Title

Genomic determination of breeding systems and trans-specific evolution of HD MAT genes in suilloid fungi.

Journal

Genetics, 224(2)

Authors

Ke, Yi-Hong

Branco, Sara

Bazzicalupo, Anna

et al.

Publication Date

2023-05-26

DOI

10.1093/genetics/iyad069

Peer reviewed

GENETICS, 2023, 224(2), iyad069
https://doi.org/10.1093/genetics/iyad069
Advance Access Publication Date: 18 April 2023
Investigation

Genomic determination of breeding systems and trans-specific evolution of *HD MAT* genes in suilloid fungi

Yi-Hong Ke [1,*], Sara Branco [2], Anna L. Bazzicalupo [3], Nhu H. Nguyen [4], Hui-Ling Liao [5,6], Peter Kennedy [7], Thomas D. Bruns [8], Alan Kuo [9], Kurt LaButti [9], Kerrie Barry [9], Igor Grigoriev [8,9], Rytas Vilgalys [1]

[1]Biology Department, Duke University, 130 Science Dr, Durham, NC 27708, USA
[2]Integrative Biology, University of Colorado, 1151 Arapahoe St, SI 2071, Denver, CO 80204, USA
[3]Department of Zoology, University of British Columbia, 4200—6270 University Blvd, Vancouver, British Columbia, Canada, V6T 1Z4
[4]Department of Tropical Plant and Soil Sciences, University of Hawaii at Manoa, 3190 Maile Way, Honolulu, HI 96822, USA
[5]North Florida Research and Education Center, University of Florida, 155 Research Rd, Quincy, FL 32351, USA
[6]Soil and Water Sciences Department, University of Florida, 1692 McCarty Dr, Room 2181, Building A, Gainesville, FL 32611, USA
[7]Department of Plant and Microbial Biology, University of Minnesota, 1475 Gortner Ave, Saint Paul, MN 55108, USA
[8]Department of Plant and Microbial Biology, University of California at Berkeley, 111 Koshland Hall, Berkeley, CA 94720, USA
[9]Lawrence Berkeley National Laboratory, US Department of Energy Joint Genome Institute, 1 Cyclotron Road, Berkeley, CA 94720, USA

*Corresponding author: Biology Department, Duke University, 130 Science Dr, Durham, North Carolina 27705, USA. Email: yk158@duke.edu

Abstract

Studying the signatures of evolution can help to understand genetic processes. Here, we demonstrate how the existence of balancing selection can be used to identify the breeding systems of fungi from genomic data. The breeding systems of fungi are controlled by self-incompatibility loci that determine mating types between potential mating partners, resulting in strong balancing selection at the loci. Within the fungal phylum Basidiomycota, two such self-incompatibility loci, namely *HD MAT* locus and *P/R MAT* locus, control mating types of gametes. Loss of function at one or both *MAT* loci results in different breeding systems and relaxes the *MAT* locus from balancing selection. By investigating the signatures of balancing selection at *MAT* loci, one can infer a species' breeding system without culture-based studies. Nevertheless, the extreme sequence divergence among *MAT* alleles imposes challenges for retrieving full variants from both alleles when using the conventional read-mapping method. Therefore, we employed a combination of read-mapping and local *de novo* assembly to construct haplotypes of *HD MAT* alleles from genomes in suilloid fungi (genera *Suillus* and *Rhizopogon*). Genealogy and pairwise divergence of *HD MAT* alleles showed that the origins of mating types predate the split between these two closely related genera. High sequence divergence, trans-specific polymorphism, and the deeply diverging genealogy confirm the long-term functionality and multiallelic status of *HD MAT* locus in suilloid fungi. This work highlights a genomics approach to studying breeding systems regardless of the culturability of organisms based on the interplay between evolution and genetics.

Keywords: fungi, mating, breeding system, balancing selection, *Suillus*, *Rhizopogon*, self-incompatibility, evolution, homeodomain protein

Introduction

Self-incompatibility loci in plants and fungi directly influence mating by imposing constraints on selfing and promoting outcrossing. The alleles at such loci are grouped into two or more self-incompatible allele types, which determine the mating types and the mating compatibility between gametes (Kahmann and Bolker 1996). Two such loci are present in species of Basidiomycota, namely *HD MAT* locus and *P/R MAT* locus (James *et al.* 2013; Coelho *et al.* 2017). However, not all *HD MAT* loci and *P/R MAT* loci control mating compatibility in all species. In some species, alleles at one or both *MAT* loci have lost their function in controlling mating types. The breeding systems of basidiomycetes are classified according to the number of loci controlling mating types, of which tetrapolar breeding system is controlled by two *MAT* loci, bipolar breeding systems is controlled by two loci, and homothallic (self-fertile) breeding system has no self-incompatibility locus (Whitehouse 1949; Koltin *et al.* 1972; Kües *et al.* 2011). In the classic basidiomycete life cycle, meiotic basidiospores germinate to produce a mycelium comprised of identical haploid nuclei (the monokaryon) which can mate with another monokaryon to yield a dikaryotic mycelium whose hyphal cells each possesses two paired haploid nuclei with compatible alleles at *MAT* loci (Kües *et al.* 2011). Even though the paired nuclei are separate, from a genome sequencing standpoint dikaryons are essentially the same as diploid.

In most basidiomycete fungi, determination of the breeding systems requires extensive crosses among gametes (monokaryotic colonies from basidiospores) of the same parent as well as between different dikaryotic individuals. Using this approach, Kniep (1928), Buller (1941), and Raper (1953) demonstrated that most fungi in Basidiomycota exhibit multi-allelic breeding systems controlled by one or two loci. Though traditional crossing studies can be applied to species which are easy to grow in pure culture, breeding systems for most unculturable fungi in Basidiomycota are still unknown. Study of breeding systems of symbiotic

Received: February 24, 2023. **Accepted:** March 23, 2023

ectomycorrhizal (ECM) fungi in particular poses difficulties since basidiospores of most ECM fungal species fail to germinate or else grow poorly in the laboratory condition (Fries 1978; 1987; Nara 2009; Cairney and Chambers 2013).

As an alternative to culture-based studies of mating compatibility, studying the signatures of evolution in sequence data can be used to reveal the structure and function of fungal breeding systems. When alleles at a *MAT* locus determine mating types, the constant heteroallelic status and/or the advantage of the rare alleles result in strong balancing selection (May *et al.* 1999; James 2015), forming distinct signatures in sequences. The increased coalescent time of alleles under balancing selection is predicted to cause an excess in nucleotide diversity and trans-specific polymorphism at such *MAT* loci (May *et al.* 1999). In the context of self-incompatibility loci, trans-specific polymorphism manifests in alleles not clustered by species but by allele types (Wang and Mitchell-Olds 2017). These signatures of balancing selection at functional *MAT* loci can be used to infer breeding systems without the need for culturing and crossing. Sequencing alleles at *MAT* locus in Basidiomycota, however, also poses difficulties, since extreme sequence divergence among different *MAT* alleles and the large size of the loci (typically over 10 kb) often prevent the use of universal PCR primers to study the sequences of *MAT* loci even at the population level (Badrane and May 1999; Coelho *et al.* 2010; van Diepen *et al.* 2013).

We explored an approach to infer breeding systems based on signatures of balancing selection from genome shotgun sequences, which requires no primer designs. Abundant shotgun genome sequence data is now available for many fungi whose breeding systems are unknown. However, the extremely high sequence divergence of the *HD MAT* locus complicates the detection of variants (SNPs) using Illumina short reads (Supplementary Fig. 1), since reads from different alleles are too divergent to map to the mating type allele in the reference genome by regular bioinformatic tools, such as BWA (Li and Durbin 2010) and bowtie2 (Langmead and Salzberg 2012). To accurately infer the sequence evolution of the highly divergent *HD MAT* alleles, we applied an approach that employs local *de novo* assembly of target *HD MAT* locus using automated Target Restricted Assembly Method (aTRAM) (Allen *et al.* 2015), which utilizes iterative read mapping and *de novo* assembly to correctly assemble alleles on *HD MAT* locus (Supplementary Fig. 2). This genome-based approach can be applied regardless of the cultivability and lack of PCR primer specificity for different *HD MAT* alleles. We used this approach to reassemble *HD MAT* alleles for a group of ectomycorrhizal fungi in the genera *Suillus* and *Rhizopogon* in order to infer their breeding systems and elucidate the evolution of their *HD MAT* alleles.

