. Dobberstein. 1998. Signal sequences: more than just greasy peptides. Trends Cell Biol. **8**:410–415.

Miller, S. R. 2003. Evidence for the adaptive evolution of the carbon fixation gene rbcL during diversification in temperature tolerance of a clade of hot spring cyanobacteria. Mol. Ecol. **12**:1237–1246.

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

Mouchiroud, D., C. Gautier, and G. Bernardi. 1995. Frequencies of synonymous substitutions in mammals are gene-specific and correlated with frequencies of nonsynonymous substitutions. J. Mol. Evol. **40**:107–113.

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

Ohno, S. 1970. Evolution by gene duplication. Springer-Verlag, Berlin.

Ohta, T. 1993. An examination of the generation-time effect on molecular evolution. Proc. Natl. Acad. Sci. USA **90**:10676–10680.

Pan, S. C., L. Sigler, and G. T. Cole. 1994. Evidence for a phylogenetic connection between *Coccidioides immitis* and *Uncinocarpus reesii* (Onygenaceae). Microbiology **140**:1481–1494.

Pappagianis, D. 1994. Marked increase in cases of coccidioidomycosis in California—1991, 1992, and 1993. Clin. Infect. Dis. **19**:S14–S18.

———. 1988. Epidemiology of coccidioidomycosis. Curr. Top. Med. Mycol. **2**:199–238.

———. 2001. Seeking a vaccine against *Coccidioides immitis* and serologic studies: expectations and realities. Fungal Genet. Biol. **32**:1–9.

Parmley, S. F., U. Gross, A. Sucharczuk, T. Windeck, G. D. Sgarlato, and J. S. Remington. 1994. 2 alleles of the gene encoding surface-antigen P22 in 25 strains of *Toxoplasma gondii*. J. Parasitol. **80**:293–301.

Peng, T., K. I. Orsborn, M. J. Orbach, and J. N. Galgiani. 1999. Proline-rich vaccine candidate antigen of *Coccidioides immitis*: conservation among isolates and differential expression with spherule maturation. J. Infect. Dis. **179**:518–521.

Peng, T., L. Shubitz, J. Simons, R. Perrill, K. I. Orsborn, and J. N. Galgiani. 2002. Localization within a proline-rich antigen (Ag2/PRA) of protective antigenicity against infection with *Coccidioides immitis* in mice. Infect. Immun. **70**:3330–3335.

Perfect, S. E., R. J. O'Connell, E. F. Green, C. Doering-Saad, and J. R. Green. 1998. Expression cloning of a fungal proline-rich glycoprotein specific to the biotrophic interface formed in the Colletotrichum-bean interaction. Plant J. **15**:273–279.

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

Reboux, G., S. Comparot, V. Kirchgesner, and T. Barale. 1995. A study of 19 species of *Chrysosporium* isolated in Besancon University Hospital (1984–1994). J. Mycol. Med. **5**:105–110.

Renia, L., I. T. Ling, M. Marussig, F. Miltgen, A. A. Holder, and D. Mazier. 1997. Immunization with a recombinant C-terminal fragment of *Plasmodium yoelii* merozoite surface protein 1 protects mice against homologous but not heterologous *P. yoelii* sporozoite challenge. Infect. Immun. **65**:4419–4423.

Rossman, M. G., A. Liljas, C.-I. Branden, and L. J. Banaszak. 1975. Evolutionary and structural relationships among dehydrogenases. Pp. 61–102 *in* P. D. Boyer, ed. The enzymes. Academic Press, New York.

Saubolle, M. 1996. Life cycle and epidemiology of *Coccidiodes immits*. Pp. 1–8 *in* H. E. Einstein, and A. Catenzaro, eds. Coccidioidomycosis. National Foundation of Infectuous Diseases, Washington, DC.

Schaner, P., N. Richards, A. Wadhwa, I. Aksentijevich, D. Kastner, P. Tucker, and D. Gumucio. 2001. Episodic evolution of pyrin in primates: human mutations recapitulate ancestral amino acid states. Nat. Genet. **27**:318–321.

Shpaer, E. G., and J. I. Mullins. 1993. Rates of amino acid change in the envelope protein correlate with pathogenicity of primate lentiviruses. J. Mol. Evol. **37**:57–65.

Staab, J. F., S. D. Bradway, P. L. Fidel, and P. Sundstrom. 1999. Adhesive and mammalian transglutaminase substrate properties of *Candida albicans* Hwp1. Science **283**:1535–1538.

Swofford, D. L. 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 3.1. Illinois Natural History Survey, Champaign, Ill.

Templeton, M. D., E. H. Rikkerink, S. L. Solon, and R. N. Crowhurst. 1992. Cloning and molecular characterization of the glyceraldehyde-3-phosphate dehydrogenase-encoding gene and cDNA from the plant pathogenic fungus *Glomerella cingulata*. Gene **122**:225–230.

Vidal, P., and J. Guarro. 2002. Identification and phylogeny of *Chrysosporium* species using RFLP of the rDNA PCR-ITS region. Studies Mycol. **47**:189–199.

Vidal, P., M. A. Vinuesa, J. M. Sanchez-Puelles, and J. Guarro. 2000. Phylogeny of the anamorphic genus *Chrysosporium* and related taxa based on rDNA internal transcribed spacer sequences. Pp. 22–29 *in* R. K. S. Kushwaha, and J. Guarro, eds. Biology of dermatophytes and other keratinophilic fungi. Revista Iberoamericana de Micologia, Bilbao, Spain.

Vissiennon, T., K. F. Schuppel, E. Ullrich, and A. F. A. Kuijpers. 1999. Case report. A disseminated infection due to *Chrysosporium queenslandicum* in a garter snake (Thamnophis). Mycoses **42**:107–110.

Williamson, M. P. 1994. The structure and function of proline-rich regions in proteins. Biochem. J. **297**:249–260.

Wolfe, K. H., and P. M. Sharp. 1993. Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. J. Mol. Evol. **37**:441–456.

Woodgett, J. R., K. L. Gould, and T. Hunter. 1986. Substrate specificity of protein kinase C—Use of synthetic peptides corresponding to physiological sites as probes for substrate recognition requirements. Eur. J. Biochem. **161**:177–184.

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

———. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. **15**:568–573.

Yang, Z. H., and J. P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. **15**:496–503.

Yang, Z. H., and R. Nielsen. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J. Mol. Evol. **46**:409–418.

———. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. **19**:908–917.

Yang, Z. H., R. Nielsen, N. Goldman, and A. M. K. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155**:431–449.

Yang, Z., and W. J. Swanson. 2002. Codon-substitution models to detect adaptive evolution that account for heterogenous selective pressures among site classes. Mol. Biol. Evol. **19**:49–57.

Yang, Z. H., W. J. Swanson, and V. D. Vacquier. 2000. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. Mol. Biol. Evol. **17**:1446–1455.

Zhang, J., H. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95**:3708–3713.

Zhu, Y. F., V. Tryon, D. M. Magee, and R. A. Cox. 1997. Identification of a *Coccidioides immitis* antigen 2 domain that expresses B-cell-reactive epitopes. Infect. Immun. **65**:3376–3380.

Zhu, Y., C. Yang, D. M. Magee, and R. A. Cox. 1996*a*. Molecular cloning and characterization of *Coccidioides immitis* antigen 2 cDNA. Infect. Immun. **64**:2695–2699.

———. 1996*b*. *Coccidioides immitis* Antigen 2: Analysis of gene and protein. Gene **181**:121–125.

Scott Edwards, Associate Editor

Accepted February 17, 2004