Extensive Trans-Specific Polymorphism at the Mating Type Locus of the Root Decay Fungus *Heterobasidion*

Linda T.A. van Diepen,[†,‡,1] Åke Olson,[†,2] Katarina Ihrmark,[2] Jan Stenlid,[*,2] and Timothy Y. James[*,1]

[1]Department of Ecology and Evolutionary Biology, University of Michigan

[2]Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala, Sweden

[‡]Present address: Department of Natural Resources and the Environment, University of New Hampshire, Durham, NH

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: tyjames@umich.edu; Jan.Stenlid@slu.se.

Associate editor: Naoki Takebayashi

The sequence data reported in this article have been submitted in the GenBank (accession nos. KF280347–KF280390). The coding DNA sequence alignments used for this study have been deposited in the Dryad Repository under http://dx.doi.org/10.5061/dryad.r7nt4.

Abstract

Incompatibility systems in which individuals bearing identical alleles reject each other favor the maintenance of a diversity of alleles. Mushroom mating type loci (*MAT*) encode for dozens or hundreds of incompatibility alleles whose loss from the population is greatly restricted through negative frequency selection, leading to a system of alleles with highly divergent sequences. Here, we use DNA sequences of homeodomain (HD) encoding genes at the *MAT* locus of five closely related species of the root rot basidiomycete *Heterobasidion annosum* sensu lato to show that the extended coalescence time of *MAT* alleles greatly predates speciation in the group, contrasting loci outside of *MAT* that show allele divergences largely consistent with the species phylogeny with those of *MAT*, which show rampant trans-species polymorphism. We observe a roughly 6-fold greater genealogical depth and polymorphism of *MAT* compared with non-*MAT* that argues for the maintenance of balanced polymorphism for a minimum duration of 24 My based on a molecular-clock calibrated species phylogeny. As with other basidiomycete HD genes, balancing selection appears to be concentrated at the specificity-determining region in the N-terminus of the protein based on identification of codons under selection and the absence of recombination within the region. However, the elevated polymorphism extends into the nonspecificity determining regions as well as a neighboring non-*MAT* gene, the mitochondrial intermediate peptidase (*MIP*). In doing so, increased divergence should decrease recombination among alleles and as a by-product create incompatibilities in the functional domains not involved in allele recognition but in regulating sexual development.

Key words: **mating type, incompatibility proteins, balancing selection.**

Article

Introduction

Incompatibility proteins play a fundamental role in structuring genotypes in natural populations by control of genetic admixture during mating or encounters of somatic tissues. Because of this fundamental role, incompatibility proteins are heavily exposed to selection and comprise some of the best examples of both rapid evolution and extensive polymorphism (Swanson and Vacquier 2002; Clark et al. 2006). Some incompatibility systems strongly select for high allele diversity; for example, larger numbers of flowering plant self-incompatibility (SI) alleles maximize the frequency of successful outcrossing during outbreeding (Wright 1939). In systems where allelic diversity is favored, rare alleles have an advantage through negative frequency selection. Negative frequency selection retards allele loss by genetic drift and in doing so may extend coalescence times of alleles deeper than the divergence times between species, causing trans-specific polymorphism of allelic lineages. Trans-species polymorphisms have been widely observed in other proteins under balancing selection such as the major histocompatibility complex (MHC) of vertebrates, SI loci of flowering plants, complementary sex determination in social hymenoptera, and vegetative incompatibility genes in fungi (Figueroa et al. 1988; Richman et al. 1996; Wu et al. 1998; Cho et al. 2006). Trans-species polymorphism is also expected for the highly polymorphic mating incompatibility (*MAT*) proteins in the Basidiomycota fungi (including mushrooms and smuts), which, unlike other fungal mating type genes, are known to be highly polymorphic with multiple alleles and subject to negative frequency selection (May et al. 1999; Billiard et al. 2011; Kües et al. 2011). However, lack of knowledge on the extent of trans-specific polymorphism among basidiomycete MAT proteins currently limits the ability to infer the ages of alleles and their relative rates of turnover.

The multiallelic basidiomycete *MAT* proteins have been most frequently compared to flowering plant SI proteins, to which they display a number of similarities but also some key differences. Both systems are multiallelic homoallelic incompatibility systems in which individuals or gametes sharing alleles are incompatible, favoring the evolution of a large

number of incompatibility alleles (Iwasa and Sasaki 1987; Uyenoyama et al. 2001; Billiard et al. 2011). Both systems involve interactions between two nonhomologous proteins encoded from the same haplotype in which the protein of one of the genes must distinguish homoallelic protein (same haplotype) from one of a large number of other possible compatible haplotypes. The function of plant SI proteins, like many incompatibility proteins, is limited to the control of fertilization through a lock and key type interaction (Charlesworth et al. 2005; McClure and Franklin-Tong 2006). However, mating type proteins of fungi not only determine the compatibility among mating cells, but also function after cell fusion (Brown and Casselton 2001). A high level of postfusion nuclear signaling is necessitated by the heterokaryotic life cycle of most basidiomycete fungi in which after cell fusion the component nuclei exist spatially separated but within the same cell. Postfusion control by *MAT* proteins includes the movement and coordinated division of mated nuclei and the expression of genes limited to sexual reproduction or sporulation (Kües et al. 2002). Other incompatibility proteins beyond fungal *MAT* also appear to function in regulation of development. Examples include MT proteins of Volvocine algae that are known transcription factors (TFs) (Ferris et al. 2010), proteins involved in chordate histocompatibility, such as involved in the fusion of colonial organisms (Nyholm et al. 2006), or complementary sex-determining proteins in hymenoptera which function postfertilization in the development of the embryo into males or females (Gempe et al. 2009). In this manner, some incompatibility proteins have a dual nature in both incompatibility and gene regulation, and are likely to display complex patterns of evolution because selection acts on multiple levels—both rejection of homoallelic partners and cooperation with heteroallelic partners postfertilization.

Basidiomycetes in the Agaricomycetes (mushrooms) with bipolar mating systems have a single *MAT* locus encoding one type of protein, the homeodomain (HD) TFs (Kües et al. 2011). Each mating type haplotype typically encodes pairs of divergently transcribed TFs, an HD1 and an HD2 type (Hiscock and Kües 1999). Mating compatibility is determined postfusion by the formation of heteroallelic heterodimers of HD1 + HD2 from compatible mates. The two types of HD proteins can be distinguished by a 60–75 amino acid region with a DNA binding motif containing three alpha-helices (Kües and Casselton 1992). Both HD1 and HD2 proteins may be divided into multiple regions of known function (fig. 1). The N-terminus typically contains dimerization motifs (Banham et al. 1995), is highly variable in sequence (Badrane and May 1999), and based on domain swapping among alleles imparts allele specificity (Wu et al. 1996; Yue et al. 1997; Yee and Kronstad 1998). The C-terminus of *Coprinopsis cinerea* HD1 proteins encode nuclear localization sequences as well as serine-, proline-, and threonine-rich regions that are putative transactivation domains functioning in sexual developmental regulation by the heterodimer (Asante-Owusu et al. 1996; Spit, et al. 1998).

The HD genes at Agaricomycete *MAT* loci are known to be among the most polymorphic of all loci and have been shown to display all the hallmarks of balancing selection (May et al. 1999). Specifically, frequency dependent selection for rare variants can be detected among DNA sequences as an increase of nonsynonymous substitutions over that expected at random, as observed for positive directional selection. Recurrent positive selection will in turn increase the rate of allele turnover, as new allele specificities are formed and replace older specificities. An outstanding question is whether the constraint requiring postfusion functioning among all pairs of compatible HD alleles should reduce the mutation rate of new fungal *MAT* alleles and slow their rate of turnover. Specifically, because the functions of HD proteins reach beyond a simple lock and key mechanism, will the evolution of a new mating allele require a large number of mutations? Evidence to suggest a slow rate of turnover of *MAT* alleles includes extensive levels of synonymous polymorphism between alleles ($\pi_S > 1$) (Stankis et al. 1992; Badrane and May 1999; James et al. 2006, 2011), suggesting new alleles rarely arise. When differences in amino acid sequences between *MAT* alleles accumulate due to ancient divergence times, functional differences may exist among alleles in their abilities to activate transcription or nuclear signaling. However, this is partially alleviated by the physical separation of HD genes into multiple recombining domains that allows functions that determine allelic specificity to evolve independently from domains that function postfertilization (Badrane and May 1999).

Estimating the rate of turnover of alleles can be facilitated by using sequence comparisons of alleles among a group of related species with known divergence times (Richman and Kohn 2000). Elevated *MAT* polymorphism within Agaricomycete species suggests very ancient divergence times among alleles, but because there has been no demonstration of trans-specific polymorphism, it is unclear just how ancient these divergence times are. The only clear evidence for trans-specific inheritance of alleles in basidiomycetes is for the anther smut (*Microbotryum violaceum*) pheromone receptors of the *MAT* locus as well as genes linked to the pheromone receptors (Devier et al. 2009; Petit et al. 2012). Remarkably the same two *MAT* allele lineages have been maintained for hundreds of millions of years. However, because the smut and rust fungi have primarily biallelic *MAT* loci, trans-specific polymorphism of this nature is equivalent to the maintenance of a shared mating type determining system, such as the ancient sex chromosomal system of the mammals in which an X-specific and Y-specific pair of genes (SOX3 and SRY) diverged from a common ancestor (Waters et al. 2007). Moreover, phylogenies of *MAT* proteins from different genera of basidiomycetes showed no trans-specific alleles (James 2012). Thus, the complete turnover of *MAT* alleles could plausibly occur at a rate faster than speciation. To provide a more solid estimate of rate of turnover of *MAT* alleles, it is thus necessary to work with a closely related set of species. More closely related alleles also facilitate detecting positive selection due to reduced difficulties in amino acid alignment and lower saturation of amino acid and silent substitution.

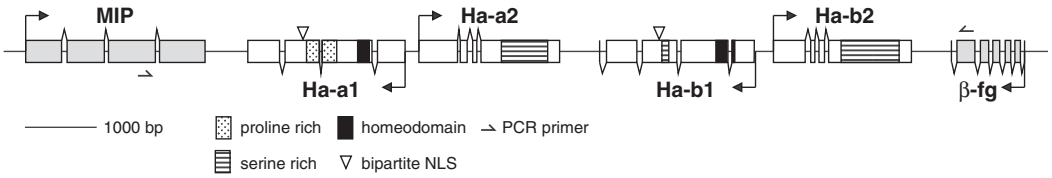

Fig. 1. Schematic of *MAT-A* locus in *Heterobasidion irregulare*. The HD1 genes are *Ha-a1* and *Ha-b1*, which appear to encode a typical HD. The HD2-like genes are *Ha-a2* and *Ha-b2*, which have a very typical gene size and intron–exon structure compared with HD2 in other Agaricomycetes, but lack a detectable HD motif. Two conserved genes flank the *MAT-A* locus in *Heterobasidion* and many other Agaricomycetes (*MIP* and *β-fg*) (James 2007; James et al. 2004). Additional motifs include bipartite NLS motifs, and proline and serine-rich regions presumably functioning in trans-activation. Motif presence and precise structure differed among strains, and only those of TC32-1 are shown. See supplementary figures S1–S4, Supplementary Material online, for alignments of proteins and domain structure.

