Repetitive DNA in the automictic fungus *Microbotryum violaceum*

Michael E. Hood
Department of Biology, University of Virginia, Charlottesville, VA 22903, USA (Phone: + 1-804-243-5077; Fax: + 1-804-982-5626; E-mail: michael.hood@virginia.edu)

Abstract

The small genomes of fungi are expected to have little repetitive content other than rDNA genes. Moreover, among asexual or highly selfing lineages, the diversity of repetitive elements is also expected to be very low. However, in the automictic fungus *Microbotryum violaceum*, a very large proportion of random DNA fragments from the autosomes and the fungal sex chromosomes are repetitive in nature, either as retro-transposon or helicase sequences. Among the retrotransposon sequences, examples were found from each major kind of elements, including copia, gypsy, and non-LTR sequences. The most numerous were copia-like elements, which are believed to be rare in fungi, particularly among basidiomycetes. The many helicase sequences appear to belong to the recently discovered *Helitron* type of transposable elements. Also, sequences that could not be identified as a known type of gene were also very repetitive within the database of random fragments from *M. violaceum*. The differentiated pair of fungal sex chromosomes and suppression of recombination may be the major forces determining the highly repetitive content in the small genome of *M. violaceum*.

Introduction

The suggestion that eukaryotic genomes are in large measure forged by repetitive genetic elements is well supported. The clearest evidence comes from the proportion of nuclear DNA that is made up of transposable elements, which in many plants and animals is greater than a third of all sequences (Kidwell, 2002). There are also large costs to carrying transposable elements because their insertion throughout the genomes can disrupt the expression of functional genes (Casacuberta & Santiago, 2003) and leads to chromosomal rearrangements between elements at non-homologous sites (Lönnig & Saedler, 2002; Shnyreva, 2003). Eukaryotic systems have in turn evolved resistance mechanisms to protect against these mutating effects, and in some cases may have co-opted repetitive elements that now perform beneficial roles as regulatory or structural sequences (Kidwell & Lisch, 2001; Hurst & Werren, 2001).

Intense interest in repetitive elements has resulted in many recent reviews (Daboussi, 1997; Kempken & Kück, 1998; Brosius, 1999; Hurst & Werren, 2001; Kidwell & Lisch, 2001; Kidwell, 2002; Lönnig & Saedler, 2002; Shapiro, 2002; Wöstemeyer & Kreibich, 2002; Casacuberta & Santiago, 2003; Gorelick, 2003; Shnyreva, 2003). However, entirely new and widespread classes of transposable elements are still being discovered in eukaryotes (e.g. *Helitron*s, the rolling circle mobile element; Kapitonov & Jurka, 2001; Poulter, Goodwin & Bulter, 2003). Moreover, detailed genomic surveys and experimental manipulations are limited to a handful of model organisms. It may therefore still be too early to determine the general

"rules" about repetitive genetic elements until there is sufficient information across broadly divergent lineages to test the proposed correlations.

The two most general rules suggested for transposable elements are that their densities increase relative to total genome size and their diversity is dependent upon outcrossing of the host genome (e.g. Hurst & Werren, 2001; Kidwell, 2002). For example, even though each major kind of repetitive element is found throughout the kingdom Fungi (Kempken & Kück, 1998), it is frequently asserted that their small genomes (on the order of a few tens of million basepairs) are "streamlined" and contain proportionally much less repetitive DNA than do most plants and animals (Kidwell, 2002; Wöstemeyer & Kreibich, 2002). Furthermore, while fungi exhibit a broad range of mating systems, many are predominantly asexual or are highly selfing, and the rate of outcrossing is also thought to be positively related to the spread and to the diversity of repetitive elements within the lineage (Hurst & Werren, 2001). However, some organisms with very rare outcrossing have considerable diversity of repetitive elements (i.e. *Cryptococcus neoformans*; Goodwin & Poulter, 2001), and how completely an organism must restrict outcrossing to decrease repetitive element diversity is unknown. Here, I describe repetitive DNA in the genome of *Microbotryum violaceum*, a highly selfing basidiomycete fungus (Antonovics, Hood & Partain, 2002; Garber & Ruddat, 2002; Hood & Antonovics, 2004). The study is based on a genome sampling approach that could also be used to investigate the nuclear content in other organisms where complete genome sequencing is not feasible. The results show that, contrary to expectation, the genome of *Microbotryum violaceum* is exceptionally rich in both the density and diversity of retrotransposons and other repetitive elements.