Materials and methods

Genomic sequencing and annotation

Genomes of eight fully assembled suilloid fungal species (*Suillus* spp. and *Rhizopogon* spp.) were examined in this study, including 7 published genomes and 1 new genome (Table 1). All eight genomes were sequenced using short-read Illumina sequencing. The new *Rhizopogon salebrosus* TDB-379 genome was sequenced with Illumina HiSeq using Illumina 270-bp fragment library sequenced in 2× 150-bp format. The raw Illumina reads were QC filtered to remove artifact/process contamination and then separated into mitochondrial and nuclear genome fraction. A subset of the nuclear genome fraction was subsequently assembled using Velvet (Zerbino and Birney 2008). The resulting assembly was used to create in silico long mate-pair library with insert 2,000 +/- 90 bp which was then assembled together with the target fastq using AllPathsLG release version R47710 (Gnerre *et al.* 2011). Another long mate-pair library with insert 3,000 +/- 90 bp was created from this initial AllPathsLG assembly. This second long mate-pair library was reassembled with the first mate-pair library along with a subset of the nuclear genome fraction to create a final AllPathsLG assembly. Contigs with BLAST+ hits to refseq.mitochondrion or have a minimum of 25 pairs and link to a contig with hits to refseq.mitochondrion were kept for mitochondrial assembly while those with hits to nt with 18S annotations were removed.

HD MAT loci of the eight suilloid species were initially identified by blastp hits against *HD* genes of *Rhizopogon* species (Mujic *et al.* 2017). The identity flanking genes surrounding the candidate *HD MAT* loci were identified by their best blastn hit. The genome coordinates of genes linked to *HD MAT* loci were listed in Table 2. Genomic alignment around *HD MAT* loci was conducted by progressive Mauve (Darling *et al.* 2004) assuming no rearrangements and visualized by Geneious R11 (https://www.geneious.com).

Genome wide variants detection

We performed sequence variant detection across all eight genomes based on mapping and variant-calling. Raw reads were trimmed by cutadapt (Martin 2011) to minimum quality 25, minimum length 80 bp, and then mapped to reference assemblies using bowtie2 with default parameters in paired-end mode. Genomic mapping depth was calculated from averaging read depth across all sites excluding 10% extreme values to avoid bias from poorly assembled regions. Raw variants were called by bcftools (Li 2011) assuming diploid ploidy. Because the length of insertions during library preparation varies in each sample, including libraries with both mate-pair and paired-end reads, anomalous read pairs (-A) and overlapped reads were allowed (-x). Total depth and minor allele depth at each variant site were as reported in DP field and the second large number in AD field, respectively. Minor allele frequency was estimated by dividing the minor allele depth by total depth. The numbers of SNPs and SNP density in each genome were estimated with the numbers of SNPs not in 10% extreme depth of each tail (Supplementary Fig. 3, Supplementary Fig. 4).

Variant detection at *HD MAT* loci

We performed the conventional variant detection approach based on mapping and variant-calling to *HD* genes in the reference genomes, including the intergenic sequence, and the 1,000-bp flanking sequence on either side (referred hence as *mapping variant detection approach*). For each genome, reads were mapped to the reference *HD MAT* sequences and the variants were called with the same parameters as in genomic statistics. The alternative *HD MAT* sequences were obtained by applying the discovered variants to the *HD MAT* sequences in the reference genomes. The reference *HD MAT* sequences and the alternative *HD MAT* sequences were treated as two alleles generated by mapping variant detection approach.

For local *de novo* assembly approach, the reference *HD MAT* sequences were treated as "bait sequences" for aTRAM using Velvet assembler (Zerbino and Birney 2008). We expanded the original aTRAM pipeline by running five independent runs with different Kmer sizes (Kmer = 30, 35, 40, 45, and 50) to produce hypothetical *HD MAT* sequences. The expected depth for Velvet was set to half of the average genome depth as that is the predicted depth of each haplotype. We then mapped reads to all hypothetical *HD MAT* sequences generated by aTRAM. The reads mapped to hypothetical

Table 1. Genome statistics of suilloid genomes used in this study.

Species (JGI portal id)	NCBI accession	Genome assembly size (Mbp)	Genome sequencing depth	Raw SNP density (per 1 kb)	Filtered SNP density (per 1 kb)[a]	Suggested Ploidy	Source	Reference
Rhizopogon salebrosus (Rhisa1)	PRJNA333299	82.29	323.3	4.07	3.898	dikaryotic	mycorrhizae culture	this article
Rhizopogon vinicolor (Rhivi1)	PRJNA196085	36.1	384.7	4.01	3.268	dikaryotic	fruiting body culture	Mujic *et al.* 2017
Suillus americanus (Suiame1)	PRJNA333476	50.81	164.2	4.86	4.551	dikaryotic	fruiting body culture	Lofgren *et al.* 2021
Suillus brevipes (Suibr2)	PRJNA218802	52.03	228.9	4.85	4.664	dikaryotic	fruiting body culture	Branco *et al.* 2015
Suillus decipiens (Suidec1)	PRJNA333477	62.78	290.3	0.85	0.0690	monokaryotic	mycorrhizae culture	Lofgren *et al.* 2021
Suillus weaverae (Suigr1)	PRJNA372929	42.34	188.1	5.85	5.606	dikaryotic	fruiting body culture	Lofgren *et al.* 2021
Suillus hirtellus (Suihi1)	PRJNA333479	49.94	219.5	0.35	0.007	monokaryotic	fruiting body culture	Lofgren *et al.* 2021
Suillus luteus (Suilu4)	PRJNA242126	44.49	169.7	0.04	0.004	mono-karyotic	single spore culture	Kohler *et al.* 2015

[a] filtering 10% sites with extreme depth in each tail.

HD MAT sequences were expected to contain all the reads from both *HD MAT* alleles. We collected those reads and conducted a final *de novo* assembling using MIRA (Chevreux *et al.* 1999) in accurate mode to produce stringent assemblies for both alleles. Newly assembled sequences were aligned to the original *HD MAT* reference sequences using MAFFT 7 (Katoh and Standley 2013) to verify their homology. The two newly assembled sequences that were homologous to *HD MAT* loci were used as alleles generated from the local *de novo* assembly approach.

Quality assessment of mapping variant detection and local *de novo* assembly method

Assuming the two sequences of the two alleles represented the sequences in the genomes, we expected when using both alleles as reference sequences for mapping, the depth drop of *HD MAT* loci as seen in genome no longer exist since the two alleles represent exactly the same sequences as which the reads were sequenced from. To compare whether the alleles meet this expectation, we mapped reads to the two alleles produced by mapping variant detection approach and to the two alleles produce by local *de novo* assembly method. For each mapping run, either two alleles of local *de novo* assembly or two alleles of mapping variant detection approach were used as reference sequences at the same time, so the reads never count twice in each method. The average depth and the depth distribution as calculated for genome wide variants detection were plotted from the mapping results and were compared to the genome-wide distribution by R language (R Core Team 2022) with package *ggplot2* (Wickham 2016).

Knowing that trans-specific polymorphisms were likely to exist and the high dissimilarity between two alleles in the mapping process, we speculated the reads of *HD MAT* locus would carry enough information to generate haplotypes of the entire *HD MAT* locus. To test whether the two local *de novo* assembled alleles represent two unambiguous haplotypes, we used HAPCUT2 (Edge *et al.* 2017) to investigate whether the phases of the two *de novo* assembled alleles could be confidently assigned. HAPCUT2 iteratively proposed new haplotypes to minimize mismatches and splits of the reads. Normally, HAPCUT2 infers haplotypes by mapping reads from two alleles against a single reference sequence. In our case, since the reads were known unable to map to the other allele properly, we first mapped the reads to two alleles separately as described above, and then merged the two alleles by adjusting the read position and CIGAR code according to the MAFFT 7 alignment of the two alleles (align_sam.py). The merged mapping results and the variants defined by *bcftools* were used as input for HAPCUT2. The haplotype blocks inferred by HAPCUT2 were then aligned to the two local *de novo* assembled alleles by MAFFT 7 to confirm their consistency.

As both *HD* genes in each allele were expected to produce functional proteins, we also used the translated amino acid sequences to assess the correctness of the haplotypes. Premature ORF or frame shifting indels suggests an incorrectly inferred haplotype. Coding sequences (CDSs) of the assembled local *de novo* assembled haplotypes were defined by transferring the annotated CDSs in the reference genomes to their homologous region in their MAFFT 7 alignments. The translated amino acid sequences were aligned to known *HD* proteins (Mujic *et al.* 2017) to assess the existence of premature ORF or frameshifting. The translated amino acid sequences were deposited in Genbank ON315855-ON315867.

Genetic diversity and genealogy

In order to test whether *HD MAT* locus fell into the high genetic diversity tail in the genome, we calculated the variant density of the filtered SNPs in 1,000-bp non-overlapping windows across each genome. The variant density between the two locally *de novo* assembled alleles was calculated as the numbers of mismatches (excluding gaps) in their pairwise alignments produced by MAFFT 7.