In this article, we develop the basidiomycete tree pathogen *Heterobasidion annosum* sensu lato as a model system for studying intra and interspecific allele inheritance among multiallelic *MAT* loci. The five *Heterobasidion* species that were formerly all considered *H. annosum* are closely related and have speciated both allopatrically and sympatrically, with species restricted to either North America or Eurasia on a specific set of host taxa (Johannesson and Stenlid 2003; Dalman et al. 2010). *Heterobasidion annosum* sensu stricto (*s.s.*) is restricted to Eurasia, where it is primarily associated with *Pinus*, but can colonize other conifers and deciduous taxa. *Heterobasidion parviporum* and *H. abietinum* are Eurasian taxa associated primarily with *Picea* and *Abies*, respectively (Korhonen and Stenlid 1998). Finally, *H. irregulare* and *H. occidentalis* are restricted to North America on divergent host species (Otrosina and Garbelotto 2010). Importantly, divergence times have been estimated between the taxa and range between 1.3 and 75 Ma depending on calibration of the molecular clock (Dalman et al. 2010) (fig. 2). The species are primarily intersterile, but intersterility is known to be greater between sympatric species pairs than allopatric pairs (Stenlid and Karlsson 1991). Genetic data demonstrate hybridization and introgression from *H. annosum s.s.* into the *H. irregulare* genome in southern Europe following human mediated colonization of the area by *H. irregulare* (Gonthier and Garbelotto 2011).

Heterobasidion annosum s.l. is heterothallic and bipolar with a single *MAT* locus. A complete genome sequence of an isolate of *H. irregulare* (TC32-1) (Olson et al. 2012) demonstrated a region on the largest chromosome with a structure and gene content similar to that of other multiallelic bipolar species such as *Coprinellus disseminatus* (James et al. 2006) with two pairs of divergently transcribed HD genes, that are presumably redundant in function (fig. 1). Genetic analyses confirmed that this gene region co-segregates with mating type and thus represents the entire *MAT* locus (Olson et al. 2012). One aspect of *H. irregulare* that differed from other basidiomycetes is that the DNA-binding HD of the HD2 homolog appeared to be abolished or highly divergent. This finding is surprising because previous studies with the model basidiomycete species *C. cinerea* and *Ustilago maydis* have found the HD2 DNA binding capacities to be more critical than the HD1 motif for function of the heterodimer (Kües, Göttgens, et al. 1994; Asante-Owusu et al. 1996; Schlesinger et al. 1997).

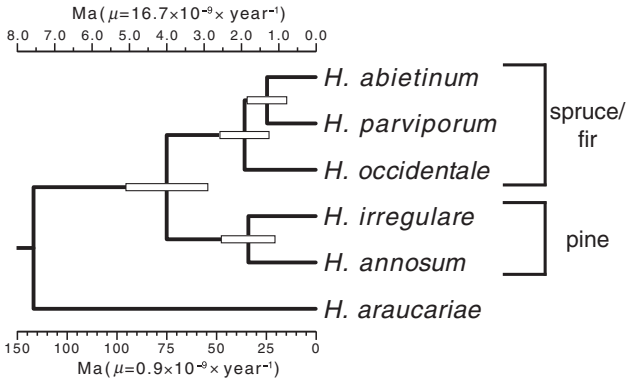

Fig. 2. Species phylogeny of the *Heterobasidion annosum* species complex based on four nuclear loci (Dalman et al. 2010). Divergence times are shown using the both the conservative ($0.9 \times 10^{-9} \times \text{year}^{-1}$) and fast ($16.7 \times 10^{-9} \times \text{year}^{-1}$) mutation rates, with boxes at nodes indicating the 95% highest posterior density of divergence times under the conservative estimate.

The goal of this study was to leverage the genome of *H. irregulare* to investigate the evolution of the *MAT* locus in the species complex. We searched for evidence of the ultimate sign of balancing selection, trans-specific polymorphism, and provide the first in depth study of the phenomenon in the multiallelic basidiomycete *MAT* system. We also demonstrate evidence for both positive selection and recombination acting on gene evolution and identify codons under selection. These analyses are compared with results obtained using marker loci not likely to be subject to balancing selection as well as a gene tightly linked to the *MAT* locus. By comparing the neutral coalescence times with those of the *MAT* genes, we are able to scale the approximate date of divergence and suggest that trans-species polymorphisms of multiallelic *MAT* alleles may be maintained for at least 24 My.

Results and Discussion

The Structure and Function of MAT Genes in *H. annosum s.l.*

Using long distance polymerase chain reaction (PCR), we amplified and sequenced the ≈11 kb MAT locus from 44 strains (table 1). The *MAT* locus of each strain demonstrated the conserved gene number and arrangement shown in figure 1

Table 1. Isolates of *Heterobasidion* spp. used in This Study.

Isolate ID	Geographic Origin	Host	Coll.[a]	IS[b]	Species	Code
Hirr_TC_32-1	Vermont, USA	*Pinus resinosa*	TC	P	*H. irregulare*	21
Hirr_MON113_sci3	Montana, USA	*P. ponderosa*	TC	P	*H. irregulare*	33
Hirr_ORE104_sci13	Oregon, USA	*P. ponderosa*	TC	P	*H. irregulare*	13
Hirr_UM236	Michigan, USA	*P. resinosa*	TJ	P	*H. irregulare*	19
Hirr_UM185	Michigan, USA	*P. resinosa*	TJ	P	*H. irregulare*	23
Hirr_UM202	Michigan, USA	*P. resinosa*	TJ	P	*H. irregulare*	12
Hirr_UM269	Michigan, USA	*P. resinosa*	TJ	P	*H. irregulare*	44
Hirr_UM274	Michigan, USA	*P. resinosa*	TJ	P	*H. irregulare*	22
Hirr_Dlc99_7a_sp C3	Wisconsin, USA	*Pinus* sp.	DL	P	*H. irregulare*	6
Hirr_Dlc99_7a_sp C7	Wisconsin, USA	*Pinus* sp.	DL	P	*H. irregulare*	20
Hocc_B1142	Ajusco, Mexico	*Abies religiosa*	DR	S/F	*H. occidentale*	18
Hocc_Bc_4b-2	British Colombia, Canada	*Tsuga heterophylla*	JS	S/F	*H. occidentale*	17
Hocc_Bc_5b-15	British Colombia, Canada	*T. heterophylla*	JS	S/F	*H. occidentale*	41
Hocc_Faf_10-7	California, USA	*A. concolor*	KK	S/F	*H. occidentale*	24
Hocc_MON_109_sci2	Idaho, USA	*Thuja plicata*	TC	S/F	*H. occidentale*	37
Hocc_ORE_102_sci2	Oregon, USA	*A. concolor*	TC	S/F	*H. occidentale*	3
Hocc_ORE_103_sci14	Oregon, USA	*A. concolor*	TC	S/F	*H. occidentale*	45
Hann_87068/6	Pistoia, Italy	*P. nigra*	PC	P	*H. annosum* s.s.	35
Hann_90231/2	Stolptsky, Belarus	*Juniperus*	KK	P	*H. annosum* s.s.	5
Hann_92153/1	Bayern, Germany	*Picea abies*	OH	P	*H. annosum* s.s.	26
Hann_93014/1	Bayern, Germany	*Pic. abies*	OH	P	*H. annosum* s.s.	9
Hann_93030/1	Bayern, Germany	*Pic. abies*	OH	P	*H. annosum* s.s.	36
Hann_FW2-2	Scotland	*Pic. sitchensis*	JS	P	*H. annosum* s.s.	42
Hann_Sä_16-4	Sätuna, Sweden	*Pic. abies*	JS	P	*H. annosum* s.s.	31
Hann_V5:91_C4	Vedby, Sweden	*Larix eurolepis*	JS	P	*H. annosum* s.s.	34
Hann_V13:53_C2	Vedby, Sweden	*L. eurolepis*	JS	P	*H. annosum* s.s.	27
Hpar_87-124-2	Oslo, Norway	*Pic. abies*	HS	S	*H. parviporum*	28
Hpar_93134/1	Aalborg, Denmark	*Pic. abies*	IT	S	*H. parviporum*	29
Hpar_95151	Ural, Russia	*Pic. abies*	KK	S	*H. parviporum*	32
Hpar_95156	Ural, Russia	*Pic. abies*	KK	S	*H. parviporum*	4
Hpar_Fas_10	Vicenza, Italy	*Pic. abies*	KK	S	*H. parviporum*	43
Hpar_Fas_11	Italy	*Pic. abies*	KK	S	*H. parviporum*	10
Hpar_Fas_13	Vicenza, Italy	*Pic. abies*	KK	S	*H. parviporum*	40
Hpar_HR20	Greece	Unknown	PT	S	*H. parviporum*	11
Hpar_HR32	Greece	Unknown	PT	S	*H. parviporum*	2
Habi_87064/2	Pistoia, Italy	*A. alba*	PC	F	*H. abietinum*	38
Habi_92137/4	Voiron, Isére, France	*Picea or Abies* sp.	KK	F	*H. abietinum*	16
Habi_92144/1	Grande Chartreuse, France	Unknown	KK	F	*H. abietinum*	8
Habi_92179/3	Potenza, Italy	Unknown	PC	F	*H. abietinum*	1
Habi_FA8	France	*A. grandis*	JS	F	*H. abietinum*	7
Habi_FB18	France	*A. grandis*	JS	F	*H. abietinum*	39
Habi_Faf 4-2	Toscana, Italy	*A. alba*	PC	F	*H. abietinum*	14
Habi_Faf 4-6	Toscana, Italy	*A. alba*	PC	F	*H. abietinum*	15
Habi_Faf 5-3	Toscana, Italy	*A. alba*	KK	F	*H. abietinum*	25
Hara_88-9-6	New Zealand	*Agathis australis*	JS	—	*H. araucariae*	30

NOTE.—Code refers to numbers used in figures 3–6.

[a]Collectors: TC, T. Chase; JS, J. Stenlid; DR, D. Rizzo; KK, K. Korhonen; TJ, T. James; DL, D. Lindner.

[b]Intersterility group affinity based on result of pairings with tester strains.

except for isolate Hpar_95156 which lacks the *Ha-b2* gene. The intron–exon structure predicted for all of the gene sequences of the strains was completely conserved with the following exceptions: 17 haplotypes of *Ha-b1* lacked the last intron and the last exon (of 4 amino acids). Hocc_ORE_102_sci2 lacked both the entire first intron and second exon of *Ha-a2*, Hirr_MON113_sci3 lacked the first intron of *Ha-b2*, and Hann_87068/6 and Hann_93030/1 both lacked the second intron of *Ha-b2*.