Methods and materials

Study system

Microbotryum violaceum causes anther-smut disease in many members of the plant family Caryophyllaceae. Diploid spores of the fungus are produced in the anther tissues of infected plants and are carried between plants by insect pollinators. As the spores germinate, they undergo meiosis, and the haploid products then mate prior to infection of the new host. However, the haploid products can be cultured in the laboratory, where they grow as yeast-like sporidia. In nature, mating is very rapid after meiosis and often is between products of the same tetrad. This kind of mating system, in conjunction with a centromere-linked mating type locus, results in automixis (i.e. intra-tetrad mating) with "central fusion", a form of selfing that has the surprising effect of maintaining heterozygosity in regions linked to centromeres throughout the genome because of the segregation of centromeres at meiosis I (Hood & Antonovics, 2004). Furthermore, the chromosome pair containing the mating type locus has characteristics similar to sex chromosomes in diploid-mating plants and animals (i.e. size dimorphism, suppressed recombination, accumulation of transposable elements relative to autosomes, and unequal density of functional genes; Hood, 2002; Hood, Antonovics & Koskella, 2004). These chromosomes containing the mating-type locus are referred to here as fungal sex chromosomes, which are analogous to the sex chromosomes in other organisms that determine mating compatibility in the haploid stage (i.e. in bryophytes) (Hood et al., 2004). *Microbotryum violaceum* is a species complex consisting of highly differentiated lineages that are host-specific to varying degrees. For this study, a single genotype of *M. violaceum* was collected in Italy (Lamole nr. Greve in Chianti) from the host *Silene latifolia*. The fungal sex chromosomes are approximately 3.1 mega-basepairs (mbp) for the A1 and 3.8 mbp for the A2. The total haploid genome size of *M. violaceum* is roughly 25 mbp, which is comparable to other ascomycete and basidiomycete fungi (Wöstemeyer & Kreibich, 2002).

Molecular techniques

The methods for DNA isolation and sequencing have been described previously (Hood et al. 2004) and are presented briefly as follows. A linear meiotic tetrad of *M. violaceum* was isolated by micromanipulation following spore germination on potato dextrose agar. Electrophoretic karyotypes were prepared for two haploid cultures from this tetrad that were of opposite mating types (A1 and A2). Pulsed-field electrophoresis was

conducted under standard run conditions for separating chromosomes in the range of 1–4 mbp (Hood et al., 2004). From the electrophoretic karyotypes separate samples of DNA were isolated: bands representing the A1 and A2 fungal sex chromosomes and a pooled sample of bands representing the autosomes. DNA from these genomic regions was extracted separately, incompletely triple-digested with 4–6 bp-cutting restriction enzymes (*Rsa*I, *Msc*I, and *Dra*I), size fractionated for fragments between 200 and 1000 bp, cloned, and sequenced. This sampling approach resulted in 684 DNA sequences averaging 470 bp in length, including 499 sequences from the fungal sex chromosomes (257 from the A1 and 242 from the A2 sex chromosomes) and 185 from the pool of autosomes. Sequence data are available from the GenBank Data Libraries under the accession numbers BZ781929 to BZ782612 (www.ncbi.nlm.nih.gov).

Sequence data analysis

The sequence fragments from *M. violaceum* were compared against the DNA and protein databases available through the Nation Center for Biotechnology Information (NCBI) using the BLASTn and BLASTx programs (www.ncbi.nlm.nih.gov/BLAST/). This search was used to assign *M. violaceum* fragments to categories on the basis of the type of sequence matched in the BLAST searches. The following categories were distinguished: retrotransposon-related (i.e. retrotransposons, reverse transcriptases, gag or pol genes, polyproteins, or retrovirus sequences), helicases, rDNA genes, and a catch-all category of identifiable "other genes". The cut-off for significance of BLAST hits was an E-value of 10^{-4}. E-values below 10^{-2} are nearly equivalent to P-values, and even using the Bonferroni correction for the 684 multiple comparisons, the chosen cut-off of 10^{-4} approximates a significance level of 0.06. Remaining sequences are categorized as "no-hit" sequences.

Additional database searches were used to further characterize repetitive sequences. Tandem repeat sequences were identified using the Tandem Repeats Finder 3.01 software with the default search parameters (Benson, 1999; http://tandem.biomath.mssm.edu/trf.html). All of the sequences from *M. violaceum* were compared for DNA

similarity among themselves using the Standalone BLASTn software, again with $E < 10^{-4}$ as a cut-off value for significant hits. The numbers of hits for different sequence categories were compared using the LSMEANS feature of PROC GENMOD in SAS (SAS Institute, Cary, NC). Finally, full-length protein sequences of representative transposable elements of each major kind of Class I (RNA-based) retrotransposons and Class II (DNA-based) transposons from other organisms were used to search against the *M. violaceum* database using the translated tBLASTn program. The sequences of the representative examples of transposable elements from other organisms are identified in Figure 1 and the text below.

To assess the diversity within and among Class I retrotransposons in *M. violaceum*, distance analyses between *M. violaceum* sequences and the representative Class I elements from other organisms were performed using translated protein alignments. Seven conserved domains are recognized across reverse transcriptase proteins of retrotransposon and retroviruses (Xiong & Eickbush, 1988). Ten *M. violaceum* sequences could be aligned with a section spanning ca. 100 residues of conserved regions 1 and 2 (as defined in Xiong & Eickbush, 1988), and 12 *M. violaceum* sequences could be aligned with a section spanning ca. 60 residues of conserved regions 4 and 5. Thus, two independent dendrograms were obtained. One *M. violaceum* sequence was long enough to align with both sections and was therefore included in both dendrograms. Alignments were prepared using ClustalW (www.ebi.ac.uk/clustalw/), with minor adjustments performed manually only to ensure that the universally conserved residues were aligned. Neighbor-joining analysis was performed using the Mega 2.1 software (Kumar et al., 2001) with 100 bootstrap replications.

