Patterns of Repeat-Induced Point Mutation in Transposable Elements of Basidiomycete Fungi

Felix Horns[1], Elsa Petit[1], Roxana Yockteng[2], and Michael E. Hood[1,*]

[1]Department of Biology, Amherst College

[2]UMR 7205, CNRS-MNHN, Origine, Structure et Evolution de la Biodiversite, Departement Systematique et Evolution, Paris, France

*Corresponding author: E-mail: mhood@amherst.edu.

Accepted: 9 January 2012

Abstract

Transposable elements (TEs) are ubiquitous genomic parasites that have prompted the evolution of genome defense systems that restrict their activity. Repeat-induced point mutation (RIP) is a homology-dependent genome defense that introduces C-to-T transition mutations in duplicated DNA sequences and is thought to control the proliferation of selfish repetitive DNA. Here, we determine the taxonomic distribution of hypermutation patterns indicative of RIP among basidiomycetes. We quantify C-to-T transition mutations in particular di- and trinucleotide target sites for TE-like sequences from nine fungal genomes. We find evidence of RIP-like patterns of hypermutation at TpCpG trinucleotide sites in repetitive sequences from all species of the Pucciniomycotina subphylum of the Basidiomycota, *Microbotryum lychnidis-dioicae*, *Puccinia graminis*, *Melampsora laricis-populina*, and *Rhodotorula graminis*. In contrast, we do not find evidence for RIP-like hypermutation in four species of the Agaricomycotina and Ustilaginomycotina subphyla of the Basidiomycota. Our results suggest that a RIP-like process and the specific nucleotide context for mutations are conserved within the Pucciniomycotina subphylum. These findings imply that coevolutionary interactions between TEs and a hypermutating genome defense are stable over long evolutionary timescales.

Key words: genome defense, coevolution, Pucciniomycotina.

Introduction

Transposable elements (TEs) are DNA sequences that are capable of autonomous replication and insertion at new locations in a genome (Kidwell and Lisch 2001). TEs have been found in every eukaryotic species studied so far and constitute a large fraction of many genomes (Wicker et al. 2007). The remarkable evolutionary success of TEs is attributed to their ability to become overrepresented among an organism's offspring through proliferation within the genome, even though they impose fitness costs (Doolittle and Sapienza 1980; Orgel and Crick 1980; Hickey 1982). Insertion of TEs into new genomic loci can disrupt gene expression (Wright et al. 2003) and promote chromosomal instability through ectopic recombination (Charlesworth et al. 1992; Petrov et al. 2003). The replication of TEs also increases overall genome size and thus imposes an additional cost due to DNA replication. Consequently, organisms have developed a variety of genome defenses that restrict the activity of TEs, such as DNA methylation (Miura et al. 2001), RNA interference (Buchon and Vaury 2005; Cerutti and Casas-Mollano 2006; Chung et al. 2008), and repeat-induced point mutation (RIP; Selker 1990; Selker 2002).

RIP is thought to be the most clear-cut example of a genome defense mechanism because it has no other known purpose (Daboussi and Capy 2003; Galagan and Selker 2004). RIP is a homology-dependent gene silencing process that hypermutates duplicated nuclear DNA. The RIP process detects duplicated DNA sequences and introduces C-to-T mutations in cytosines adjacent to particular nucleotides (i.e., RIP target sites). For example, the RIP process of the ascomycete *Neurospora crassa* causes C-to-T mutations in CA dinucleotides (often written as CpA) that are found within linked duplicated sequences longer than ~400 bp (Watters et al. 1999) and sharing greater than 80% nucleotide identity (Cambareri et al. 1991). Unlinked duplications are also mutated, though at lower frequencies (Watters et al. 1999). The RIP-induced mutations can inactivate TEs and reduce the risk of ectopic recombination by causing