Our sequencing also revealed evidence of degeneration of *MAT* alleles. All premature stop codons were shown to be sequencing errors once resequenced except for the three

strains Habi_FA8 (*H. abietinum*), Habi_Faf 5-3 (*H. abietinum*), and Hpar_95156 (*H. parviporum*), whose *Ha-a2* genes contained nonsense mutations at codons 42 and 146. Habi_Faf5-3 had an additional nonsense mutation in codon 424. These variants appear to be pseudo-genes. Because the nonfunctional *Ha-a2* allele is linked to a putatively functional *Ha-a1* allele, the haplotype presumably is able to form at least one version of the heterodimer and initiate normal sexual development. The fact that these nonfunctional alleles are found in multiple species suggests that the consequences of loss of function of these genes have minimal impact on the mating abilities of the strains because they have been maintained for a long time. These observations are consistent with the detection of pseudogenes in the *MAT* loci of other Agaricomycetes (Pardo et al. 1996; van Peer et al. 2011) and support a model of redundancy within the *MAT* locus. In a mating between strains with a full complement of *Heterobasidion MAT* genes, four active heterodimers would be formed, creating potentially 4-fold redundancy. It would be of interest to test whether redundancy completely or only partially offsets the lower heterodimer titer produced by haplotypes with nonfunctional variants or missing genes. Variation in the formation of clamp connections observed among heterokaryons (a sign of nuclear compatibility) comprised of different mating types (James et al. 2008) may be hypothesized to correlate with the number of heterodimers produced by pairs of mating types.

Alignments of the *MAT* proteins had a consistent motif structure as shown in supplementary figures S1–S4, Supplementary Material online. Conserved motifs in HD1 proteins included HDs and nuclear localization signals (NLS). The genes *Ha-a2* and *Ha-b2* are positionally homologous to the HD2 HD genes of other basidiomycetes yet they lack the distinctive HD DNA-binding motif found in other basidiomycete HD2 proteins (fig. 1). To further investigate these two proteins, we analyzed their primary sequence for evidence of characteristics of known HD2 proteins, such as a NLS, coiled-coils for protein dimerization, and DNA-binding motifs. Both of the HD2-like proteins had a serine-rich region as observed in other *MAT* HD proteins and assumed to function in transactivation (Tymon et al. 1992; Spit et al. 1998). Both Ha-a2 and Ha-b2 were predicted to be localized to the nucleus based on the programs Cello and WoLF PSORT. Alleles of both Ha-a2 and Ha-b2 proteins appeared to contain coiled-coil motifs in the N-terminus (encoded by exon 1). However, no evidence for DNA-binding motifs was found. Ha-a2 and Ha-b2 are thus predicted proteins that localize to the nucleus, bind with other proteins, and may act as TFs but lack the canonical HD DNA-binding domain of their positional homologues. This observation of an atypical HD2 protein suggests *Heterobasidion MAT* genes may have diverged in function compared with most other mushroom HD loci which otherwise have DNA binding domains in both HD1 and HD2 proteins of the heterodimer (Brown and Casselton 2001). Based on the detected motifs of the HD proteins we can propose the following model: dimerization of the HD1 and HD2-like proteins should allow the complex to enter the nucleus through the bipartite-NLS of the HD1 protein. Binding of specific promoter regions by the HD of the HD1 component of the dimer would then allow the serine-rich putative activation domain (Tymon et al. 1992) of the HD2-like protein to stimulate transcription.

Divergence and Trans-Species Polymorphism Is Highly Elevated at the MAT Locus

The polymorphism at synonymous sites over sequence pairs within each species is on average 50-fold higher at the *MAT* locus than among non-*MAT* genes in all species (table 2). The polymorphism of the *Ha-b* genes is twice as high as it is for the *Ha-a* genes, which is in turn 5- to 10-fold higher than for the mating type-linked mitochondrial intermediate peptidase (*MIP*) gene (table 2).

Gene trees of the *MAT* amino acid sequences revealed a striking amount of trans-species polymorphism evidenced by complete incongruence between the *MAT* protein trees and the estimated species tree (fig. 3). Many nodes that united pairs of *MAT* alleles from different species were observed and were also strongly supported by bootstrap values. Characterizing the amount of trans-species polymorphism as deep coalescences revealed a similar number across the four proteins (table 3). The phylogeny for the Ha-a1 protein sequences, however, contrasts with the other three *MAT* proteins in displaying clades mostly consistent with the earliest divergence in the species tree. Specifically, the pine infecting *Heterobasidion* species (*H. annosum s.s.* and *H. irregulare*) form a clade in this tree while the spruce/fir infecting *Heterobasidion* species (*H. abietinum*, *H. occidentale*, *H. parviporum*) form a separate clade. Strain Hirr_MON113_sci3 labeled 33 and indicated with an arrow in figure 3*A* is the only exception of a pine clade strain grouping among the spruce/fir lineage for Ha-a1.

In contrast, the phylogeny of the non-*MAT* genes showed the expected pattern with most markers showing reciprocal monophyly of species (fig. 4). There were few deep coalescences (table 3) found within the non-*MAT* genes. *MIP*, the gene tightly linked to the *MAT* locus was an exception and displayed 18 deep coalescences and nonmonophyly of each species (fig. 4*A*). However, spruce/fir infecting *Heterobasidion* species formed a clade distinct from the pine infecting *Heterobasidion* species in the *MIP* tree. As with locus *Ha-a1*, strain Hirr_MON113_sci3 was exceptional in its placement among spruce/fir clade sequences.

Our data therefore show a specific pattern of trans-specific polymorphism at the *MAT* locus of *H. annosum s.l.*, with diversification of each *MAT* alleles in each species occurring before the spruce/fir and pine divergence. The extended coalescence time of *MAT* alleles by negative frequency selection can explain the high polymorphism of the *MAT* loci we observed and is consistent with the forces of balancing selection demonstrated to act on *MAT* loci in other basidiomycete species (May et al. 1999; James et al. 2006; Devier et al. 2009; Engh et al. 2010).

Table 2. Estimates of Average Divergence at Synonymous Sites (π_s) of Sequence Pairs within Species.

Species	*a1*	*a2*	*b1*	*b2*	*MIP*	*TF*	*G3P*	*EFA*	*GST1*
Heterobasidion annosum	0.178	0.262	0.653	0.573	0.033	0.009	0.027	0.005	0.009
H. irregulare	0.209	0.267	0.486	0.579	0.032	0.010	0.000	0.021	0.028
H. parviporum	0.217	0.286	0.609	0.600	0.031	0.000	0.000	0.004	0.000
H. abietinum	0.189	0.235	0.452	0.547	0.031	0.000	0.005	0.012	0.008
H. occidentale	0.343	0.338	0.587	0.638	0.032	0.000	0.017	0.000	0.000
Mean	0.227	0.278	0.557	0.587	0.032	0.004	0.010	0.008	0.009

Lack of Introgression at the MAT Locus

An alternative explanation to balancing selection for the pattern of trans-specific polymorphism is that recurrent introgression has led to the introduction of alleles from one species into another, creating a phylogenetic pattern similar to balancing selection with sequences from different species grouping together in a phylogeny. Moreover, because introgression is likely to lead to the introduction of a novel mating type in the recipient species, negative frequency selection could lead to selective introgression of alleles at *MAT* even in the face of a genomic background largely devoid of evidence of introgression (James et al. 1999), as in our case (fig. 4). If the *Heterobasidion MAT* phylogenetic patterns were largely driven by introgression, then it is expected that in some instances very closely related *MAT* sequences from different species would be detected. Identifying closely related sequence at the *MAT* locus requires identifying an expected distribution of sequence divergence within species for comparison. This follows because if introgression is frequent, pairs of related sequences in donor and recipient species would show very recent divergence, as introgressed alleles increase in frequency in the recipient species.

To test whether any pairs of trans-specific alleles were similar enough in divergence to indicate that they may have been introgressed, we created a null expectation using synonymous divergences among alleles within species from the non-*MAT* genes, excluding *MIP* due to an excess of observed deep coalescences. Pairwise synonymous divergence within a species ranged from 0 to 0.074 within species for these 4 genes. Comparison of pairwise synonymous divergence within a species for *MAT* alleles largely exceeded but occasionally ($n = 24/626 = 3.8\%$) overlapped with that of the non-*MAT* genes (fig. 5), while the pairwise synonymous divergence of *MAT* alleles between species almost never overlapped with that of intraspecific divergence ($n = 14/2{,}872 = 0.5\%$), with the most similar pair being *Ha-b1* alleles of Habi_92179-3 (no. 1) and Hpar_HR32 (no. 2) at $dS = 0.014$. As this pair of species is the most recently diverged (fig. 2) and is also sympatric, in is impossible to determine if these highly related alleles are similar due to introgression or lineage sorting. Overall, because the interspecific comparisons of *MAT* allele sequences in general show a large divergence at synonymous sites, these data provide little evidence for a role of introgression in the phylogenetic patterns that we observed.

This absence of detectable introgression is somewhat surprising given the reasonable amount of compatibility in the lab (Stenlid and Karlsson 1991) and the well documented introgression observed in nature following an accidental intercontinental dispersal event (Gonthier and Garbelotto 2011). On the other hand, the sympatric species pair *H. parviporum* and *H. abietinum* are diverging to the point of reciprocal monophyly of many genes (fig. 4), suggesting that hybridization in the field may be very low. It is tempting to hypothesize that the exchange of *MAT* alleles between diverging species is prevented by reproductive incompatibilities of *MAT* proteins through mismatch of the rapidly diverging trans-activation domains and the promoter regions of the target genes.

MAT Genes Show Intragenic Recombination

According to our understanding of recombination within the *MAT* locus of a bipolar species with two redundant subloci (James et al. 2006), recombination should not occur between *Ha-a1* and *Ha-a2* or between *Ha-b1* and *Ha-b2* while recombination between *Ha-a2* and *Ha-b1* should be frequent. By sequencing all of the four loci of *MAT*, we were able to test for cophylogeny of the alleles of pairs of divergently transcribed *MAT* genes. Under a model in which the N-termini of HD1 and HD2 genes of the same subloci encode specificity, recombination between N-termini of the two proteins is not possible. Overall there was similarity in the terminal branches of the phylogenies when comparisons are made between Ha-a1 and Ha-a2 or between Ha-b1 and Ha-b2 trees, though both Ha-b1 and Ha-b2 demonstrated higher bootstrap support for nodes than Ha-a1 and Ha-a2, including support for a number of long internal branches (fig. 3). As expected for loci that are undergoing recombination, the phylogenies among all *MAT* loci were statistically incongruent when assessed using the approximately unbiased (AU) test ($P < 0.001$). In phylogenetic comparisons restricted to solely the N-termini, the phylogenies among all *MAT* loci were again statistically incongruent ($P < 0.005$). Despite the statistically significant topological differences, we identified a high level of phylogenetic consistency between the phylogenies of the N-termini of Ha-b1 and Ha-b2 proteins with 7 clades of identical strain composition for this region (fig. 6). These data suggest that the putative N-terminal specificity regions function together as a unit, as in the HD1 + HD2 heterodimer model (Kües, Asante-Owusu, et al. 1994), and that recombination is largely if not entirely suppressed in this region.