A distance analysis was also used to compare helicase fragments from *M. violaceum* to *Helitron*-like transposable elements from several other organisms. Because *Helitron* sequences are less diverse than retrotransposons, a larger region of protein sequence alignment was necessary. First, a composite *M. violaceum* sequence was made by aligning the fragments with a representative *Helitron* transposable element from *Oryza sativa* (protein sequence accession number BAA92200 translated from DNA sequence
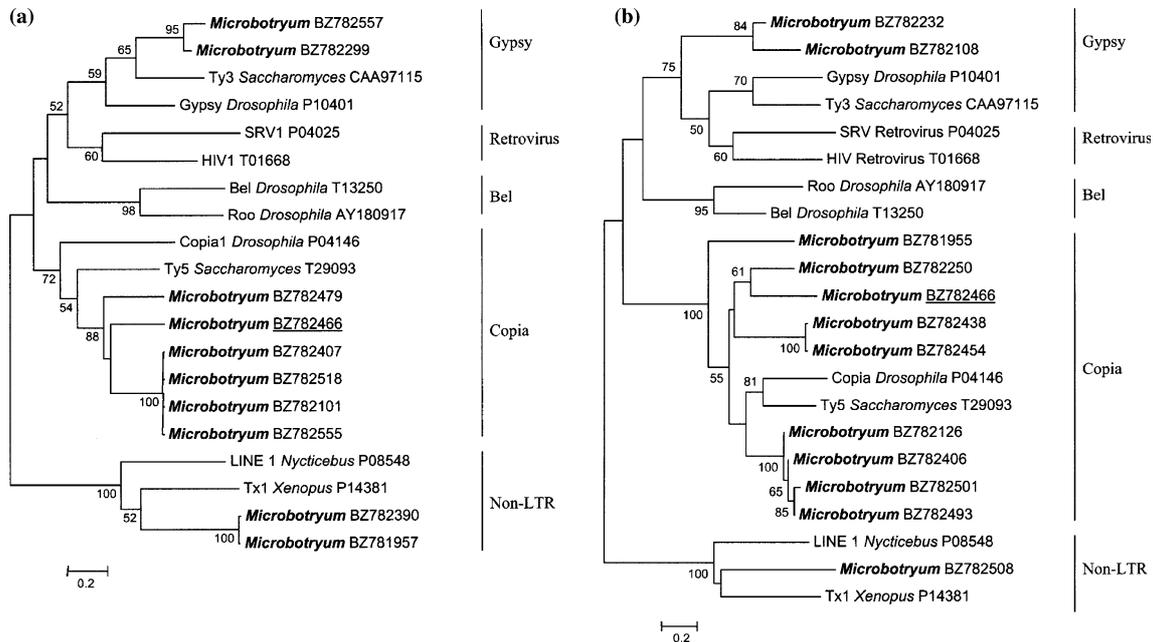
Figure 1. Neighbor-joining analysis of Class I retrotransposon sequences from *M. violaceum* aligned with conserved regions of reverse transcriptase protein sequences (as defined in Xiong and Eickbush, 1988). (a) Alignment with conserved regions 1 and 2. (b) Alignment with conserved regions 4 and 5. One sequence from *M. violaceum* (accession number BZ782466; underlined) is included in both analyses. Other sequences include two examples from each major kind of Class I element. All fragments are identified by accession numbers. Node support reflects 100 bootstrap replication, and dendrograms are midpoint rooted.

accession AP001278 as in Poulter et al., 2003). The *M. violaceum* fragments provided ca. 90% coverage of a 900 residue region of the *Helitron* protein from *Oryza*. Using ClustalW, a protein sequence from each host species in the *Helitron* (helicase) phylogeny of Poulter et al. (2003) was then aligned with the composite *M. violaceum* sequence, and a neighbor-joining analysis was performed using Mega 2.1. Because this analysis of *Helitron* sequences was unconventional in combining multiple fragments from the *M. violaceum* genome, two control tests were conducted to increase confidence in the results. First, where multiple *M. violaceum* fragments overlapped in the alignment, their separate distances from all other hosts species were estimated using only the region of overlap. Second, all combinations of substituting different overlapping fragments into the composite *M. violaceum* sequence were tested for effects on the overall results. Moreover, in the previous analysis by Poulter et al. (2003), multiple helicase sequences from the same species were monophyletic with 100% bootstrap support in all cases ($n = 6$).

Results

The random sampling approach to genome analysis of *M. violaceum* provided approximately 3 to 4% coverage of the A1 and A2 sex chromosomes and approximately 0.4% coverage of the autosomes based upon sampling roughly equal numbers of clones from each of the different genomic regions. Relative coverage of these regions was used to estimate the contributions of each of the sequence categories (identified by the BLAST searches) to the total genome size of *M. violaceum* (Table 1).

Among the categories of repetitive genetic elements (rDNA genes, retrotransposons, and helicases), the largest contribution to the genome of *M. violaceum* was made by retrotransposon-related sequences (Table 1). The retrotransposons included samples of each major kind of Class I elements. In both distance analyses based on the conserved regions (Figure 1a, b), *M. violaceum* sequences clustered with the copia, gypsy, and non-LTR (long-terminal repeat) elements from other organisms with strong bootstrap support.