To visualize detailed genetic divergence between each pair of *HD MAT* alleles, the sequence divergence in codon-based alignments of CDSs using prank (Loytynoja 2014) was calculated for all combinations of *HD MAT* genes. Mean pair-wise divergence was measured by the average of JC-corrected divergence (Jukes and Cantor 1969) in sliding 300-bp overlapping windows. Genealogy of *HD MAT* alleles was inferred using PhyML 3 (Guindon *et al.* 2010) with HKA substitution model with 100

Table 2. Summary of *HD MAT* depth and SNP density.

Species	*HD MAT* coordinate	*HD MAT* depth			*HD MAT* variant density (per 1 kb)	
		reference[a]	mapping variant detection[b]	local *de novo*[c]	mapping variant detection[b]	local *de novo*[c]
R. salebrosus	scaffold_4:325574-327895	291	155 + 149 = 304	172 + 177 = 349	37.2	127.7
R. vinicolor[1]	scaffold_24:106196-108049	331	171 + 168 = 339	190 + 192 = 382	40.0	121.9
S. americanus	scaffold_2:750316-752187	118	66 + 57 = 122	84 + 86 = 150	14.1	165.5
S. brevipes	scaffold_7:289073-290784	152	82 + 78 = 160	128 + 125 = 253	11.9	224.5
S. decipien	scaffold_5:375610-377851	273	NA	NA	NA	NA
S. granulatus	scaffold_1:1580479-1582647	183	95 + 92 = 187	110 + 109 = 219	29.4	104.3
S. hirtellus	scaffold_1:1479724-1482059	210	NA	NA	NA	NA
S. luteus	scaffold_1:1133054-1134939	164	NA	NA	NA	NA

[a] When reads mapped to single reference sequence.
[b] When reads mapped to two alleles from mapping variant detection approach. Each read only mapped the best matched allele (no reads are counted twice).
[c] When reads mapped to two alleles from local *de novo* assembly approach. Each read only mapped the best matched allele (no reads are counted twice).

bootstraps from 400 bp equally distributed blocks in CDSs of *HD1* and *HD2* genes. Rooting of *Suillus* HD genes was performed based on the minimal clades that contained all *Rhizopogon* alleles.

Results

We investigated the breeding systems of eight species of *Suillus* and *Rhizopogon*. We used seven previously sequenced genomes that were fully assembled and annotated from Illumina short reads, with average sequencing depths from 170× to 385× (Table 1). We sequenced the *Rhizopogon salebrosus* TDB-379 genome with Illumina to 120.8× read depth and assembled in 1,976 scaffolds totaling 82.29 Mbp (N50 = 40 and L50 = 106,002 bp) and 5,710 contigs totaling 42.82 Mbp (N50 = 559 and L50 = 17,961 bp). We also sequenced the *R. salebrosus* TDB-379 transcriptome with Illumina and assembled it into 83,186 contigs, which aided the prediction of 18,900 protein coding genes.

Ploidy of genomes

Ploidy of all 8 genomes was examined by the patterns of SNP sites. The observed SNP density within the sequenced genomes of three species *S. luteus*, *S. decipiens*, and *S. hirtellus* was two orders of magnitude lower than in the other five species (Table 1). Also, when depicting mapping depths of putative heterozygous sites (SNPs), whose depth should be close to genome average (Supplementary Fig. 5), the majority of the putative heterozygous sites in the three genomes had depths close to twice genome average. This suggests the majority of the putative heterozygous sites from those genomes were likely the result of mapping two paralogs to one gene instead of true signals of heterozygotes. Based on the combination of lower counts of detected heterozygous sites and abnormal depths at those heterozygous sites, we conclude that *S. luteus*, *S. decipiens*, and *S. hirtellus* genomes were monokaryotic (haploid) since the signal of heterozygotes in the genome was not distinguishable from noise or else absent.

For the five genomes determined to be dikaryotic (diploid), we observed a consistent decrease in mapping depth at *HD MAT* locus when reads were mapped against the *HD MAT* sequences in reference genomes (Supplementary Fig. 1). This suggests substantial numbers of reads at *HD MAT* locus failed to map to the reference sequences. An additional decrease of mapping depth around the *HD MAT* locus of *S. weaverae* was also observed, which was likely due to structural variation of partial duplicated genes and the subsequent poly-N uncertain sequences in that region (see discussion).

De novo assembly of *HD MAT* locus

We locally *de novo* assembled two *HD MAT* alleles from each dikaryotic genome using aTRAM (Allen *et al.* 2015) (Supplementary Fig. 2). For genomes identified as monokaryons in the previous section, the *HD MAT* sequences in the reference genomes were treated as the only allele. The total length of the *de novo* assembled alleles varied from 4,485 to 4,612 bp, which included *HD1*, *HD2*, and the intergenic regions. Mapping reads to the *HD MAT* sequences in the respective reference genomes revealed a reduction in read depth and distorted depth distribution (Fig. 1; Supplementary Fig. 6). Less than 70% of reads in *Suillus* and less than 90% reads in *Rhizopogon* at *HD MAT* locus could be mapped to their reference genome (Table 2; Fig. 1). This deficiency of mapping depth disappeared when mapping was performed to two locally *de novo* reassembled alleles, to which more than 98% of reads could be mapped. Overall, the mapping depth profile for local *de novo* assembled alleles showed smoother mapping depths and the average mapping depths were closer to the genome average (Supplementary Fig. 6; Supplementary Fig. 7). We interpret this as the evidence that *de novo* assembly using aTRAM fully captured all reads belonging to both alleles of *HD MAT* locus while the sequences in the reference genomes or the sequences generated from mapping variant detection missed a substantial amount of reads from the alternative alleles. We further verified whether *de novo* assembled alleles represent valid haplotypes using HAPCUT2. This read-based haplotype inference generates complete haplotype blocks across *HD MAT* locus in each diploid genome. There was no conflict between the *de novo* assembled alleles and the HAPCUT2 phased alleles when they were aligned to each other (Table 3).

All local *de novo* assembled haplotypes possessed open reading frames and no premature stop codons in the two homeodomain genes. The length of the amino acid sequences varied from 642 to 665 aa in *HD1*, and 575 to 561 aa in *HD2*. Sequences of all *de novo* assembled *HD* genes were also readily alignable to each other (Supplementary Fig. 8). We conclude that the aTRAM assembled alleles are the optimal haplotypes of the sequenced reads and are likely to be the actual haplotype in the chromosomes.

Genetic diversity and genealogy of *HD MAT* haplotypes in suilloid genomes

The number of variants detected by the *de novo* assembled *HD MAT* alleles was 3–18-fold higher than the mapping variant detection approach (Table 2) which failed to detect variants at the center of *HD MAT* locus (Fig. 2). All the variants detected by the mapping variant detection approach were included in the *de*

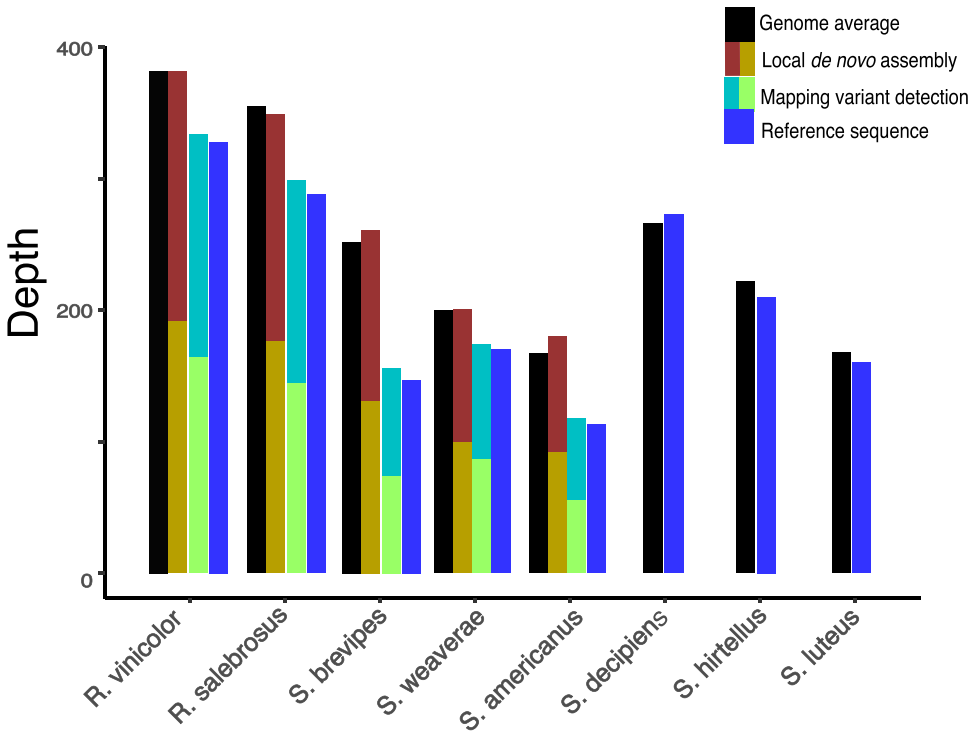

Fig. 1. Average read depths mapped to *HD MAT* sequences generated from different approaches. The average depths of reads mapping to *HD MAT* sequences generated from different approaches. More reads are mapped to local *de novo* assembly alleles than to the other sequences, whose read depths are also closer to genome averages. This suggests some reads of alternative alleles failed to map to the single reference sequence, so those alternative alleles are undetectable using mapping variant detection. Genome average: Read depths across genome; Local *de novo* assembly: Read depths mapped to two alleles generated from local *de novo* assembly. Depths to each haplotype shown in different colors. Mapping variant detection: Read depths mapped to two alleles generated from mapping variant detection. Depths to each haplotype shown in different colors. Reference: Read depths mapped to the single *HD MAT* sequence in the reference genome. Only mapping to reference sequence is applicable to homokaryontic (haploid) genomes since sequences of alternative alleles do not exist.