Comparison of phylogenies for a single protein divided into two regions, N-terminal and C-terminal, suggested

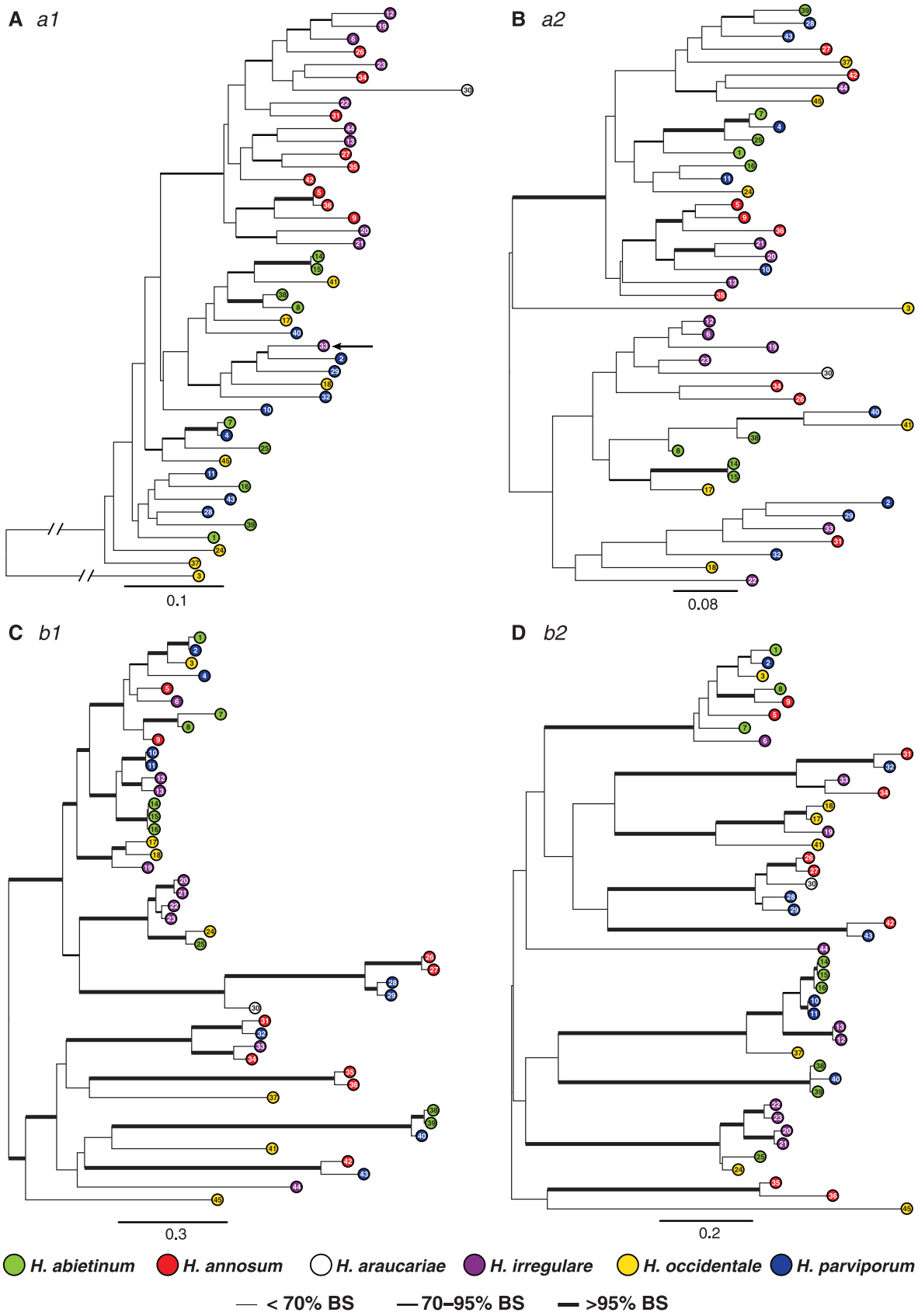

FIG. 3. Phylogenies of the four mating proteins show extensive trans-species polymorphism. Bolded branches represent those with bootstrap support >70%. Branch lengths leading to strain 3 in (A) have been shortened 4×. Arrow in (A) indicates strain discussed in text that is the only one from the pine clade that groups among the spruce/fir clade. See table 1 for additional information, including the code used for each strain.

Table 3. Polymorphism of Loci in *Heterobasidion annosum s.l.*

Locus	*n*	*N*	π_n	π_s	DC	4SD
a1	45	1,962	0.092	0.255	63	0.965 (0.124)
a2	42	1,518	0.170	0.276	65	0.258
b1	45	1,413	0.265	0.601	59	0.580
b2	44	903	0.302	0.616	64	0.611
MIP	45	627	0.009	0.066	18	0.601 (0.113)
EFA	40	234	0.005	0.075	0	0.036
G3P	44	336	0.000	0.114	1	0.157
GST1	45	297	0.016	0.115	5	0.088 (0.070)
TF	45	294	0.001	0.044	0	0.083 (0.004)

NOTE.—*n*, number of sequences analyzed; *N*, number of basepairs (excluding poorly aligned regions); π_n, average diversity of nonsynonymous sites per nucleotide; π_s, average diversity of synonymous sites per nucleotide; DC, number of deep coalescences (trans-species coalescence); 4SD, genealogical depth (under a molecular clock) at 4-fold silent sites of amino acids that are invariant over all sequences. 4SD for *a1* is shown excluding strain Hocc_ORE_102_sci2 in parentheses. Values in parentheses for *MIP*, *GST1*, and *TF* are without *H. araucariae*. Loci *EFA* and *G3P* are missing data for *H. araucariae*.

recombination or gene conversion within genes was common but also showed that the number of deep coalescences was lower in the C-terminus relative to the N-terminus for only the HD1 proteins and not the HD2-like proteins. Specifically, the number of deep coalescences for the proteins were as follows: Ha-a1 (N-term)= 73, Ha-a1 (C-term)= 43, Ha-b1 (N-term)= 71, Ha-b1 (C-term)= 52, Ha-a2 (Nterm)= 71, Ha-a2 (C-term)= 74, Ha-b2 (N-term)= 68, and Ha-b2 (C-term)= 63. Ha-a1, which was distinctive in having a protein phylogeny that grouped according the species phylogeny when using the full protein sequence (figs. 2 and 3*A*), showed a species-specific pattern only for the C-terminus and not the N-terminus when the phylogenies were estimated separately (supplementary fig. S5, Supplementary Material online). These data show that recombination is frequent within the *Ha-a1* gene and that trans-specific polymorphism is greater in the N-terminus where specificity is likely determined. Recombination between the N-terminal region of the protein under balancing selection and the C-terminus would separate incompatibility functions from those that function in downstream events such as trans-activation of sexual genes (Badrane and May 1999). Nonetheless, the C-terminus of the *MAT* proteins displays very high polymorphism as well as large numbers of deep coalescences. One possibility is that the C-terminus is also under balancing selection, perhaps by interacting with specific residues of the N-terminal region. Alternatively, recombination between N- and C-termini may be infrequent such that diversity is maintained by tight linkage to the region under balancing selection. Our finding that trans-specific polymorphism has spread from the presumed specificity region of $\approx$150 amino acids into the neighboring trans-activating domain may suggest that the local recombination suppression needed to maintain a self-incompatible haplotype has led to an expansion of recombination suppression resulting in a pattern of trans-specific polymorphism much larger than expected solely on the basis of linkage without recombination suppression (Wiuf et al. 2004; Uyenoyama 2005).

We scanned for evidence of recombination or gene conversion using the programs GENECONV and MAXCHI implemented in RDP4. No intragenic recombination events were detected in any of the non-*MAT* genes including the *MIP* gene. On the other hand, we detected multiple recombination events for each of the *MAT* locus genes. The numbers of detected recombination or gene conversion events for the *MAT* genes were as follows: 11 at *Ha-a1*, 16 at *Ha-a2*, 6 at *Ha-b1*, and 6 at *Ha-b2* (supplementary fig. S6, Supplementary Material online). The recombination break points were usually found between the domains identified in the proteins and, most importantly, not within the putative specificity determining N-terminal region (supplementary fig. S6, Supplementary Material online). Most of the recombinants consist of parental sequences from isolates belonging to the same intersterility group, but there are a few exceptions, including examples of recombination between isolates present on different continents (supplementary fig. S6, Supplementary Material online). This indicates that the detected recombination events are older than speciation, and it partially contradicts any model in which residues of the C-terminal region interact in an allele-specific manner.

Evidence for Positive Selection at MAT

Evolutionary analyses were conducted to identify codons under positive selection in the genes. Codeml implemented in the PAML package (Yang 2007) was used for *MIP* and the other non-*MAT* loci. The log likelihood values and parameter estimates under the different models allowing variable ω (dN/dS) ratios among sites for the *MIP* and non-*MAT* loci are listed in table 4 and supplementary table S1, Supplementary Material online. For the four non-*MAT* genes, all model comparisons suggested no action of positive selection (table 4). The model M8 (selection) fit the *MIP* gene data set better than the M7 (neutral) model while M2 was not better than M1 (table 4). However, no sites were identified to be under positive selection in the *MIP* gene (supplementary table S1, Supplementary Material online), but both model M2 and M1 suggest that approximately 1.7% of the sites fall in the categories of sites neutrally evolving with an ω close to 1 (supplementary table S1, Supplementary Material online).

For the *MAT* loci, we performed a Bayesian inference on ω using omegaMap (Wilson and McVean 2006), because this method allows for the identification of positively selected codons in sequences even in the presence of recombination. The model parameters (table 5) show higher substitution rate and higher insertion–deletion rate at the *b* loci compared with the *a* loci. Two and three gene regions were detected with a posterior probability >90% of positive selection in *Ha-a1* and *Ha-a2*, respectively (fig. 6). The numbers of codons under positive selection, posterior probability more than 95%, were 20 and 8 for the genes *Ha-a1* and *Ha-a2*, respectively (supplementary table S2, Supplementary Material online). A majority of codons under selection were found in the N terminal region of *Ha-a1* and in both in the N-terminus and within the serine-rich region of *Ha-a2* (fig. 7). No codons