Table 1. Contribution of DNA sequence categories to the genome of *M. violaceum*

	Proportion of genome	Number of sequences	rDNA genes	Retro-transposons	Helicases	Other genes	No-hit sequences
Fungal sex chromosomes	0.14	499	15 (3.0%)	72 (14.4%)	14 (2.8%)	48 (9.6%)	350 (70.1%)
Autosomes	0.86	185	10 (5.4%)	12 (6.5%)	3 (1.6%)	39 (21.1%)	121 (65.4%)
Genome			5.1%	7.6%	1.8%	19.4%	66.1%
Repetitiveness within database			3.68[A]	1.82[B]	1.47[B(C)]	0.07[D]	0.84[C]

Estimates of total genomic contribution are based on the percentage of fungal sex chromosome and autosome sources of DNA made up by the different sequence categories, and the relative contribution of these regions to genome size. Sequence fragments matching rDNA genes were identified by the BLASTn search, while those matching other type of sequences were identified by BLASTx search against global databases at NCBI. Repetitiveness measure indicates the number of Standalone BLASTn hits (with the $e < 10^{-4}$) for sequences within each category (combining sex chromosome and autosome sequences) to other sequences within the database of fragments; significant differences ($P < 0.05$) between categories as determined by LSMEANS procedures in SAS are indicated by the superscript letters.

The single *M. violaceum* sequence that was included in both distance analyses was consistently associated with the copia elements. In cases where bootstrap support for grouping *M. violaceum* elements with particular element kind were less than 100 (i.e. the putative Gypsy elements in Figure 1a, b), the strongest BLASTx match in the NCBI databases were also to members of the same kind of element.

Searching the *M. violaceum* database with each of the Class I elements from other organisms led to a categorization of a much larger number of sequences because it provided a method for characterizing sequences that could not be aligned with the conserved regions of reverse transcriptase. The validity of this approach was supported by the fact that in all cases where the same sequence were involved ($n = 18$), the distance analyses (Figure 1) produced the same categorization as the BLAST search against the *M. violaceum* database. Among the 84 retrotransposon-related sequences, there were 49 matches to copia elements (a list of accession numbers by element category is available by request). Also, there were 15 gypsy elements, 5 non-LTR elements. The remaining 15 sequences could not be confidently classified to any particular retrotransposon type by the methods described above, but the strongest BLASTx matches in the NCBI databases were copia elements for 9 sequences and were gypsy elements for 6 sequences. The searches against the *M. violaceum* database

using the retroviral or Bel retrotransposons from Figure 1 yielded no significant hits. Similarly, searches using various representatives of the DNA-based Class II transposons yielded no significant hits, including the sequences of *Mariner* (accession number AAC28384), the *Mariner*-type *Impala* (AAB33090), *Fot1* (CAA46025), and the *hAT*-type *Tascot* (CAA68959).

Helicase sequences were very common among the isolated fragments (Table 1; accession numbers available by request). The distance analysis
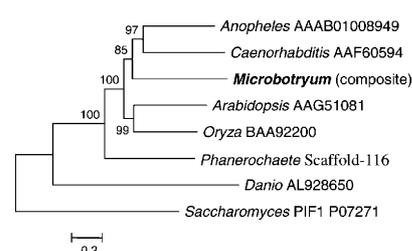

Figure 2. Neighbor-joining analysis of helicase protein sequences from *M. violaceum* and other model *Helitrons*. The sequence from *M. violaceum* represents a composite of fragments totaling ca. 900 amino acids. Fragments are identified by accession numbers of protein sequences, but DNA sequences are indicated for *Anopheles* AAAB01008949 (bases 255804 to 251093), *Danio* AL928650 (bases 48162 to 51749), *Phanerochaete* Scaffold-116 (bases 77157 to 74875; www.jgi.doe.gov). Node support reflects 100 bootstrap replication, and the dendrogram was rooted to the yeast mitochondrial PIF1 (P07271) of the RAD3/XPD helicase subfamily (as in Poulter et al., 2003).

clustered helicase of *M. violaceum* among the *Helitron* transposable elements from unrelated organisms and distant from the outgroup, a non-*Helitron* helicase of *Saccharomyces* (Figure 2). The control tests to support the use of a composite helicase sequence showed that overlapping *M. violaceum* fragments clustered together when compared with the same region from the other species (in 4 out of 5 cases), and substituting any particular overlapping fragment into the composite helicase sequence had no effect on the position of *M. violaceum* in the dendrogram nor a large effect on bootstrap values (data not shown). Also, where *Helitron* fragments overlapped they showed a high degree of DNA sequence identity (average 92%).

Tandem repeat sequences were identified within 22 *M. violaceum* fragments. These included 3 microsatellites of 10-, 12-, and 19-tandem repeat trimers and 22 satellite sequences ranging in size from 9 to 84 bp and in repeat copies from 1 to 6.