Table 3. Summary of HAPCUT2 haplotype blocks at *HD MAT* locus.

species	Haplotype block size	number of reads mapped	number of variants	number of conflicts[a]
S. brevipes	4713	4516	848	0
S. americanus	4870	2768	664	0
S. weaverae	5405	3161	434	0
R. vinicolor	4419	5521	433	0
R. salebrosus	4260	4974	493	0

[a] mismatch between alignment of HAPCUT2 haplotype block and local *de novo* assembled alleles.

novo assembled *HD MAT* alleles except for two variants in *S. weaverae* (Fig. 2). Those two variants were the artifact of mismapping reads from the paralogous partial *HD2* sequence only found in *S. weaverae* genome. The alignments of two locally *de novo* assembled *HD MAT* haplotypes and the sequences in the reference genomes showed the sequences in the reference genome contained segments of variants from different *HD MAT* haplotypes (Fig. 2). This suggests the *HD MAT* assemblies in the reference genomes are chimeras of two *HD MAT* haplotypes from the two alleles.

The SNP density of *HD* genes and their intergenic regions fell into the 0.1% highest regions in the sequenced diploid genomes (Supplementary Fig. 9). The sequence divergence was highest at the N-termini within genus (non-green lines in Fig. 3), suggesting the N-termini were under the strongest balancing selection. The sequence divergence between some pairs of *HD* haplotypes from *Rhizopogon* and *Suillus* species (non-green lines in Fig. 3) was less than other intra-genus comparisons at the N-termini (green lines in Fig. 3), suggesting that the origin of those mating-type alleles predates the split between these two genera. The sequence divergence of C-termini of *HD* gene haplotypes was higher between genera than within-genera, indicating they were closer to the pattern of neutral evolution and they had a more recent divergence time than the N-termini did. The polyphyletic genealogy of alleles of *Rhizopogon* at the N-termini (Fig. 4) also suggests the maintained polymorphism since the common ancestor of *Suillus* and *Rhizopogon*. Furthermore, differences in well-supported branches in the genealogy of *HD1* and *HD2* suggest recombination had occurred within the *HD MAT* locus between *HD1* and *HD2* genes. Surprisingly, although the divergence time of N-termini was greater than that of C-termini, the sequence divergence of N-termini between *Rhizopogon* and *Suillus* (green lines in Fig. 3) was less than that of the C-termini, which implies the sequences at N-termini were more conserved in the long term.

The inference of breeding system

Two factors determine the pattern of sequence divergence: the divergence time and the mutation rate. Both higher mutation rates and longer coalescent times lead to higher sequence divergence of the sampled alleles. However, a higher mutation rate does not distort genealogy, and cannot lead to more frequent trans-specific polymorphisms than loci with lower mutation rates. Balancing selection increases the absolute coalescent time of the sampled

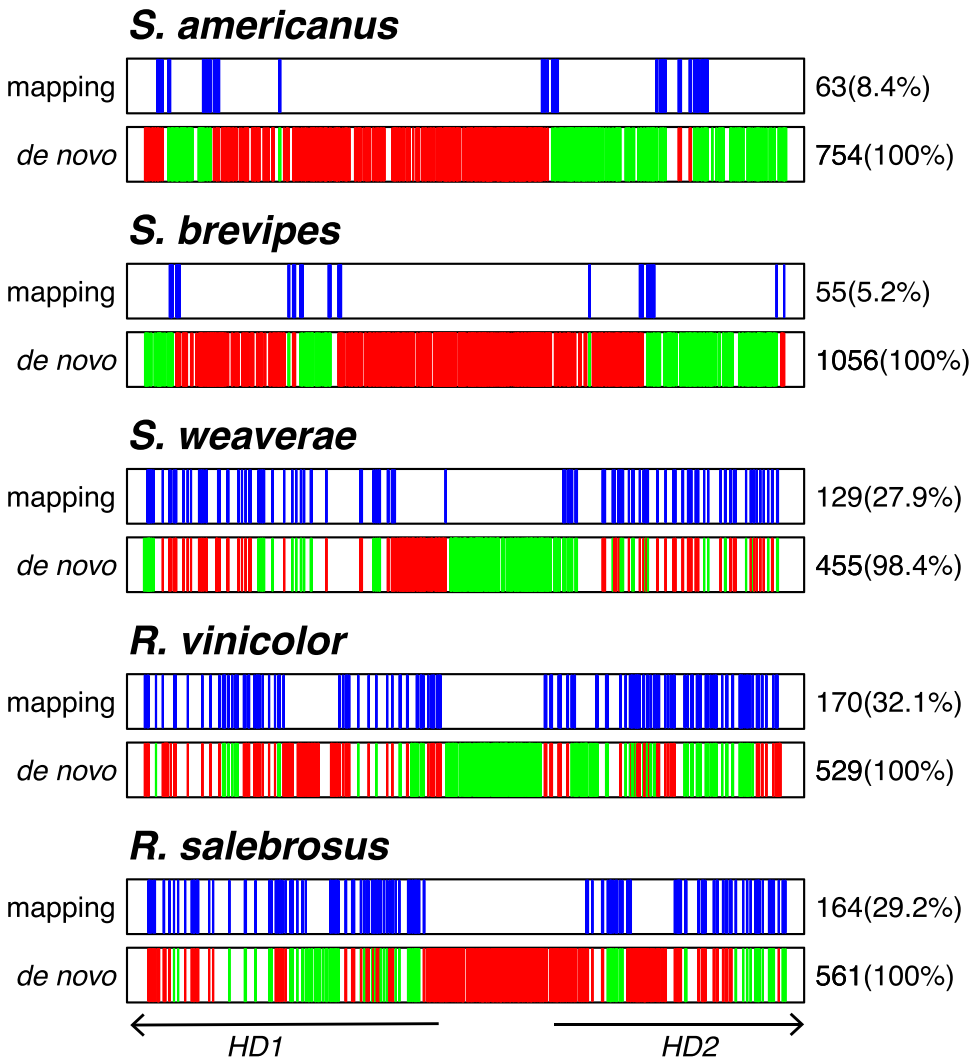

Fig. 2. Variants distribution across *HD MAT* locus in suilloid genomes. Colored bars represent the positions of detected variants across sites at *HD MAT* locus by mapping variant detection approach (blue) or local *de novo* assembly approach (red and green). Red or green bars indicate the sequences in the reference genomes are the variants from one haplotype or the other. The numbers beside each row indicate the total number of variants detected by the approach. The percentage in parentheses shows how many variants in the other approach are included in this approach. The pattern suggests the sequences in the reference genomes are chimeras from two haplotypes. The lack of blue bars at the center of *HD MAT* locus suggests that mapping variant detection approach fails to capture variants at the center.

alleles, resulting in both higher sequence divergence and more trans-specific polymorphisms when the expected coalescent time predates the split of the taxa (May *et al.* 1999; Wang and Mitchell-Olds 2017). Mating-type determining genes of Basidiomycota should be under balancing selection since their constant heterozygous state and the rare alleles have an advantage in a system where only different alleles are mating compatible. All of our results are consistent with this expectation. First, a high level of variants was observed at the *HD MAT* locus. Second, both genealogy and the patterns of sequence divergence suggest that the time since divergence of the center region of *HD MAT* alleles predates their speciation and even predates the split between these two closely related genera. Thus, we conclude the *HD MAT* locus is under strong balancing selection, as predicted for a functional mating locus determining mating specificity (May *et al.* 1999; James 2015; Wang and Mitchell-Olds 2017).