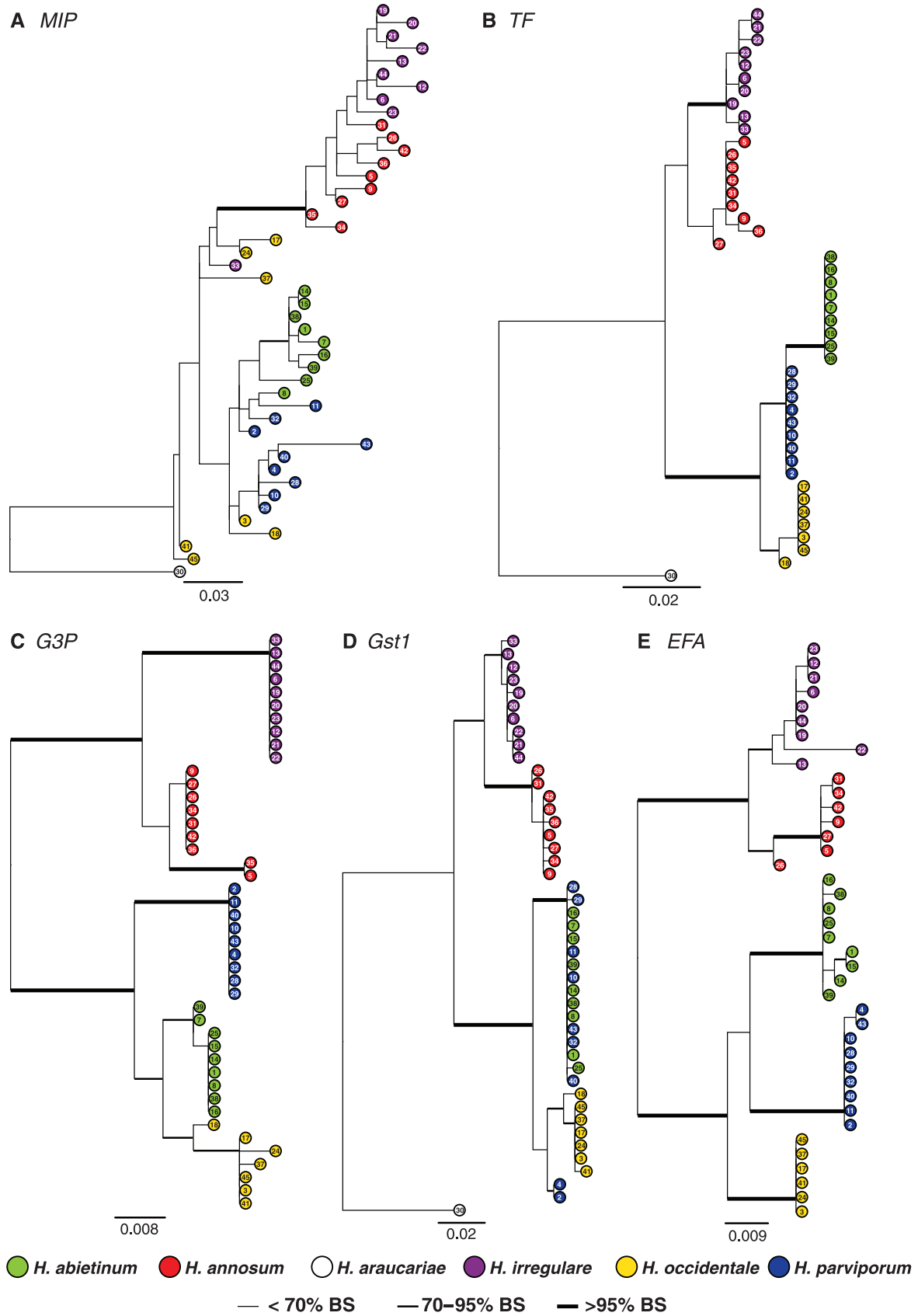

FIG. 4. Phylogenies of the four nonmating type proteins are consistent with the species phylogeny. Bolded branches represent those with bootstrap support >70%. See table 1 for code used for each strain.

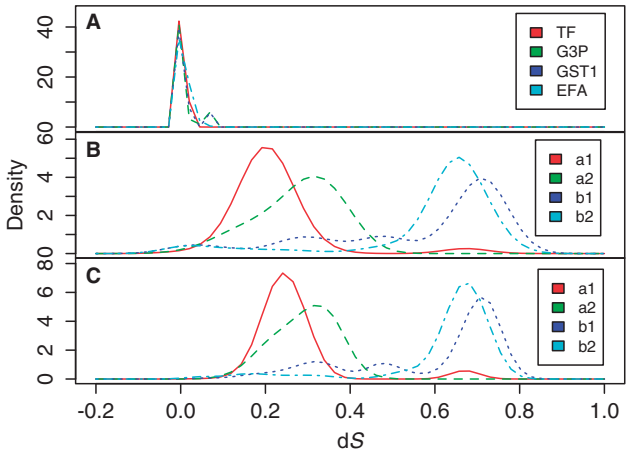

Fig. 5. Distribution of synonymous differences between sequences at the non-*MAT* and *MAT* loci provides no evidence of introgression. (*A*) Distribution of synonymous differences between intraspecific sequence pairs for the non-*MAT* genes. (*B*) Synonymous differences between intraspecific sequence pairs for the *MAT* genes. (*C*) Synonymous differences between interspecific sequence pairs for the *MAT* genes. The values were not computed using *MIP* because of the extensive deep coalescences observed at this locus (table 3).

under selection were identified for the *Ha-b1* and *Ha-b2* genes, and this result agreed with that calculated using codeml without considering recombination (data not shown).

The codons evolving under positive selection in this study are similar in location to those observed in the inky cap mushroom *C. cinerea* (Badrane and May 1999), namely amino acids the N-terminal specificity region that are likely involved in recognition by heterodimer formation. However, we have also identified codons in the C-terminal region that also appear to be under positive selection. What is the effect of the heightened polymorphism of the C-terminal regions of the *MAT* proteins on mating? Pine and spruce infecting lineages show high but not complete incompatibility (Stenlid and Karlsson 1991), and the hypothesis that directional selection in the C-terminus reduces inter-species compatibility could be tested using crosses of strains that have C-termini more similar to those in other species, such as strain Hirr_MON113_sci3 (fig. 3*A*). Regardless of the cause of hyperpolymorphism of the C-termini of the *MAT* loci, it seems very plausible that not all heterodimers will perform the same function, leading to a possible means of sexual selection among mating types.

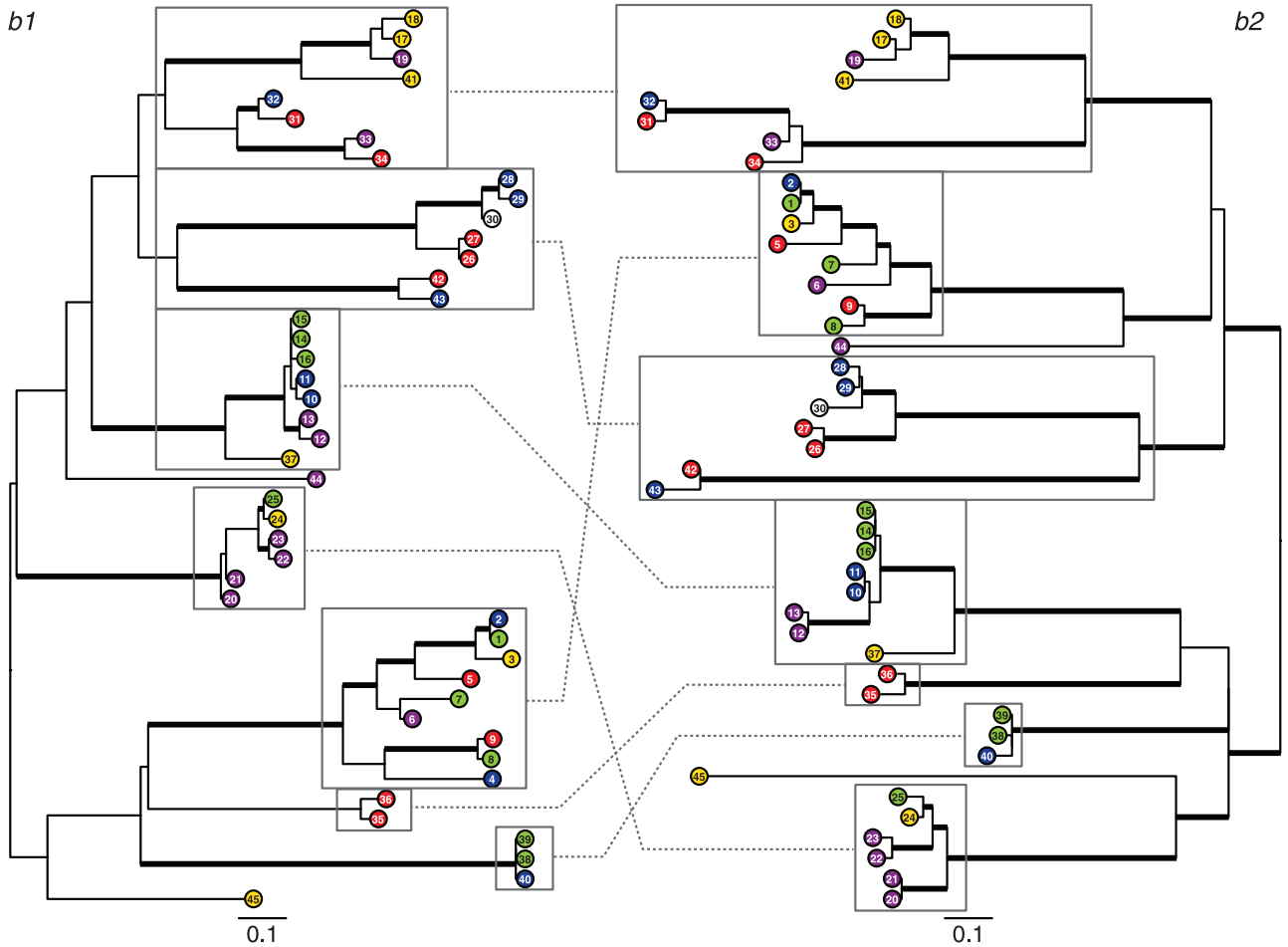

Fig. 6. Comparison of the phylogenies of Ha-b1 and Ha-b2 based solely on the N-terminus of the proteins indicates consistent clades in the two linked proteins. Bolded branches represent those with bootstrap support >70%. See table 1 for code used for each strain. Boxes indicate identical clades found in the two trees.

Table 4. Likelihood Values for Different Codon-Substitution Models the Mating Type, *MIP* and Non-*MAT* Genes and Likelihood Ratio Statistics (Delta L) for Model Fit for the Mating Type, *MIP* and Non-*MAT* genes.

	M0	M1	M2	M2 vs. M1	M7	M8	M8 vs. M7
MIP	−1,707.221	−1,680.410	−1,680.410[a]	0	−1,688.148	−1,680.259	15.78*
EFA	−392.496	−392.495	−392.496	0	−392.501	−392.501	0
GST1	−644.343	−644.343	−644.342	0	−644.336	−644.337	0
G3P	−582.415	−582.401	−582.400	0	−582.400	−582.401	0
TF	−559.882	−559.882	−559.882	0	−559.894	−559.894	0

[a]Proportion of sites in the category $\omega > 1$ is zero.
*Significant at $P < 0.001$ level with df = 2.

Table 5. Model Parameters for the *MAT* Loci Inferred by Bayesian Analysis of Recombination and Substitution Using OmegaMap.

	θ_{syn}	k	φ
Ha-a1	0.87	4.52	4.56
Ha-a2	1.08	5.84	2.74
Ha-b1	3.36	4.06	14.99
Ha-b2	3.58	4.34	9.52

NOTE.—θ_{syn} is the per base synonymous mutation rate per base pair, k is the transition-transversion ratio, and φ is the insertion/deletion rate.

Estimating the Ages of the MAT Alleles

Statistics that measure the shape of a genealogy can be used to formulate hypotheses on rates and causes of allele turnover (Uyenoyama 1997; Richman and Kohn 1999). Dramatic changes in population sizes, such as recovery from a severe bottleneck, create a star-like phylogeny with longer than expected terminal branches (Uyenoyama 1997). Terminal branch lengths are those connecting leaves, or sequences, to the first internal node. The statistic R_{SD} can be used to measure whether the ratio of the sum of terminal branches to the depth of the genealogy (i.e., earliest divergence time) is longer than expected under the neutral coalescent, where R_{SD} is expected to be 1.25.