In the comparison of *M. violaceum* sequence fragments against each other using Standalone BLASTn, the various categories of fragments differed in the average number of hits to other *M. violaceum* sequences. The average significant hit in this search was an alignment over 280 bp of sequence lengths and had 96% basepair identity ($n = 490$ significant BLASTn hits). The category that was not expected to be repetitive in nature (i.e. the "other genes" category) generally consisted of single-copy sequences within the database (i.e. they hit no other sequence fragments). rDNA genes, retrotransposons, and helicases frequently had multiple hits, significantly more so than the "other genes" category (Table 1). "No-hit" sequences also had more hits to other sequences than did "other genes." The data suggest that some of these sequences belong to one of the other repetitive categories because, although these "no hits" did not match any such sequences in the NCBI databases, several did match other sequences in *M. violaceum* that had been identified as repetitive categories by the BLAST search against NCBI databases. Among these Standalone BLASTn matches, 15 "no-hit" sequences had significant similarity to some of the 84 *M. violaceum* retrotransposons, 3 had significant similarity to the 17 helicase fragments, but none showed similarity to the 86 sequences in the "other genes" category. However, even by

removing these 18 "no-hit" fragments that matched retrotransposon or helicase sequences in Standalone BLAST, the average number of matches to other sequences for "no-hit" fragments (0.7 hits) remained significantly higher than that for the "other genes" category (0.1 hits) ($X^2 = 18.9$, d.f. $= 1$, $P < 0.001$). Fifteen "no-hit" sequences had 4 or more matches only to other "no-hit" sequences.

Within the category of retrotransposons, those originating from the fungal sex chromosomes had significantly more Standalone BLASTn hits within the *M. violaceum* database than retrotransposons originating from the autosomes: sex chromosome retrotransposons averaged 2 hits to other sequences ($n = 73$), and autosome retrotransposons averaged 1 hit to another sequence ($n = 12$; $X^2 = 4.8$, d.f. $= 1$, $P = 0.029$). A significant difference in the same direction was also found between "no-hit" sequences originating from the fungal sex chromosomes (1.9 hits [$n = 350$]) and those from autosomes (1.6 hits [$n = 121$]) ($X^2 = 14.7$, d.f. $= 1$, $P < 0.001$). Helicase sequences from the fungal sex chromosomes were also more repetitive, but not significantly so.

Discussion

The small eukaryotic genomes of fungi are considered, as a rule, to contain little repetitive DNA. The general assertion has been that repetitive sequences make up 5 to 10% of the nuclear DNA in ascomycetes and basidiomycetes, to which rDNA genes make the largest contribution (Wöstemeyer & Kreibich, 2002). The present study shows that while the contribution of rDNA genes to the genome of *M. violaceum* is indeed in this range, additional types of elements, including retrotransposons and helicases, bring the repetitive content of the genome to 15% or more.

While this study provides simply a rough estimate of repetitive content in *M. violaceum*, it is likely to be conservative. First, the figure does not include the many "no-hit" fragments that had significant DNA similarity to other *M. violaceum* sequences in the repetitive categories (retrotransposons or helicases). Second, the level of significance used to initially categorize sequences against the NCBI databases was fairly strict; the Bonferroni correction is generally considered to be overly

conservative (Sokal & Rohlf, 1995; Perneger, 1998). For example, by relaxing the cutoff E-value from 10^{-4} to a value of 1, an additional 45 sequences can be cautiously assigned to the category of retrotransposons, and this would raise the estimated genomic contribution of this category alone from 7.6 to 11.9%. Sequences that are weakly related to retrotransposons may be older copies that have experienced substantial mutational decay. Third, many "no-hit" sequences had matches within the *M. violaceum* database only to several other "no-hit" sequences. This sequence redundancy within the database, even though it represented a small fraction of the total genome and the sequences were from different genomic regions (i.e. fungal sex chromosomes versus autosomes), suggests that some "no-hit" fragments are members of a repetitive category of sequences that has yet to be identified. This possibility will need to be addressed in future studies. Finally, it should be noted that tandem repeat sequences were not included in the estimated percentage of the genome made up by repetitive elements. However, by relaxing the alignment parameters in Tandem Repeats Finder to permit more single base pair mismatches and indels, three times as many sequences were found to contain tandem repeats ($n = 61$, or roughly 10% of the sequences).

Support for the methods and results of the current study come from a previous report where AFLP products linked to the mating type locus of *M. violaceum* were sequenced (Hood, 2002). Four out of 20 AFLP products were retrotransposons most closely related to copia elements in BLASTx searches. One additional AFLP product was a helicase that Poulter et al. (2003) recognized as a *Helitron*-like transposable element. Although a much smaller sample, this is comparable to the frequency of these types of elements on the fungal sex chromosomes in the current study. Because the AFLP techniques used different sets of restriction enzymes to digest the genomic DNA (i.e. *Eco*RI, *Mse*I, *Xho*I, and *Msp*I), the results of the current study are unlikely to reflect a strong restriction site bias of repetitive elements in *M. violaceum* to the *Rsa*I, *Msc*I, and *Dra*I enzymes used to obtain the cloned sequences. Assuming a random sampling of the genome, the probability of repeatedly obtaining the same locus by chance alone is small, roughly equal to the percent coverage of the sex chromosome (3 to 4%) and the autosomes (0.4%),

respectively, for sequences from these sources. Additionally, by excluding rDNA and tubulin genes that have been used for phylogeny reconstruction (Perlin et al., 1997; Freeman et al., 2002), only 56 DNA sequences from *M. violaceum* (averaging 356 bp in length) had been deposited by other researchers into the NCBI databases. Ten of the cloned fragments from the current study matched one or more of these previously deposited sequences, further illustrating the high level of redundancy within the *M. violaceum* genome.