The numbers and the turnover of alleles at a functional *MAT* locus may also be determined by the genealogy of sampled alleles. A functional *MAT* locus may have two or more alleles, i.e. either biallelic or multiallelic (Nieuwenhuis *et al.* 2013; Kues 2015). In a persisting biallelic system without the formation of new mating types, only two fixed allele types exist, presumably due to losing the ability to produce new mating types or failure of the new mating types to cross the local optimum of two alleles status (Nieuwenhuis *et al.* 2013; Krumbeck *et al.* 2020). In this case, two clusters of alleles diverged from an early point are expected, with alleles in either group diverging in concordance with speciation events (Fig. 5a). Additionally, all dikaryotic (diploid) individuals must possess the same two alleles from each cluster. In a multiallelic system, new mating types must consistently form to counter allele loss through drift (Kues and Casselton 1993; Kahmann and Bolker 1996). Even if at some point there are only two alleles co-existing in certain species or sampled individuals, the genealogical structure can still be distinguished from persisting biallelic by not having the same pairs of alleles in the closely related species (e.g. species B in Fig. 5b). Since the *HD MAT* alleles sampled in *Suillus* species did not form a deep bisecting split and no pairs of sampled species share the same pair of *HD MAT* alleles,

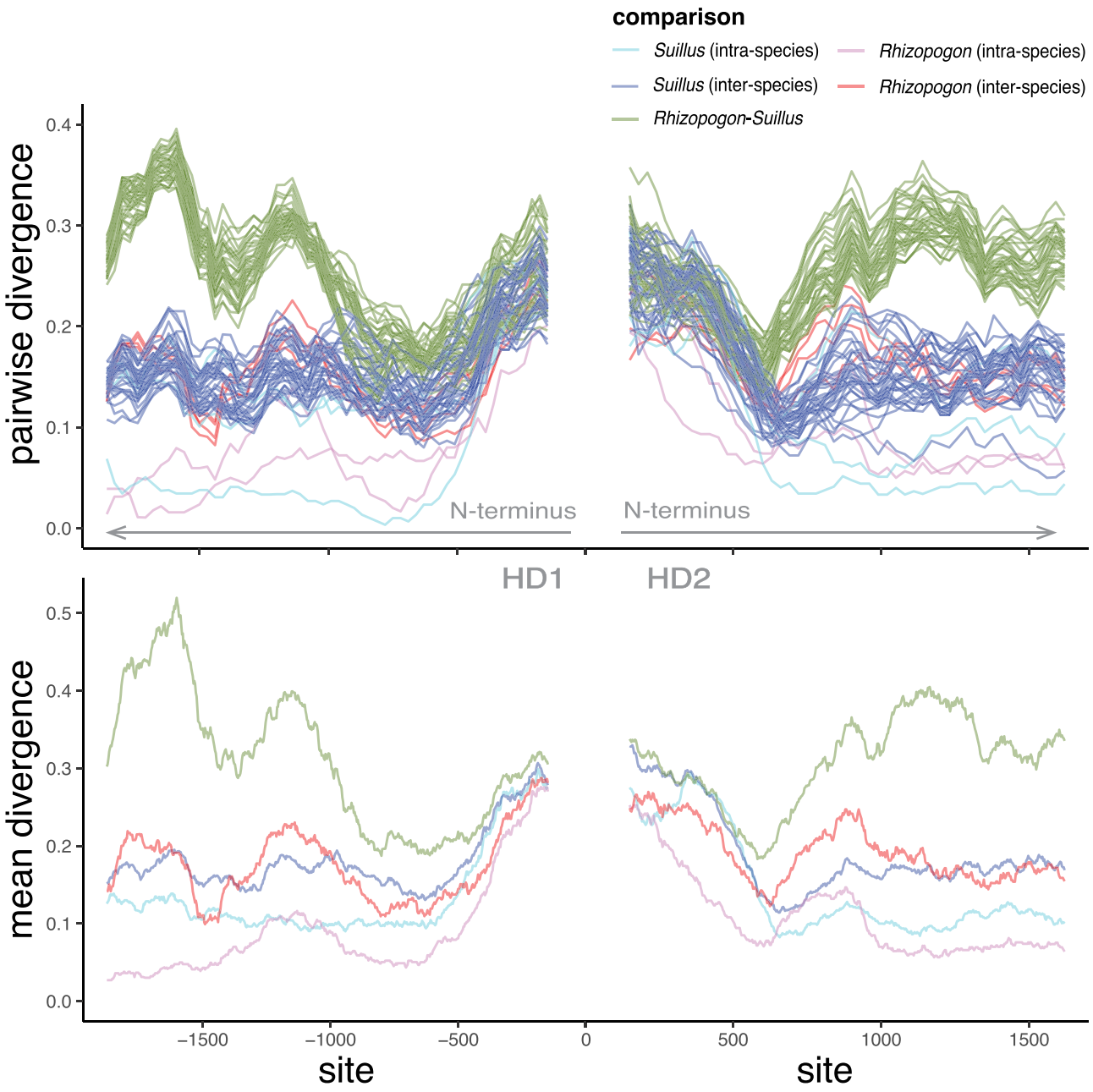

Fig. 3. Pairwise divergence of coding regions for *HD1* and *HD2* alleles from different suilloid fungi. Pairwise divergence is calculated from 500-bp overlapping windows from the alignments of locally *de novo* assembled haplotypes. At the center (N-terminus) of *HD MAT* locus, the pairwise divergence is similar regardless the species and the genus, suggesting strong balancing selection and extended trans-specific polymorphism. Toward the end (C-terminus) of *HD MAT* locus, the alleles from the more closely related taxa tend to be more similar, suggesting less trans-specific polymorphism in those regions. However, the sequence divergence between *Suillus* and *Rhizopogon (green lines)* is lower at the center region, suggesting the sequences in the center are more conserved in the long term.

genealogy of *HD* alleles rejects the hypothesis of a persisting biallelic system in *Suillus*. We conclude that the *HD MAT* locus of the common ancestor of suilloid fungi was multiallelic and has not changed to biallelic system before any common ancestors of the sampled diploid genomes.

Synteny of *HD MAT* loci

The synteny of the four previously identified genes located adjacent to the *HD MAT* locus (*MIP*, *HD1*, *HD2*, and beta-flanking gene) was maintained in *Suillus* and *Rhizopogon* spp. (Figure 6) (Coelho *et al.* 2017; Mujic *et al.* 2017). Although the beta-flanking gene was not annotated in *S. weaverae* due to the poly-N uncertain sequence in that region, it was later identified in both *de novo* assembled alleles from their alignment to beta-flanking genes in other species. Glycosyltransferase family 8 protein *(GLGEN)* was another gene always detected near *HD* genes in the sampled suilloid genomes, but its position relative to the *HD* genes was not conserved across different species. Two additional, unidentified genes were detected within the *HD MAT* region in more than one genome. One of the two genes was consistently present in all species and BLASTs to other fungal genomes, which implies it is likely a functional gene of unknown function. Other genes were only

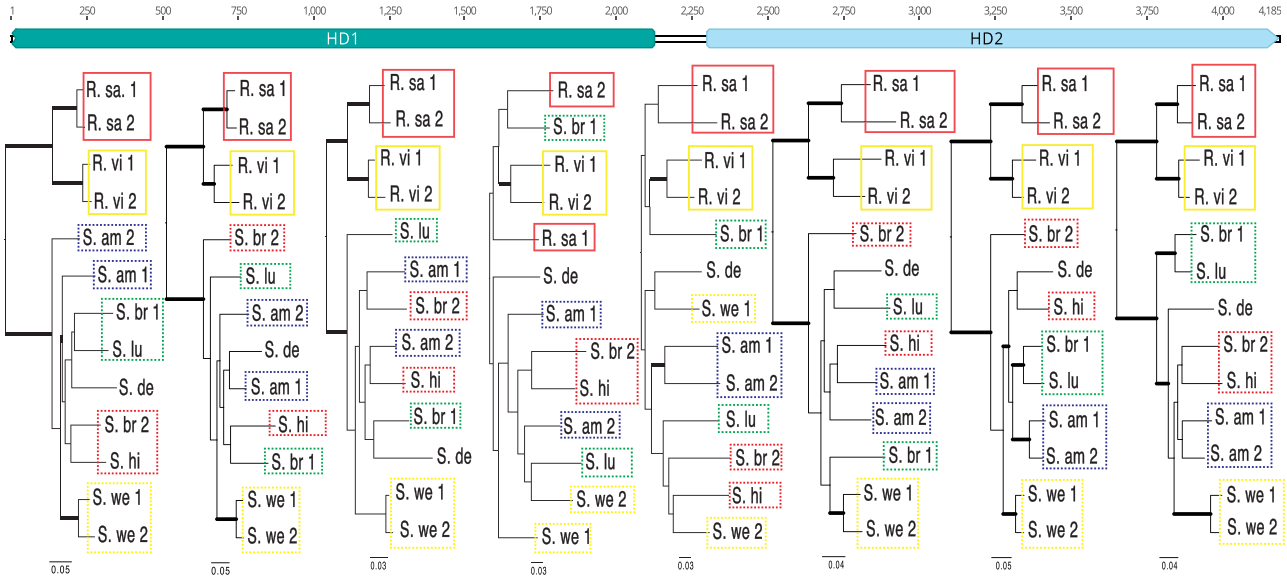

Fig. 4. Genealogy of *HD MAT* haplotypes. Genealogy of different segments of HD MAT locus. The branches of bootstrap values >70 is shown in bold. The colored squares indicate haplotypes in the same tip bipartitions at the *HD2* end of the sequences (not necessarily from the same species). The topology is not consistent across the locus, implying recombination during the evolution.