Estimation of R_{SD} for our data required constructing phylogenies of each of the five species with molecular clock-enforced branch lengths. Two of the *MAT* gene data sets failed a clock-like rate of evolution (table 6). Although most values of R_{SD} exceeded the neutral coalescent expectation, only the *Ha-b2* locus of *H. occidentale* demonstrated an R_{SD} value significantly greater than expected under neutrality (table 6). In our analyses, R_{SD} values of *H. abietinum* differed from the other four species, with a mean R_{SD} value of 0.95. The lower value in *H. abietinum* is likely due to repeated *MAT* alleles that shorten the average terminal branch lengths. Among the other four species, mean values of R_{SD} are similar and show slightly longer terminal branches than expected under neutrality. The similar values of R_{SD} suggest that none of the species have undergone dramatic changes in population size relative to each other, consistent with similar levels of neutral variation across species, for example, π_s values (table 2). Two explanations for the general excess in terminal branch lengths in four of the five species are either a shared population bottleneck in the ancestral population or the specific retention of divergent alleles over time. Significantly excessive R_{SD} was also observed for the *b1* MAT locus of the inky cap mushroom *C. cinerea*, which showed an $R_{SD} > 5$ (May et al. 1999), for MHC class II *Ab* genes in *Mus musculus* with an $R_{SD} = 4.2$ (Richman and Kohn 1999), and for SI genes in the Solanaceae with an $R_{SD} > 5.4$ (Richman and Kohn 2000). This consistent recovery of longer than expected terminal branches at loci subject to balancing selection across three kingdoms of eukaryotes suggests that the maintenance of divergent alleles at incompatibility loci may be common. A possible explanation for why this occurs is that heterozygotes possessing divergent alleles may show a stronger recognition response or may show increased heterozygote advantage, as the case expected for MHC genes (Penn et al. 2002).

Finally, we attempted to quantify more precisely how long alleles have been maintained at *MAT* relative to those evolving without balancing selection. To do so, we specifically focused on comparing sites across these two gene partitions that are absent from selection. We compared the genealogical depth of gene phylogenies based solely on 4-fold degenerate codon positions (4SD) at invariant amino acids (table 3). Results are comparable with π_s and reveal values between 0.258 and 0.965 for the *MAT* genes and 0.036–0.157 for the non-*MAT* genes. The 4SD for *MIP* was 0.601, however eliminating the divergent sequence *H. araucariae* reduced 4SD at *MIP* to 0.113. Deleting outliers Hocc_ORE_102_sci2 and *H. araucariae* from the analyses provides a conservative estimate that the genealogical depth of the *MAT* genes is extended ≈5.9-fold over that of the neutral genes. As the age of the *H. annosum s.l.* group is estimated to be 4.0–75 My old (Dalman et al. 2010), these data suggest that *MAT* alleles may have remained polymorphic for at least 24–450 My. This range of dates spans an order of magnitude because the molecular dating work of Dalman et al. (2010) used a neutral mutation rate of $0.9–16.7 \times 10^{-9}$ for introns of four nuclear genes. These rates were obtained from a published meta-analysis of synonymous sites of coding regions in Eurotiomycete fungi. Because this group of fungi is unrelated to Agaricomycetes, and mutation rates are often greater at shorter introns than those observed at synonymous sites (Kasuga et al. 2002; Halligan et al. 2004), we favor the faster clock calibration and the younger *H. annosum* divergence of 4 My as a conservative estimate. These results also do not require a divergence of *MAT* alleles in *H. annosum* that predates

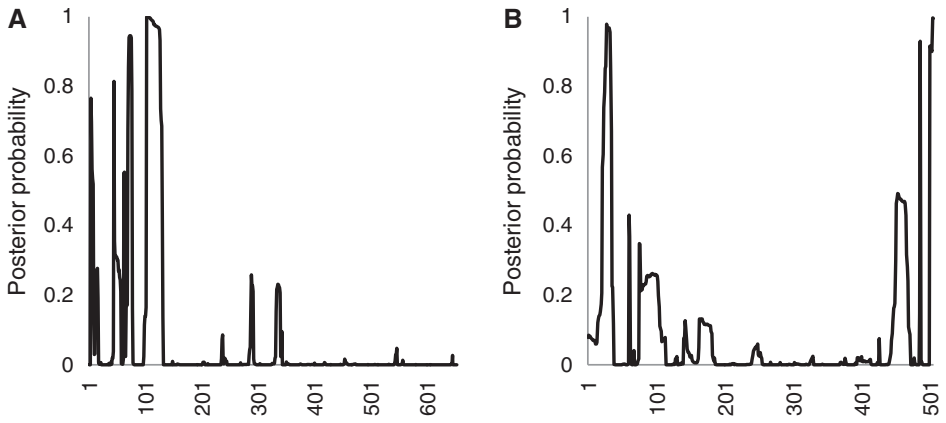

FIG. 7. Posterior probabilities of positive selection ($\omega > 1$) for the (*A*) *Ha-a1* and (*B*) *Ha-a2* genes using a block model. The *X* axis shows the codon position in the Gblocks alignment, and the corresponding amino acid position in strain TC32-1 is given in supplementary table S2 and in supplementary figures S1 and S2, Supplementary Material online.

Table 6. Values of the Relative Terminal Branch Lengths to Coalescent Depth (R_{SD}) Computed for the N-Termini of *MAT* Proteins.

Locus	*Ha-a1*	*Ha-a2*	*Ha-b1*	*Ha-b2*	Mean
Heterobasidion abietinum	0.59	0.70	1.26	1.24[a]	0.95
H. annosum	2.33	1.93	1.68	1.55	1.88
H. irregulare	2.48	1.88	2.88	2.52	2.44
H. occidentale	2.39[a]	1.59	2.66	3.26*	2.47
H. parviporum	2.44	1.72	2.32	2.75	2.31

[a]Data sets that failed the test for a clock-like constant rate of evolution over the tree.

*Significantly different from the neutral coalescent at $P < 0.05$.

the evolution of the Agaricomycetes, some 300 Ma (Floudas et al. 2012).

Conclusion

Many mushroom fungi typically have large population sizes and geographic patterns of speciation, both of which are expected to increase the probability of observing trans-species polymorphism at loci under strong negative frequency selection (Vilgalys and Sun 1994; James et al. 1999). Why has it taken so long to provide definitive evidence of trans-species polymorphism at *MAT* in mushroom fungi? First, experimental means to access the DNA sequences have been hard to secure because of their high polymorphism and uncertain genetic structure. We now have ample genomic evidence that suggests most of the *MAT* loci in both bipolar and tetrapolar species share a high degree of gene order conservation, allowing assessment of homology among sequences (James et al. 2011; van Peer et al. 2011). In this study, by using flanking non-*MAT* genes as the anchors for PCR primers, we are able to unequivocally assign our sequences to homologous genes. A second reason that trans-species polymorphism has seldom been recorded in Agaricomycetes is that most studies have focused on individual model species, whereas this is the only one to study multiple species of close enough divergence to show *MAT* allele turnover on a time scale greater than that of species divergence.

By choosing a group of five closely related species, we were able to capture *MAT* allele divergences that predated speciation. Here, we showed that many of the *MAT* locus alleles diverged before the first divergence in the species complex (the pine and spruce/fir split). Our evidence for balancing selection acting specifically on the *MAT* genes is clear when comparing the flanking genes, which show much lower polymorphism than the *MAT* genes (table 3). Using neutral variation at 4-fold synonymous sites, we estimate that the divergence of the *MAT* genes began at a time point roughly 6 times the age of the divergence of the species complex. These results suggest that mushroom *MAT* loci should be included amongst those examples of where trans-species polymorphism is the rule rather than exception, such as SI alleles in plants and MHC alleles in vertebrates (Figueroa et al. 1988; Richman et al. 1996).

Materials and Methods

Isolate Collection and Culture Methods

We utilized 45 homokaryotic isolates of *Heterobasidion* with approximately equal representation of the five species (table 1). All of the European and many of the North American isolates have been used in a wider phylogeographic study (Dalman et al. 2010). *Heterobasidion* species were assigned on the basis of geographic origin, tree host, and intersterility group (table 1), and these results are confirmed by phylogenetic placement. Homokaryotic strains from Michigan were generated in this study from dilution plating of basidiospores onto malt extract agar (MEA) containing 1.5% malt and 1.5% agar, from which single basidiospore isolates were selected. Basidiospore isolates were inspected for clamp connections to ensure isolates were homokaryotic. Our sequence analysis of the isolates predicted that only Habi_Faf4_2 and Habi_Faf4_6 would share mating types, and the incompatibility of the two isolates was confirmed by pairing the strains and scoring the cross for signs of positive mating (e.g., clamp connections).

Molecular Methods

Each isolate was grown on the surface of a 9 cm Petri dish of 1.5% MEA overlain with cellophane until the isolate had colonized most of the surface. Fresh mycelia was collected and ground in 2X cetyltrimethylammonium bromide buffer using a microcentrifuge pestle or glass beads (3 mm diameter). DNA was extracted from the homogenate following established protocols (James et al. 2008). Amplification of the *MAT-A* locus was accomplished by long distance PCR from the *MIP* to the beta-flanking gene (*β-fg*) genes (fig. 1) using the enzyme mix LA Taq (TaKaRa). Amplification conditions included 2.5 mM $MgCl_2$, 0.4 mM dNTPs, 0.5 μM each of forward and reverse primer, 5% DMSO, and 1.25 units of LA Taq in a 25 μl reaction. The primer sequences were Ha-MAT-A-F (5′-GTTGACCCTTTTCCACGAGA-3′) and Ha-MAT-A-R (5′-AACAAGCAGGATGCAGTCAA-3′). The PCR amplification profile was an initial denaturation step of 1 min at 94 °C followed by 35 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 6 min and a final extension step of 72 °C for 10 min. Successful amplification products were either cloned into PCR-XL-TOPO (Invitrogen) or sequenced directly following treatment with ExoSAP-IT (USB). Amplification of non-*MAT* genes encoding a TF, glutathione-*S*-transferase 1 (*GST1*), elongation factor 1-α (*EFA*), and glyceraldehyde 3-phosphate dehydrogenase (*G3P*) were conducted according to Dalman et al. (2010).

Products that were cloned were first gel purified using crystal violet staining following by DNA purification using the QIAquick Gel Extraction Kit (Qiagen). Transformation utilized One Shot TOP10 (Invitrogen) electrocompetent cells. From each sample one clone was selected, grown in 2 ml of LB broth with 50 μg/ml kanamycin at 37 °C at 250 rpm overnight, and plasmid DNA was extracted using the QIAprep Spin miniprep Kit (Qiagen). The *MAT-A* clones were then used to generate random transposon insertion libraries. Transposons encoding chloramphenicol resistance were randomly inserted into the plasmids using the HyperMu <CHL-1> Insertion kit (Epicentre Biotechnologies). Plasmid DNA was extracted from 24 putative clones using a QIAprep 96 Turbo Miniprep Kit and sequenced using transposon specific primers supplied with the HyperMu <CHL-1> Insertion kit. Sequencing was performed using an ABI3730 at the U. Michigan DNA Sequencing Core or at Macrogene Inc. (Seoul, Korea).