Even though highly selfing organisms are expected to have a low diversity of repetitive transposable elements (Hurst & Werren, 2001), it is clear that *M. violaceum*, an automictic fungus, contains multiple types of Class I retrotransposons and also a large contribution by *Helitron*-type transposable elements. Again, this estimation of repetitive element diversity in *M. violaceum* is probably conservative because of the nature of the random sampling. Typical DNA-based Class II transposons (i.e. *Mariner*, *Fot1*, or *hAT*-type elements) were not identified among the sequence fragments. However, Class II elements often have a lower copy number and make a smaller genomic contribution because of their "cut and paste" replication (Kidwell, 2002). In the current study, where 684 fragments were characterized, and assuming a Poisson distribution, we could be 95% confident of detecting a type of repetitive element representing at least 0.5% of the total genome (Feller, 1957). Roughly twice as many fragments would be needed for equal confidence in detecting an element representing 0.25% of the genome. In fungi examined to date, Class II elements are generally below this threshold (Kidwell, 2002; Mannhaupt et al., 2003), and therefore their presence in *M. violaceum* cannot be excluded without a more detailed study.

Previous studies by Garber and Ruddat (1994, 1998, 2000, 2002) provide evidence of continued activity by transposable elements in *M. violaceum* through mutations in color, morphology, and nutritional requirements. Sequences of retrotransposon RNA products can be obtained by reverse transcriptase (rt) PCR (Hood, unpublished data). Finally, among the fragments characterized in this study, enzymes such as DNA repair proteins, DNA lyase proteins, and topoisomerase proteins (e.g. accessions BZ781946, BZ781953, and BZ781954) are common and may be associated

with repetitive element activity. One remaining sequence of interest was related to suppressor of Gypsy elements in yeasts (BZ782397 had 4e-05 match to the yeast sequence NP_013740). However, it remains possible that some fraction of the transposable element copies are the result of non-autonomous mechanisms of sequence duplication, and the extent to which elements are active in *M. violaceum* is unknown.

An essential question remains: why does *M. violaceum* have so many repetitive elements, especially when compared to the numbers estimated for other organisms with similar genome sizes and low rates of outcrossing? Part of the answer may lie in the fact that large regions of the *M. violaceum* genome appear to be in a condition of fixed heterozygosity and/or suppressed recombination. In *M. violaceum*, the dimorphic fungal sex chromosomes exhibit suppressed recombination throughout much or all of their lengths, and are known to retain a much greater proportion of heterozygosity than the autosomes (Hood & Antonovics, 2004). Repetitive elements are expected to accumulate in such regions because they cannot be excised by ectopic crossing over, they may become fixed via Muller's ratchet, and the deleterious mutations they cause in such regions are likely to remain sheltered by the permanent heterozygosity (Charlesworth & Langley, 1989). Possibly as a result of this process, the density of transposable elements is much higher on the fungal sex chromosomes than on the autosomes of *M. violaceum*, and this is similar to X chromosomes versus autosomes in other organisms where the X shows suppressed recombination (Hood et al., 2004). In a comparison across organisms, the biased distribution of transposable elements on the sex chromosomes was also correlated with a greater overall density on the autosomes. This result suggests that by providing a large and permanent 'reservoir', the dimorphic sex chromosomes increase the genome-wide contribution of repetitive elements (Hood et al., 2004).

Further evidence that the fungal sex chromosomes of *M. violaceum* play a special role in the dynamics of repetitive elements was provided by the greater average number of Standalone BLASTn hits within the *M. violaceum* database for retrotransposon and ''no-hit'' sequences isolated from the fungal sex chromosomes versus the autosomes. The basis for this higher repetitiveness

of sex chromosome elements is not clear, but seems not to be an artifact of sampling. Further analysis of the data did not indicate any correlated difference between sex chromosome and autosome fragments in the strength of BLASTx hits or in the position of the fragments within the full-length retrotransposon (data not shown). However, further studies are needed because the overall content of fungal sex chromosomes has been contrasted with autosome content in very few species (i.e. Hood et al., 2004). It remains to be determined whether the divergence of fungal sex chromosomes from autosomes and thus their impact on genomic composition may be greater in *M. violaceum* than in other model species (e.g. *Saccharomyces cerevisiae*, *Neurospora crassa*, etc.; Wöstemeyer & Kreibich, 2002).

In *M. violaceum* there is also probably a degree of suppressed recombination on the autosomes that could further promote the accumulation of repetitive elements. Garber et al. (1987) reported a deficiency of loci greater than 10 cm from centromeres, and in this study, a DNA sequence (accession BZ782096) was identified with significant similarity to the yeast recombination suppressors *hist4p* and *Sirtuin*. Suppressed recombination on autosomes in *M. violaceum* is likely to be related to its mating system, specifically because automixis with central fusion leads to the preservation of heterozygosity in regions linked with the centromeres (Lewis & John, 1963; Hood & Antonovics, 2004). Genetic load associated with this heterozygosity then favors the spread of recombination suppressors (Antonovics & Abrams, 2004). Others have suggested that the accumulation of repetitive elements is itself a suppressor of recombination through effects mediated by increased methylation (Gorelick, 2003), and it has even been argued that suppressed recombination is a means of preserving genomic stability under pressure from repetitive elements that otherwise might promote repeat-induced ectopic crossing over (Wöstemeyer & Kreibich, 2002).