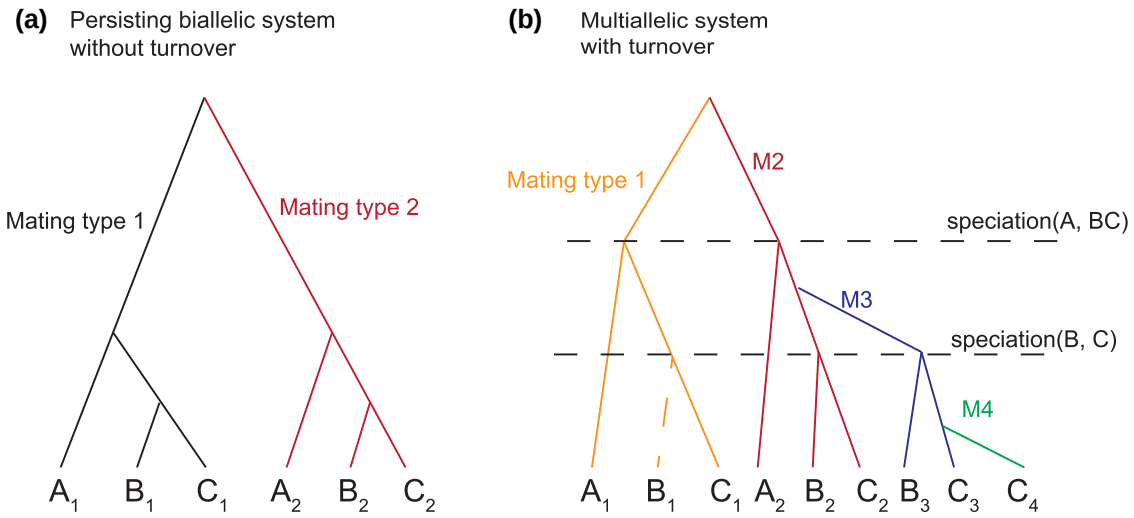

Fig. 5. Possible evolutionary scenarios for different breeding systems across species. A, B, C indicate different species with mating type 1 to 4. (a) Without new mating type production, species share the same two alleles. Each individual must have alleles from each clade. (b) With allele turnover, genealogy of alleles does not form a deep bisect, even with two species only have two alleles (species A and species B). B_1 allele is lost in the latter scenario, but the existence of B_3 enables the persistence of forming heterozygotes at the *MAT* locus. However, the two alleles from species B (B_2 and B_3) coalesces to each other earlier than to any alleles of species A, unlike the bisecting structure found in the first scenario.

found beside *HD MAT* in one genome, which include actin-related protein in *S. hirtellus*, aminoacyl-tRNA synthetase in *S. weaverae*, ATP-dependent DNA helicase in *S. americanus*, as well as several other genes of unknown function. In the *S. weaverae* assembly, we discovered a structural variant with a large insertion between *beta-flanking* gene and *GLGEN* which contained two partial *HD2* and partial *beta-flanking genes*. This insertion was only observed in one haplotype, whose mapping depth was also half compared to other regions (Supplementary Fig. 1). This implies the duplication of *HD* genes was involved in the structural variation near *HD MAT* locus of suilloid fungi.

Discussion

Monokaryotic and dikaryotic genomes in suilloid fungi

In the course of mapping Illumina reads onto their reference genomes, we observed contrasting patterns of read-mapping depth at SNP sites among eight suilloid fungi, which demonstrated that three genomes were monokaryotic. The *S. luteus* UH-Slu-Lm8-n1 genome was first isolated as a single-basidiospore isolate and thus was expected to be monokaryotic (Kohler *et al.* 2015). The genomes of *S. decipiens* and *S. hirtellus* also showed little SNPs in

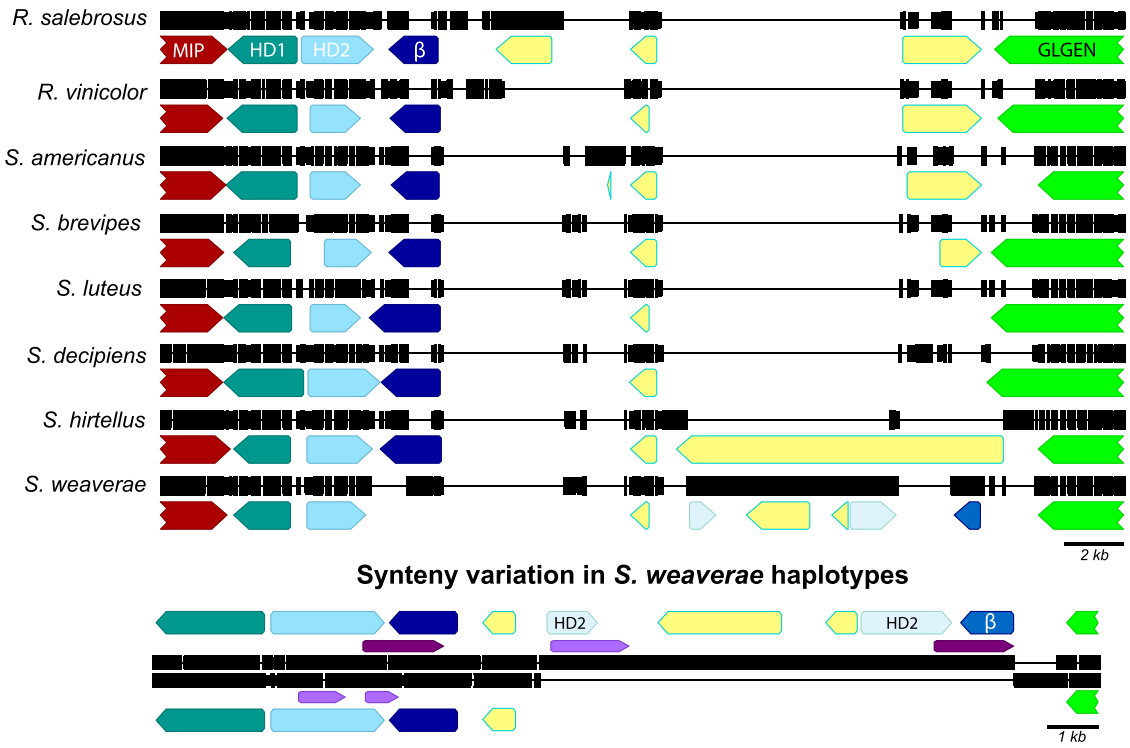

Fig. 6. Synteny of *HD MAT* locus in suilloid species. Synteny of HD MAT locus in suilloid species based on the alignments and the annotations in reference genomes (upper panel) and haplotypes (lower panel). Upper panel: The synteny is presented as alignment across reference sequences. Tall and thin columns indicate nucleotides and gaps in the alignment. Lower panel: structural variation between two alleles in *S. weaverae*, where one allele has a long insertion homologous to partial *HD2* and *beta-flanking* genes from both alleles (light and dark purple arrows indicate the respective similar regions). The other allele has the structure similar to other suilloid genomes. *MIP*: mitochondrial intermediate peptidase; *HD1*: homeodomain encoding protein type 1; *HD2*: homeodomain encoding protein type 2; β: beta flanking gene; *GLGEN*: Glycosyltransferase family 8 protein.

the sequenced strains and possessed a single *HD MAT* allele. Both were first isolated as pure cultures from fruiting bodies (*S. hirtellus*) or mycorrhizal root tips (*S. decipiens*) (Liao HL, pers. Comm.) and thus were initially presumed to be dikaryotic. Microscopic observation of clamp-connections and numbers of nuclei per cell are the indicators of the dikaryotic stage in fungi of Basidiomycota. However, those microscopic methods are difficult to apply to suilloid fungi, as most species do not consistently form clamp-connections as in other species (Pantidou and Groves 1966; Fries and Neumann 1990) and the numbers of nuclei in dikaryotic mycelia vary (Horton 2006; Sawada *et al.* 2014). We suspect that these isolates were dikaryotic when first isolated but later became monokaryotic via dedikaryotization due to losing one nucleus during the successive culturing (Santra and Nandi 1977; Leal-Lara and Eger-Hummel 1982). Additional study of ploidy and *MAT* locus in fruiting bodies, mycorrhizas and cultures of those species is required to address the ploidy variation during different phases of their life cycles.