For PCR products that were not cloned or for gap-filling of the DNA sequences of clones from the transposon insertion libraries, we utilized a primer walking strategy. Primers were designed using the software Primer3 v0.4.0 (Rozen and Skaletsky 2000). In the case of several proteins, we identified potential nonsense mutations. To distinguish potential nonsense mutations of pseudo-genes from sequencing errors, we re-sequenced the haplotypes by PCR amplification from genomic DNA and Sanger sequencing with custom primers.

Gene Assembly and Annotation

Assembly of the complete *MAT-A* locus sequence from the Sanger sequencing reads for each strain was accomplished using Sequencher v5.10.1 (Gene Codes Corp.) or SeqMan Pro (DNASTAR Lasergene 9). Complete *MAT-A* sequences consist of four mating type genes *Ha-a1*, *Ha-a2*, *Ha-b1*, and *Ha-b2* and are flanked by the nonmating type specific genes *MIP* and *β-fg* (fig. 1). Using conservation of sequence location, intron positions, and similarity to sequences deposited in GenBank and the full genome of *H. irregulare* assembly v2.0, we extracted the coding sequences from each assembled *MAT-A* haplotype and translated the coding sequences into hypothetical protein sequences. The alignments were divided into N and C terminal domains, which roughly correspond with specificity determining and nonspecificity determining domains, respectively (Badrane and May 1999). For HD1 containing proteins Ha-a1 and Ha-b1, the N terminus represents the amino acids N-terminal to the HD region, and the C-terminus the remainder of the protein. For proteins Ha-a2 and Ha-b2, the N terminus represents the amino acids encoded by the first exon (fig. 1), and the C-terminus the rest of the protein. Identification of domains in each of the HD proteins was tested using WoLF PSORT (Horton et al. 2007), Coils (Lupas et al. 1991), Motif Scan (Sigrist et al. 2010), and Cello (Yu et al. 2003).

Alignment and Phylogenetic Analysis

For each of the four genes of the *MAT-A* locus, the translated amino acid sequences were aligned using Muscle (Edgar 2004) followed by manual refinement in MacClade v4.08 (Maddison and Maddison 2000) or in MEGA v5.05 (Tamura et al. 2011). Tranalign (http://emboss.bioinformatics.nl/cgi-bin/emboss/tranalign, last accessed August 1, 2013) was then used to align the coding sequences according to the amino acid alignments. Prior to all the following analyses, the ambiguously aligned regions of each protein were masked using Gblocks (Castresana 2000). We determined the most appropriate model of protein evolution for each data set using Prottest (Abascal et al. 2005). The best model for Ha-a1 was JTT + G + F, Ha-a2 was JTT + I + G, and JTT + I + G + F for Ha-b1 and Ha-b2. Models for nucleotide analyses of the non-*MAT* genes and the flanking *MIP* gene were performed using jmodeltest v0.1.1 (Posada 2008). Best models were TIM1 + I + G, TIM2, TMP1, K80 + G, and K80 for the *MIP*, *TF*, *G3P*, *GST1*, and *EFA* genes, respectively. Gene trees were constructed using PhyML 3.0 with support estimated by 1,000 bootstrap pseudo-replicates. Phylogenetic incongruence at *MAT* loci was tested by the AU test implemented in CONSEL (Shimodaira and Hasegawa 2001) using site-wise log-likelihoods for each tree estimated by PAML 4.5 (Yang 2007). The phylogenies generated for each locus were compared using each sequence alignment such that each protein topology was evaluated for all four *MAT* locus alignments.

We compared gene trees with the species tree (fig. 2) to estimate the minimal number of deep coalescence events needed to reconcile a given gene tree with the known species divergence. Using the method of minimizing deep coalescence (MDC) events (Maddison and Knowles 2006), the maximum parsimony solution for the number of deep coalescences is the minimum number of trans-specific lineages in

each branch of the species tree that are required to explain the gene tree. We estimated the number of DC for the *MAT* and non-*MAT* genes using Mesquite 2.7.4 (Maddison and Maddison 2003). Gene trees and MDC estimates were also obtained for N and C termini of the MAT proteins separately.

Tests for clock-like evolution were conducted by a likelihood ratio test of topologies in which a molecular clock was enforced to one without a clock using PAML 4.5. Ultrametric trees were then used to compute the genealogical depth of trees and the scaled ratio of terminal branch lengths to total genealogical depth (R_{SD}) according to (Uyenoyama 1997). Significant differences of R_{SD} values from the neutral coalescent were determined by generating a null distribution of R_{SD} values calculated on 1,000 simulated neutral coalescent trees generated using Mesquite v2.7.4.

Analysis of Divergence, Selection, and Recombination

Average codon-based evolutionary divergence over sequence pairs were estimated using MEGA 5.05 (Tamura et al. 2011). Analyses were conducted with the Nei-Gojobori model (Nei and Gojobori 1986) using proportion of differences (π_S), on the Gblocks filtered alignments. To test for amino acids under positive selection in each gene, various codon substitution models were fit to the data using the CODEML program of PAML 4.5. The site models M0, M1, M2, M7, and M8 as suggested by Yang et al. (2000) use a statistical distribution to describe the random variation in ω (=dN/dS) ratio among sites, where dN is the percentage of nonsynonymous divergence and dS is the percentage of synonymous divergence between pairs of sequences. M0 assumes constant ω ratio across all sites. Model M1 (neutral) assumes two site categories: sites with a ω ratio lower than 1 (purifying selection) and sites with ω ratio close to or equal to 1 (neutral selection). In addition to the two site categories of model M1, the model M2 assumes a third category with a ω ratio higher than 1 (positive selection). Model M7 assumes a beta distribution between 0 and 1 on the substitution type parameters p and q. In addition to these parameters, model M8 assumes an extra category of sites with a ω ratio higher than 1. Only models M2 and M8 allow for sites under positive selection. The following settings were used: Model = 0, NSites = 0 1 2 7 8, and CodonFreq = 2. For the *Ha-a2* gene the estimates of selection were conducted following removal of three putative pseudogene sequences.

Codon positions in the alignment with missing or ambiguous data were completely deleted in all sequences. The fit of two model pairs (M1 vs. M2 and M7 vs. M8) were tested using a likelihood ratio test (LRT). Twice the difference in log likelihood ratio between the models was compared with a chi-squared (χ^2) distribution with two degrees of freedom. Because PAML does not account for recombination and *MAT* genes were shown to undergo recombination, we estimated positive selection among the *MAT* genes using the population genetics method implemented in the software omegaMap (Wilson and McVean 2006). This Bayesian method jointly infers ω and population recombination (ρ), allowing each to vary along the sequence following independent change-point processes sampled through Markov chain Monte Carlo (MCMC). In our analysis, we used an exponential ratio distribution for μ (rate of synonymous transversions), and κ (transition–transversion ratio) and φ (insertion–deletion rate). The parameters ω and ρ were set to fit the exponential ratio distribution and the prior expected length of blocks was set to 30 codons. Two independent MCMC chains were run for 500,000 iterations with thinning set to 1,000 and a burn-in of 10%. Upon convergence, the two chains were merged to infer ω.

Intragenic recombination was analyzed using the methods RDP (Martin and Rybicki 2000), GENECONV (Padidam et al. 1999) and MAXCHI (Padidam et al. 1999) implemented in the software package RDP4 (Martin et al. 2010). Default settings were used except that disentangle overlapping signal was not applied. For RDP, a window size of 30 polymorphic sites within the triplet and no reference sequence were used, and for MAXCHI a window size of 70 was used. For GENECONV the default settings were used. In addition to the above methods Bootscan/Rescan (Martin et al. 2010) was used in secondary scans of the recombination signals detected by the primary scan.

Supplementary Material

Supplementary tables S1 and S2 and figures S1–S6 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the University of Michigan and the Swedish Foundation for Strategic Research. Kajsa Himmelstrand is acknowledged for help with the sequence analysis. The authors thank Stephen A. Smith for providing a Python script for calculating tree depth and terminal branch length. They also thank Georgiana May for many helpful suggestions that greatly improved the manuscript.

References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.

Asante-Owusu RN, Banham AH, Bohnert HU, Mellor EJC, Casselton LA. 1996. Heterodimerization between two classes of homeodomain proteins in the mushroom *Coprinus cinereus* brings together potential DNA-binding and activation domains. *Gene* 172:25–31.

Badrane H, May G. 1999. The divergence-homogenization duality in the evolution of the *b1* mating type gene of *Coprinus cinereus*. *Mol Biol Evol.* 16:975–986.

Banham AH, Asanteowusu RN, Gottgens B, Thompson SAJ, Kingsnorth CS, Mellor EJC, Casselton LA. 1995. An N-terminal dimerization domain permits homeodomain proteins to choose compatible partners and initiate sexual development in the mushroom *Coprinus cinereus*. *Plant Cell* 7:773–783.

Billiard S, Lopez-Villavicencio M, Devier B, Hood ME, Fairhead C, Giraud T. 2011. Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. *Biol Rev.* 86:421–442.

Brown AJ, Casselton LA. 2001. Mating in mushrooms: increasing the chances but prolonging the affair. *Trends Genet.* 17:393–400.