Retrotransposons are generally viewed as sexually transmitted nuclear parasites that spread into new genomes through outcrossing. As with any vertically transmitted parasite that causes even slightly harmful effects, this transmission between hosts is essential for its maintenance. In reference to the paucity of active transposable elements in the ''ancient asexual'' taxa, it has been suggested

that transposable elements in clonal or strictly inbreeding organisms must either be eliminated or "domesticated" to prevent a rapid extinction of the host lineage (Arkhipova & Meselson, 2000; Nuzhdin & Petrov, 2003). In fungi such as *Candida albicans* and *C. neoformans*, and particularly here in *M. violaceum*, the rarity of outcrossing does not appear to have limited the accumulation of transposable elements. In fact, Morgan (2001) has pointed out that under some situations, the density of transposable elements may be expected to increase for organisms on the cusp between rare outcrossing and absolutely none. In *M. violaceum* from *S. latifolia*, levels of heterozygosity and the maintenance of deleterious recessive alleles suggests that automixis is the primary reproductive mode (Hood & Antonovics, 2004), and that when mating outside the meiotic tetrad occurs, it is usually among tetrads from the same diploid individual. Given that genetic markers may be in regions of fixed heterozygosity, estimates of outcrossing based on allelic distributions have not been possible. However, it is conceivable that spore mixing from different genotypes during the transmission process could lead to rare outcrossing and the spread of repetitive elements. Other lineages of *Microbotryum* (e.g. the lineage found on *Silene caroliniana*) appear to exhibit more strict automixis, and it would be interesting to compare the number and diversity of repetitive elements between the *Microbotryum* lineages with contrasting degrees of outcrossing. This is particularly true as the genomic consequences of automictic reproduction have yet to be thoroughly investigated with models that predict transposable element density. The host-plant may also be a source of novel transposable elements. Interestingly, copia-like elements were the most numerous kind of retrotransposons in the genome of *M. violaceum*, but this is reported to be the rarest kind of Class I element in fungi, especially among basidiomycetes (Kempken & Kück, 1998; Wöstemeyer & Kreibich, 2002; Díez et al., 2003). Furthermore, in *M. violaceum* there appears to be substantial variation among fragments of this kind (see Figure 1b). Copia-like elements have been found in three other basidiomycetes (*Cryptocossus neoformans*, *Laccaria bicolor*, and *Pisolithus microcarpus*). In these cases, the elements were reported to be more similar to transposable elements from plants than from ascomycete fungi (Goodwin &

Poulter, 2001; Díez et al., 2003). For *M. violaceum*, it was also true that the strongest BLASTx hits for the copia-like elements were often sequences originating from plants. All of these particular basidiomycetes have significant ecological relationships with plants, and while the cross-kingdom transfer of infectious genetic elements cannot be excluded, this interpretation requires an appropriate degree of caution (Díez et al., 2003). A much broader survey of fungal genomes is needed to begin resolving these interesting relationships.

This study has provided an informative overall characterization of nuclear content in *M. violaceum*. This was accomplished without the tremendous cost of whole genome sequencing, and illustrates the value of a genome sampling approaches for characterizing nuclear composition. Of course, this has only become a reasonable approach due to the vast international databases of identified genes and proteins and the tools available for comparing sequence similarity. With fungi in particular, where individual chromosomes can be separated by electrophoresis and isolated from other genetic components of the cell (i.e. mitochondria), this approach provides a simple and inexpensive means to comparative genomics.

References

Antonovics, J. & J.Y. Abrams, 2004. Intra-tetrad mating and the evolution of linkage relationships. Evolution.

Antonovics, J., M. Hood & J. Partain, 2002. The ecology and genetics of a host shift: *Microbotryum* as a model system. American Naturalist 160: S40–S53.

Arkhipova, U. & M. Meselson, 2000. Transposable elements in sexual and ancient asexual taxa. Proceeding of the National Academy of Science USA 97: 14473–14477.

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27: 573–580.

Brosius, J., 1999. Genomes were forged by massive bombardments with retroelements and retrosequences. Genetica 107: 209–238.

Casacuberta, J.M. & N. Santiago, 2003. Plant LTR-retrotransposons and MITEs: control of trasposition and

10

impact on the evolution of plant genes and genomes. Gene 311: 1–11.

Charlesworth, B. & C.H. Langley, 1989. The population genetics of *Drosophila* transposable elements. Ann. Rev. Genet. 23: 251–287.

Daboussi, M.J., 1997. Fungal transposable elements and genome evolution. Genetica 100: 253–260.

Díez, J., T. Béguiristain, F. Le Tacon, J.M. Casacuterta & D. Tagu, 2003. Identification of Ty1-copia retrotransposons in three ectomycorrhizal basidiomycetes: evolutionary relationships and use as molecular markers. Curr. Genetics 43: 34–44.