Breeding system of suilloid fungi

The inferred multi-allelic *HD MAT* locus aligned with the known breeding systems in other suilloid fungi. For a few species of suilloid fungi whose sexual spores were able to germinate as pure cultures, their breeding systems were reported in previous studies. The breeding systems of *Rhizopogon rubescens* (Kawai *et al.* 2008), *Suillus variegatus* (Fries 1994), *S. bovinus* (Fries and Sun 1992), *S. luteus*, and *S. granulatus* (Fries and Neumann 1990) were all reported to be bipolar and multi-allelic. Which *MAT* locus is responsible for self-incompatibility is unknown since no further sequencing of *MAT* loci was performed in those culture-based studies. Our inference and the reported breeding systems in those culture-based studies showed that suilloid fungi may be generally multi-allelic at *HD MAT* locus across the clade.

Patterns of genetic variation at *HD MAT* locus

The synteny around suilloid *HD MAT* locus follows the classic form of *MIP-HD1-HD2-beta-flanking* configuration reported in other fungi in Basidiomycota (Kües *et al.* 2011; Coelho *et al.* 2017; Mujic *et al.* 2017). Structural variation arising from gene duplication of *HD genes* resulting in deviation from the canonical gene order has been reported in some basidiomycete genomes, e.g. *Coprinopsis cinereus* (Kues and Casselton 1993) and *Schizophyllum commune* (Ohm *et al.* 2010). This was observed here for one *HD MAT* allele in *S. weaverae* as well, in which one of the *HD MAT* alleles was partially duplicated. The extent and the frequency of non-canonical synteny variants for *HD MAT* locus in natural populations are unclear, but their rarity in suilloid fungi suggests that those types of structural change are likely to be deleterious, causing purifying selection to eventually eliminate the duplicated *HD* gene fragments. The coincidence of low variation at C-terminus of the *S. weaverae HD MAT* locus relative to other suilloid fungi implies the recombination around the *HD MAT* locus in *S. weaverae* may occur more frequently than that in other species.

The sequence divergence and the genealogy revealed different evolutionary processes in different regions of *HD MAT* locus. Inter-specific variation among suilloid *HD MAT* haplotypes was highest in the center region located between *HD1* and *HD2*, including the N-termini regions of each gene. This region appears to be under the strongest balancing selection and thus is likely to directly determine mating types. This agrees with molecular biology

experiments in *C. cinerea* (Banham *et al.* 1995) and *Ustilago maydis* (Kahmann and Bolker 1996) that the *HD* protein recognition is located at the N-terminus and the changes of sequences in this region can sometimes alter mating specificity. The absence of trans-specific polymorphism toward the C-terminus suggests this region is linked to but not directly under balancing selection. The strongest contrast between C- and N-terminus divergence was observed in *S. weaverae*. Although the N-terminus has equally diverged and trans-specific polymorphic as other pairs of alleles, the divergence of its C-terminus is more recent than all other pairs of alleles. A similar pattern was observed in *Sporidiobolus salmonicolor* (Coelho *et al.* 2010), *C. cinereus* (Badrane and May 1999), and *Heterobasidion* (van Diepen *et al.* 2013) where the N-terminus from different haplotypes was highly diverged and loaded with mutations while the sequence divergence of C-terminus is lower. Although stronger trans-specific polymorphism at the N-terminus region appears in the sampled suilloid fungi, the older divergence time of N-terminus regions however corresponds to lower sequence divergence at the genus level. This is contrary to the expectation that longer divergence time results in higher sequence divergence given similar mutation rates. This conflict implies the mating type determining N-terminus has stronger functional constraints than the C-terminus does, so the sequences of the N-terminus within each mating type are more conserved than those of C-terminus. Thus, according to our data in the suilloid clade, the origin of N-terminus in different mating types of *HD* genes is ancient yet the sequences in each mating type are conserved.

Local *de novo* assembly approach for *HD MAT* locus

High sequence divergence between two alleles at *HD MAT* locus impedes the common mapping variant detection approach from capturing the full set of variants at *HD MAT* locus. When reads from highly divergent alleles cannot be completely mapped to the sequence in the reference genome, variants are overlooked and polymorphisms are underreported. The local *de novo* assembly approach (aTRAM) provides a practical procedure and haplotype-level resolution to investigate the sequence evolution at *HD MAT* locus from genome shotgun sequencing. Phasing SNPs from next-generation sequencing reads to haplotypes requires information on neighbor variants spanned by paired reads. Longer reads and higher read variant density increase the ability to phase the haplotype blocks. Due to the nature of high variant density at *HD MAT* locus, phasing is achievable with short reads. Although *de novo* assembly or phasing is not normally required for describing short variants in genomes, in this case, the local *de novo* assembled haplotypes provided valuable information in understanding the evolution of *HD MAT* locus at the allele level. Accurate genetic diversity, different patterns of evolution within N-terminus vs C-terminus, and genealogy at the allele level would not have been possible with the common mapping variant detection approach.

Studying the signatures of evolution can help to understand genetic processes. Here, we demonstrate how the existence of balancing selection can be used to identify breeding systems of fungi from genomic data. This approach using local *de novo* assembly to study *HD MAT* locus should be widely applicable for analysis of genetic diversity, genealogy, recombination, and patterns of selection at mating-type determining loci. For example, the actual strength of balancing selection at *MAT* loci can be estimated from sequence data as the studies in *C. cinereus* (May *et al.* 1999) and in *Trichaptum* species (Peris *et al.* 2022). Such calculation can also be achieved with this culture-free genomic approach given enough sample size for each species. This genome-based method for assessing sequence divergence and genealogy provides an alternative approach to understanding the breeding systems and evolution of mating-type alleles in fungi even in the absence of crossing studies. This opens a path to expand our knowledge of breeding systems in non-culturable organisms based on the interplay between evolution and genetics.

Data availability

The genome shotgun sequencing reads and assembly of the new genome *Rhizopogon salebrosus* TDB-379 are available through BioProjects PRJNA333299. Other genomes are accessible from NCBI with the respective accession numbers in Table 1. *De novo* assembled *HD MAT* sequences are available in GenBank with accession numbers ON315855-ON315867. Strains are available upon request. Customized scripts for data analysis are available at https://github.com/KeFungi/SUIMAT_pub.

Supplemental material available at GENETICS online.

Acknowledgments

We are grateful to Lotus Lofgren for providing advice and useful suggestions for this manuscript. We thank the Office of Information Technology and Research Computing team at Duke University for maintaining and providing computational resources.

Funding

This work was supported by the National Science Foundation Grants (DEB 1554181) to RV and (IOS 2029168) to SB and H-LL.

Conflicts of interest

The author(s) declare no conflict of interest.

Literature cited

Allen JM, Huang DI, Cronk QC, Johnson KP. Atram—automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. BMC Bioinformatics. 2015;16(1):98. doi:10.1186/s12859-015-0515-2.

Badrane H, May G. The divergence-homogenization duality in the evolution of the b1 mating type gene of *Coprinus cinereus*. Mol Biol Evol. 1999;16(7):975–986. doi:10.1093/oxfordjournals.molbev.a026187.

Banham AH, Asante-Owusu RN, Gottgens B, Thompson S, Kingsnorth CS, Mellor E, Casselton LA. An N-terminal dimerization domain permits homeodomain proteins to choose compatible partners and initiate sexual development in the mushroom *Coprinus cinereus*. Plant Cell. 1995;7(6):773–783. doi:10.1105/tpc.7.6.773.

Branco S, Gladieux P, Ellison CE, Kuo A, LaButti K, Lipzen A, Grigoriev IV, Liao HL, Vilgalys R, Peay KG, *et al.* Genetic isolation between two recently diverged populations of a symbiotic fungus. Mol Ecol. 2015;24(11):2747–2758. doi:10.1111/mec.13132.

Buller AHR. The diploid cell and the diploidisation process in plants and animals, with special reference to the higher fungi. Bot Rev. 1941;7(8):389–431. doi:10.1007/BF02872469.

Cairney JW, Chambers SM. Ectomycorrhizal Fungi: Key Genera in Profile. Berlin: Springer; 2013.

Chevreux B, Wetter T, Suhai S. Genome sequence assembly using trace signals and additional sequence information. Proceedings of the German Conference on Bioinformatics (GCB). 1999;99(1):45–56.

Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T. Fungal sex: the basidiomycota. Microbiol Spectr. 2017;5(3):3. doi:10.1128/microbiolspec.FUNK-0046-2016.

Coelho MA, Sampaio JP, Goncalves P. A deviation from the bipolar-tetrapolar mating paradigm in an early diverged basidiomycete. PLoS Genet. 2010;6(8):e1001052. doi:10.1371/journal.pgen.1001052.

Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004;14(7):1394–1403. doi:10.1101/gr.2289704.

Edge P, Bafna V, Bansal V. Hapcut2: robust and accurate haplotype assembly for diverse sequencing technologies. Genome Res. 2017;27(5):801–812. doi:10.1101/gr.213462.116.

Fries N. Basidiospore germination in some mycorrhiza-forming hymenomycetes. Trans Br Mycol Soc. 1978;70(3):319–324. doi:10.1016/S0007-1536(78)80128-4.

Fries N. Ecological and evolutionary aspects of spore germination in the higher basidiomycetes. Trans Br Mycol Soc. 1987;88(1):1–7. doi:10.1016/S0007-1536(87)80179-1.

Fries N. Sexual incompatibility in *Suillus variegatus*. Mycol Res. 1994;98(5):545–546. doi:10.1016/S0953-7562(09)80477-2.

Fries N, Neumann W. Sexual incompatibility in *Suillus luteus* and *S. granulatus*. Mycol Res. 1990;94(1):64–70. doi:10.1016/S0953-7562(09)81265-3.

Fries N, Sun YP. The mating system of *Suillus bovinus*. Mycol Res. 1992;96(3):237–238. doi:10.1016/S0953-7562(09)80974-X.

Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, *et al.* High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 2011;108(4):1513–1518. doi:10.1073/pnas.1017351108.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307–321. doi:10.1093/sysbio/syq010.

Horton TR. The number of nuclei in basidiospores of 63 species of ectomycorrhizal homobasidiomycetes. Mycologia. 2006;98(2):233–238. doi:10.1080/15572536.2006.11832695.

James TY. Why mushrooms have evolved to be so promiscuous: insights from evolutionary and ecological patterns. Fungal Biol Rev. 2015;29(3–4):167–178. doi:10.1016/j.fbr.2015.10.002.

James TY, Sun S, Li W, Heitman J, Kuo HC, Lee YH, Asiegbu FO, Olson A. Polyporales genomes reveal the genetic architecture underlying tetrapolar and bipolar mating systems. Mycologia. 2013;105(6):1374–1390. doi:10.3852/13-162.

Jukes TH, Cantor CR. Evolution of Protein Molecules. Mammalian Protein Metabolism. New York: Academic Press; 1969. p. 21–132.

Kahmann R, Bolker M. Self/nonself recognition in fungi: old mysteries and simple solutions. Cell. 1996;85(2):145–148. doi:10.1016/S0092-8674(00)81091-0.

Katoh K, Standley DM. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–780. doi:10.1093/molbev/mst010.

Kawai M, Yamahara M, Ohta A. Bipolar incompatibility system of an ectomycorrhizal basidiomycete, *Rhizopogon rubescens*. Mycorrhiza. 2008;18(4):205–210. doi:10.1007/s00572-008-0167-4.

Kniep H. Die Sexualität der Niederen Pflanzen. Jena: Gustav Fischer; 1928.

Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canback B, Choi C, Cichocki N, Clum A, *et al.* Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. Nat Genet. 2015;47(4):410–415. doi:10.1038/ng.3223.

Koltin Y, Stamberg J, Lemke PA. Genetic structure and evolution of the incompatibility factors in higher fungi. Bacteriol Rev. 1972;36(2):156–171. doi:10.1128/br.36.2.156-171.1972.

Krumbeck Y, Constable GWA, Rogers T. Fitness differences suppress the number of mating types in evolving isogamous species. R Soc Open Sci. 2020;7(2):192126. doi:10.1098/rsos.192126.

Kues U. From two to many: multiple mating types in basidiomycetes. Fungal Biol Rev. 2015;29(3–4):126–166. doi:10.1016/j.fbr.2015.11.001.

Kues U, Casselton LA. The origin of multiple mating types in mushrooms. J of Cell Sci. 1993;104(2):227–230. doi:10.1242/jcs.104.2.227.

Kües U, James TY, Heitman J. Mating type in basidiomycetes: unipolar, bipolar, and tetrapolar patterns of sexuality. In: Pöggeler S, Wöstemeyer J, editors. Evolution of Fungi and Fungal-Like Organisms. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 97–160.

Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9(4):357–359. doi:10.1038/nmeth.1923.

Leal-Lara H, Eger-Hummel G. A monokaryotization method and its use for genetic studies in wood-rotting basidiomycetes. Theor Appl Genet. 1982;61(1):65–68. doi:10.1007/BF00261513.

Li H. A statistical framework for snp calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987–2993. doi:10.1093/bioinformatics/btr509.

Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–595. doi:10.1093/bioinformatics/btp698.

Lofgren LA, Nguyen NH, Vilgalys R, Ruytinx J, Liao HL, Branco S, Kuo A, LaButti K, Lipzen A, Andreopoulos W, *et al.* Comparative genomics reveals dynamic genome evolution in host specialist ectomycorrhizal fungi. New Phytol. 2021;230(2):774–792. doi:10.1111/nph.17160.

Loytynoja A. Phylogeny-aware alignment with PRANK. Methods Mol Biol. 2014;1079:155–170. doi:10.1007/978-1-62703-646-7_10.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17(1):10–12. doi:10.14806/ej.17.1.200.

May G, Shaw F, Badrane H, Vekemans X. The signature of balancing selection: fungal mating compatibility gene evolution. Proc Natl Acad Sci U S A. 1999;96(16):9172–9177. doi:10.1073/pnas.96.16.9172.

Mujic AB, Kuo A, Tritt A, Lipzen A, Chen C, Johnson J, Sharma A, Barry K, Grigoriev IV, Spatafora JW. Comparative genomics of the ectomycorrhizal sister species *Rhizopogon vinicolor* and *Rhizopogon vesiculosus* (Basidiomycota: Boletales) reveals a divergence of the mating type B locus. G3 (Bethesda). 2017;7(6):1775–1789. doi:10.1534/g3.117.039396.

Nara K. Spores of ectomycorrhizal fungi: ecological strategies for germination and dormancy. New Phytol. 2009;181(2):245–248. doi:10.1111/j.1469-8137.2008.02691.x.

Nieuwenhuis BP, Billiard S, Vuilleumier S, Petit E, Hood ME, Giraud T. Evolution of uni- and bifactorial sexual compatibility systems in fungi. Heredity (Edinb). 2013;111(6):445–455. doi:10.1038/hdy.2013.67.

Ohm RA, de Jong JF, Lugones LG, Aerts A, Kothe E, Stajich JE, de Vries RP, Record E, Levasseur A, Baker SE, *et al.* Genome sequence of the

model mushroom *Schizophyllum commune*. Nat Biotechnol. 2010; 28(9):957–963. doi:10.1038/nbt.1643.

Pantidou ME, Groves JW. Cultural studies of boletaceae: some species of *Suillus* and *Fuscoboletinus*. Can J Bot. 1966;44(10): 1371–1392. doi:10.1139/b66-151.

Peris D, Lu DS, Kinneberg VB, Methlie IS, Dahl MS, James TY, Kauserud H, Skrede I. Large-scale fungal strain sequencing unravels the molecular diversity in mating loci maintained by long-term balancing selection. PLoS Genet. 2022;18(3):e1010097. doi: 10.1371/journal.pgen.1010097.

Raper JR. Tetrapolar sexuality. Q Rev Biol. 1953;28(3):233–259. doi:10.1086/399698.

R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2022.

Santra S, Nandi B. Monokaryotization of dikaryotic mycelia of some wood-destroying basidiomycetes in presence of toxic substances. New Phytol. 1977;79(1):179–181. doi:10.1111/j.1469-8137.1977.tb02194.x.

Sawada K, Wan JN, Oda K, Nakano S, Aimi T, Shimomura N. Variability in nucleus number in basidiospore isolates of *Rhizopogon roseolus* and their ability to form ectomycorrhizas with host pine roots. Mycol Progress. 2014;13(3):745–751. doi:10.1007/s11557-013-0957-8.

van Diepen LT, Olson A, Ihrmark K, Stenlid J, James TY. Extensive trans-specific polymorphism at the mating type locus of the root decay fungus heterobasidion. Mol Biol Evol. 2013;30(10): 2286–2301. doi:10.1093/molbev/mst126.

Wang B, Mitchell-Olds T. Balancing selection and trans-specific polymorphisms. Genome Biol. 2017;18(1):231. doi:10.1186/s13059-017-1365-1.

Whitehouse HL. Heterothallism and sex in the fungi. Biol Rev Camb Philos Soc. 1949;24(4):411–447. doi:10.1111/j.1469-185X.1949.tb00582.x.

Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer; 2016.

Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de bruijn graphs. Genome Res. 2008;18(5):821–829. doi:10.1101/gr.074492.107.

Editor: J. Stajich