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

Charlesworth D, Vekemans X, Castric V, Glemin S. 2005. Plant self-incompatibility systems: a molecular evolutionary perspective. *New Phytol.* 168:61–69.
Cho SC, Huang ZY, Green DR, Smith DR, Zhang JZ. 2006. Evolution of the complementary sex-determination gene of honey bees: balancing selection and trans-species polymorphisms. *Genome Res.* 16: 1366–1375.
Clark NL, Aagaard JE, Swanson WJ. 2006. Evolution of reproductive proteins from animals and plants. *Reproduction* 131:11–22.
Dalman K, Olson Å, Stenlid J. 2010. Evolutionary history of the conifer root rot fungus *Heterobasidion annosum* sensu lato. *Mol Ecol.* 19: 4979–4993.
Devier B, Aguileta G, Hood ME, Giraud T. 2009. Ancient trans-specific polymorphism at pheromone receptor genes in basidiomycetes. *Genetics* 181:209–223.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
Engh IB, Skrede I, Saetre GP, Kauserud H. 2010. High variability in a mating type linked region in the dry rot fungus *Serpula lacrymans* caused by frequency-dependent selection? *BMC Genet.* 11:64.
Ferris P, Olson B, De Hoff PL, et al. (15 co-authors). 2010. Evolution of an expanded sex-determining locus in *Volvox*. *Science* 328: 351–354.
Figueroa F, Gunther E, Klein J. 1988. MHC polymorphism predating speciation. *Nature* 335:265–267.
Floudas D, Binder M, Riley R, et al. (71 co-authors). 2012. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* 336:1715–1719.
Gempe T, Hasselmann M, Schiott M, Hause G, Otte M, Beye M. 2009. Sex determination in honeybees: two separate mechanisms induce and maintain the female pathway. *PLoS Biol.* 7:e1000222.
Gonthier P, Garbelotto M. 2011. Amplified fragment length polymorphism and sequence analyses reveal massive gene introgression from the European fungal pathogen *Heterobasidion annosum* into its introduced congener *H. irregulare*. *Mol Ecol.* 20:2756–2770.
Halligan DL, Eyre-Walker A, Andolfatto P, Keightley PD. 2004. Patterns of evolutionary constraints in intronic and intergenic DNA of *Drosophila*. *Genome Res.* 14:273–279.
Hiscock SJ, Kües U. 1999. Cellular and molecular mechanisms of sexual incompatibility in plants and fungi. *Int Rev Cytol.* 193:165–295.
Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35:W585–W587.
Iwasa Y, Sasaki A. 1987. Evolution of the number of sexes. *Evolution* 41: 49–65.
James TY. 2007. Analysis of mating-type locus organization and synteny in mushroom fungi- beyond model species. In: Heitman J, Kronstad J, Taylor JW, Casselton LA, editors. Sex in fungi: molecular determination and evolutionary implications. Washington (DC): ASM Press. p. 317–331.
James TY. 2012. Ancient yet fast: rapid evolution of mating genes and mating systems in Fungi. In: Singh RS, Xu J, Kulathinal RJ, editors. Evolution in the fast lane: rapidly evolving genes and genetic systems. Oxford: Oxford University Press. p. 187–200.
James TY, Kues U, Rehner SA, Vilgalys R. 2004. Evolution of the gene encoding mitochondrial intermediate peptidase and its cosegregation with the *A* mating-type locus of mushroom fungi. *Fungal Genet Biol.* 41:381–390.
James TY, Lee M, van Diepen LTA. 2011. A single mating-type locus composed of homeodomain genes promotes nuclear migration and heterokaryosis in the white-rot fungus *Phanerochaete chrysosporium*. *Eukaryot Cell.* 10:249–261.
James TY, Porter D, Hamrick JL, Vilgalys R. 1999. Evidence for limited intercontinental gene flow in the cosmopolitan mushroom, *Schizophyllum commune*. *Evolution* 53:1665–1677.
James TY, Srivilai P, Kües U, Vilgalys R. 2006. Evolution of the bipolar mating system of the mushroom *Coprinellus disseminatus* from its tetrapolar ancestors involves loss of mating-type-specific pheromone receptor function. *Genetics* 172:1877–1891.
James TY, Stenlid J, Olson Å, Johannesson H. 2008. Evolutionary significance of imbalanced nuclear ratios within heterokaryons of the basidiomycete fungus *Heterobasidion parviporum*. *Evolution* 62: 2279–2296.
Johannesson H, Stenlid J. 2003. Molecular markers reveal genetic isolation and phylogeography of the S and F intersterility groups of the wood-decay fungus *Heterobasidion annosum*. *Mol Phylogen Evol.* 29: 94–101.
Kasuga T, White TJ, Taylor JW. 2002. Estimation of nucleotide substitution rates in eurotiomycete fungi. *Mol Biol Evol.* 19:2318–2324.
Korhonen K, Stenlid J. 1998. Biology of *Heterobasidion annosum*. In: Woodward S, Stenlid J, Karjalainen R, Hüttermann A, editors. *Heterobasidion annosum*: biology, ecology, impact and control. Wallingford (UK): CAB International. p. 43–70.
Kües U, Asante-Owusu RN, Mutasa ES, Tymon AM, Pardo EH, Oshea SF, Gottgens B, Casselton LA. 1994. Two classes of homeodomain proteins specify the multiple-*A* mating types of the mushroom *Coprinus cinereus*. *Plant Cell* 6:1467–1475.
Kües U, Casselton LA. 1992. Fungal mating type genes—regulators of sexual development. *Mycol Res.* 96:993–1006.
Kües U, Göttgens B, Stratmann R, Richardson WVJ, O'Shea SF, Casselton LA. 1994. A chimeric homeodomain protein causes self-compatibility and constitutive sexual development in the mushroom *Coprinus cinereus*. *EMBO J.* 13:4054–4059.
Kües U, James TY, Heitman J. 2011. Mating type in Basidiomycetes: unipolar, bipolar, and tetrapolar patterns of sexuality. In: Pöggeler S, Wöstemeyer J, editors. Evolution of fungi and fungal-like organisms, The Mycota XIV. Berlin: Springer Verlag. p. 97–160.
Kües U, Walser PJ, Klaus MJ, Aebi M. 2002. Influence of activated *A* and *B* mating-type pathways on developmental processes in the basidiomycete *Coprinus cinereus*. *Mol Genet Genomics.* 268:262–271.
Lupas A, Van Dyke M, Stock J. 1991. Predicting coiled coils from protein sequences. *Science* 252:1162–1164.
Maddison D, Maddison W. 2000. MacClade version 4: analysis of phylogeny and character evolution. Version 4.08. Sunderland (MA): Sinauer Associates.
Maddison WP, Knowles LL. 2006. Inferring phylogeny despite incomplete lineage sorting. *Syst Biol.* 55:21–30.
Maddison WP, Maddison DR. 2003. Mesquite: a modular system for evolutionary analysis. Version 1.0. [cited 2013 Aug 1]. Available from: http://mesquiteproject.org.
Martin D, Rybicki E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16:562–563.
Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26:2462–2463.
May G, Shaw F, Badrane H, Vekemans X. 1999. The signature of balancing selection: fungal mating compatibility gene evolution. *Proc Natl Acad Sci U S A.* 96:9172–9177.
McClure BA, Franklin-Tong V. 2006. Gametophytic self-incompatibility: understanding the cellular mechanisms involved in "self" pollen tube inhibition. *Planta* 224:233–245.
Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 3:418–426.
Nyholm SV, Passegue E, Ludington WB, Voskoboynik A, Mitchel K, Weissman IL, De Tomaso AW. 2006. fester, a candidate allorecognition receptor from a primitive chordate. *Immunity* 25:163–173.
Olson A, Aerts A, Asiegbu F, et al. (53 co-authors). 2012. Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. *New Phytol.* 194:1001–1013.
Otrosina WJ, Garbelotto M. 2010. *Heterobasidion occidentale* sp. nov. and *Heterobasidion irregulare* nom. nov.: a disposition of North American *Heterobasidion* biological species. *Fungal Biol.* 114:16–25.
Padidam M, Sawyer S, Fauquet CM. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* 265:218–225.
Pardo EH, Oshea SF, Casselton LA. 1996. Multiple versions of the *A* mating type locus of *Coprinus cinereus* are generated by three paralogous pairs of multiallelic homeobox genes. *Genetics* 144:87–94.

Penn DJ, Damjanovich K, Potts WK. 2002. MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc Natl Acad Sci U S A.* 99:11260–11264.

Petit E, Giraud T, de Vienne DM, et al. (14 co-authors). 2012. Linkage to the mating-type locus across the genus *Microbotryum*: insights into nonrecombining chromosomes. *Evolution* 66: 3519–3533.

Posada D. 2008. jModelTest: Phylogenetic model averaging. *Mol Biol Evol.* 25:1253–1256.

Richman AD, Kohn JR. 1999. Self-incompatibility alleles from *Physalis*: implications for historical inference from balanced genetic polymorphisms. *Proc Natl Acad Sci U S A.* 96:168–172.

Richman AD, Kohn JR. 2000. Evolutionary genetics of self-incompatibility in the Solanaceae. *Plant Mol Biol.* 42:169–179.

Richman AD, Uyenoyama MK, Kohn JR. 1996. Allelic diversity and gene genealogy at the self-incompatibility locus in the Solanaceae. *Science* 273:1212–1216.

Rozen S, Skaletsky HJ. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics methods and protocols: methods in molecular biology. Totowa (NJ): Humana Press. p. 365–386.

Schlesinger R, Kahmann R, Kamper J. 1997. The homeodomains of the heterodimeric bE and bW proteins of *Ustilago maydis* are both critical for function. *Mol Gen Genet.* 254:514–519.

Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246–1247.

Sigrist CJ, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N. 2010. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.* 38: D161–D166.

Spit A, Hyland RH, Mellor EJC, Casselton LA. 1998. A role for heterodimerization in nuclear localization of a homeodomain protein. *Proc Natl Acad Sci U S A.* 95:6228–6233.

Stankis MM, Specht CA, Yang H, Giasson L, Ullrich RC, Novotny CP. 1992. The *Aα* mating locus of *Schizophyllum commune* encodes two dissimilar multiallelic homeodomain proteins. *Proc Natl Acad Sci U S A.* 89:7169–7173.

Stenlid J, Karlsson JO. 1991. Partial intersterility in *Heterobasidion annosum*. *Mycol Res.* 95:1153–1159.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. *Nat Rev Genet.* 3:137–144.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2685–2686.

Tymon AM, Kues U, Richardson WVJ, Casselton LA. 1992. A fungal mating type protein that regulates sexual and asexual development contains a POU-related domain. *EMBO J.* 11:1805–1813.

Uyenoyama MK. 1997. Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. *Genetics* 147:1389–1400.

Uyenoyama MK. 2005. Evolution under tight linkage to mating type. *New Phytol.* 165:63–70.

Uyenoyama MK, Zhang Y, Newbigin E. 2001. On the origin of self-incompatibility haplotypes: transition through self-compatible intermediates. *Genetics* 157:1805–1817.

van Peer AF, Park SY, Shin PG, Jang KY, Yoo YB, Park YJ, Lee BM, Sung GH, James TY, Kong WS. 2011. Comparative genomics of the mating-type loci of the mushroom *Flammulina velutipes* reveals widespread synteny and recent inversions. *PLoS One* 6: e22249.

Vilgalys R, Sun BL. 1994. Ancient and recent patterns of geographic speciation in the oyster mushroom *Pleurotus* revealed by phylogenetic analysis of ribosomal DNA sequences. *Proc Natl Acad Sci U S A.* 91:4599–4603.

Waters PD, Wallis MC, Graves JAM. 2007. Mammalian sex—origin and evolution of the Y chromosome and SRY. *Semin Cell Dev Biol.* 18: 389–400.

Wilson DJ, McVean G. 2006. Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics* 172: 1411–1425.

Wiuf C, Zhao K, Innan H, Nordborg M. 2004. The probability and chromosomal extent of trans-specific polymorphism. *Genetics* 168: 2363–2372.

Wright S. 1939. The distribution of self-steriliity alleles in populations. *Genetics* 24:538–552.

Wu J, Saupe SJ, Glass NL. 1998. Evidence for balancing selection operating at the *het-c* heterokaryon incompatibility locus in a group of filamentous fungi. *Proc Natl Acad Sci U S A.* 95:12398–12403.

Wu J, Ullrich RC, Novotny CP. 1996. Regions in the Z5 mating gene of *Schizophyllum commune* involved in Y-Z binding and recognition. *Mol Gen Genet.* 252:739–745.

Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.

Yang ZH, Nielsen R, Goldman N, Pedersen AMK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449.

Yee AR, Kronstad JW. 1998. Dual sets of chimeric alleles identify specificity sequences for the bE and bW mating and pathogenicity genes of *Ustilago maydis*. *Mol Cell Biol.* 18:221–232.

Yu C-S, Wang J-Y, Yang J-M, Lyu P-C, Lin C-J, Hwang J-K. 2003. Fine-grained protein fold assignment by support vector machines using generalized *n*peptide coding schemes and jury voting from multiple-parameter sets. *Proteins* 50:531–536.

Yue CL, Osier M, Novotny CP, Ullrich RC. 1997. The specificity determinant of the Y mating-type proteins of *Schizophyllum commune* is also essential for Y-Z protein binding. *Genetics* 145: 253–260.