Feller, W., 1957. An Introduction to Probability Theory and Its Applications. Second Edition. John Wiley and Sons, Inc. New York.

Freeman, A.B., K.K. Duong, T.L. Shi, C.F. Hughes & M.H. Perlin, 2002. Isolates of *Microbotryum violaceum* from North American host species are phylogenetically distinct from their European host-derived counterparts. Mol. Phyl. Evol. 23: 158–170.

Garber, E.D. & M. Ruddat, 1994. Genetics of *Ustilago violacea*. XXXII. Genetic evidence for transposable elements. Theor. Appl. Genetics 89: 838–846.

Garber, E.D. & M. Ruddat, 1998. Genetics of *Ustilago violacea*. XXXIV. Genetic evidence for a transposable element functioning during mitosis and two transposable elements functioning during meiosis. Int. J. Plant Sci. 159: 1018–1022.

Garber, E.D. & M. Ruddat, 2000. Genetics of *Ustilago violacea*. XXXV. Transposition in haploid and diploid sporidia and germinating teliospores. Int. J. Plant Sci. 161: 227–231.

Garber, E.D. & M. Ruddat, 2002. Transmission genetics of *Microbotryum violaceum* (*Ustilago violacea*): A case history. Adv. Appl. Microbiol. 51: 107–127.

Goodwin, T.J.D. & T.M. Poulter, 2001. The diversity of retrotransposons in the yeast *Cryptococcus neoformans*. Yeast 18: 865–880.

Gorelick, R., 2003. Transposable elements suppress recombination in all meiotic eukaryotes, including automictic ancient asexuals: a reply to Schön and Martens. J. Nat History 37: 903–909.

Hood, M.E., 2002. Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*. Genetics 160: 457–461.

Hood, M.E. & J. Antonovics, 2004. Mating within the meiotic tetrad and the maintenance of genomic heterozygosity. Genetics 166: 1751–1759.

Hood, M.E., J. Antonovics & H. Heishman, 2003. Karyotypic similarity identifies multiple host-shifts of a pathogenic fungus in natural populations. Infect. Genetics Evol. 2: 167–172.

Hood, M.E., J. Antonovics & B. Koskella, 2004. Shared forces of sex chromosome evolution in haploid-mating and diploid-mating organisms. Genetics.

Hurst, G.D.D. & J.H. Werren, 2001. The role of selfish genetic elements in eukaryotic evolution. Nat. Rev. Genetics 2: 597–606.

Kapitonov, V.V. & J. Jurka, 2001. Rolling-circle transposons in eukaryotes. Proc. Nat. Acad. Sci. USA 98: 8714–8719.

Kempken, F. & U. Kück, 1998. Transposons in filamentous fungi – facts and perspectives. BioEssarys 20: 652–659.

Kidwell, M.G., 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115: 49–63.

Kidwell, M.G. & D.R. Lisch, 2001. Perspective: Transposable elements, parasitic DNA, and genome evolution. Evolution 55: 1–24.

Kumar, S., K. Tamura, I.B. Jakobsen & M. Nei, 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17: 1244–1245.

Lönnig, W.-E. & H. Saedler, 2002. Chromosome rearrangements and transposable elements. Ann. Rev. Genetics 36: 389–410.

Mannhaupt, G. et al., 2003. What's in the genome of a filamentous fungus? Analysis of the Neurospora genome sequence. Nucl. Acids Res. 31: 1944–1954.

Morgan, M.T., 2001. Transposable element number in mixed mating populations. Genetical Res. 77: 261–275.

Nouvel, P., 1994. The mammalian genome shaping activity of reverse-transcriptase. Genetica 93: 191–201.

Nuzhdin, S.V. & D.A. Petrov, 2003. Transposable elements in clonal lineages: lethal hangover from sex. Biol. J. Linnean Soc. 79: 33–41.

Perlin, M.H., C. Hughes, J. Welch, S. Akkaraju, D. Steinecker, A. Kumar, B. Smith, S. Garr, S. Brown & T. Andom, 1997. Molecular approaches to differentiate subpopulations or formae speciales of the fungal phytopathogen *Microbotryum violaceum*. Int. J. Plant Sci. 158: 568–574.

Perneger, T.V., 1998. What's wrong with Bonferroni adjustments. British Med. J. 316: 1236–1238.

Poulter, R.T.M., T.J.D. Goodwin & M.I. Bulter, 2003. Vertebrate helentrons and other novel *Helitron*s. Gene 313: 201–212.

Shapiro, J.A., 2002. Repetitive DNA, genome system architecture and genome reorganization. Res. Microbiol. 153: 447–453.

Shnyreva, A.V., 2003. Transposable elements are the factors involved in various rearrangements and modifications of the fungal genomes. Russian J. Genetics 39: 505–518. Trans. Genetika 39: 621–636.

Sokal, R.R. & F.J. Rohlf, 1995. Biometry: The Principles and Practice of Statistics in Biological Research. W.H. Freeman and Company, New York.

Wöstemeyer, J. & A. Kreibich, 2002. Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. Curr. Genetics 41: 189–198.

Xiong, Y. & T.H. Eickbush, 1988. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol. Biol. Evol. 5: 675–690.