Table 1
Taxonomy of Fungal Species and Sources of Their Genome Sequences

Phylum	Subphylum	Species	Strain	Sequence Assembly Version	Genome Source[a]
Ascomycota	Pezizomycotina	*Neurospora crassa*	OR74A		Broad
Basidiomycota	Puccinomycotina	*Microbotryum lychnidis-dioicae*	P1A1 Lamole		*Microbotryum violaceum* sequencing project, Broad
Basidiomycota	Puccinomycotina	*Puccinia graminis*	CRL 75-36-700-3		*P. graminis* sequencing project, Broad
Basidiomycota	Puccinomycotina	*Melampsora laricis-populina*	98AG31	1.0	JGI
	Pucciniomycotina	*Rhodotorula graminis*	WP1	1.1	JGI
Basidiomycota	Ustilaginomycotina	*Ustilago maydis*	521		*U. maydis* sequencing project, Broad
Basidiomycota	Agaricomycotina	*Tremella mesenterica*	Fries	1.0	JGI
Basidiomycota	Agaricomycotina	*Postia placenta*	Mad-698-R		JGI
Basidiomycota	Agaricomycotina	*Coprinopsis cinerea*	Okayama 7 (#130)		*C. cinerea* sequencing project, Broad

[a] Sources: Broad: sequences produced by the Broad Institute of Harvard and Massachusetts Institute of Technology (http://www.broadinstitute.org/); JGI: sequences produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/).

dissimilar mutations in duplicated sequences. RIP can effectively suppress mobile element activity (Kinsey et al. 1994; Cambareri et al. 1998; Margolin et al. 1998; Selker et al. 2003) but may also constrain the potential for innovation by hampering the evolution of novel gene functions through gene duplication and gradual divergence of biological function (Brookfield 2003; Galagan and Selker 2004).

RIP was first detected in *N. crassa* (Selker et al. 1987). Experimental evidence has subsequently shown that RIP, as an active process, exists in many other ascomycetes, including *Magnaporthe grisea* (Nakayashiki et al. 1999; Ikeda et al. 2002), *Podospora anserina* (Graïa et al. 2001), and *Leptosphaeria maculans* (Idnurm and Howlett 2003). Mutation patterns indicative of RIP, that is, elevated rates of transition mutations of cytosines in the context of particular surrounding nucleotides, have been observed in repetitive elements from *Fusarium oxysporum* (Hua-Van et al. 1998), *Penicillium chrysogenum* (Braumann et al. 2008), *Nectria haematococca* (Coleman et al. 2009), *Grossmannia clavigera* (DiGuistini et al. 2011), and several *Neurospora* (Kinsey et al. 1994) and *Aspergillus* species (Neuveglise et al. 1996; Nielsen et al. 2001; Montiel et al. 2006; Braumann et al. 2008; Clutterbuck 2011). Though the RIP process causes C-to-T transitions in each of these species, there is substantial variation in the preferred adjacent nucleotide context for RIP mutations (reviewed in Galagan and Selker 2004; Clutterbuck 2011), which may be a consequence of coevolutionary dynamics between the RIP defense and the TEs.

Among basidiomycetes, evidence of RIP-like hypermutations has been described in the anther-smut fungus *Microbotryum lychnidis-dioicae* (referred to as *M. violaceum* isolated from *Silene latifolia* in Hood et al. 2005), in which the target site appears to be strictly the trinucleotide combination TpCpG. To our knowledge, *Microbotryum* remains the only basidiomycete genus in which such hypermutation patterns in repetitive elements have been found. The lack of knowledge about the taxonomic distribution of RIP and variation in its preferred mutation target sites across the fungal kingdom limits our ability to understand the long-term coevolutionary interactions between RIP and TEs that have a major role in shaping genome evolution (Brookfield 2003; Galagan and Selker 2004).

In this study, we characterized the mutational patterns of nine fungal genome sequences to determine whether the signatures of RIP activity are present. We analyzed alignments of repetitive TE-like sequences from each species for evidence of C-to-T mutations and biases in adjacent nucleotide context. Using this approach, we characterized the taxonomic distribution of variation in the RIP-like processes among basidiomycetes.

Materials and Methods

Sources of DNA sequences corresponding to raw assemblies of basidiomycete genomes are listed in table 1. A database of fungal TE sequences was downloaded from REPBASE (Jurka 1998, 2000; Jurka et al. 2005). TE-like sequences were retrieved from each fungal genome using entries from the fungal TE sequences database as the queries in the translated sequence similarity search program tBLASTx (a Basic Local Alignment Search Tool implemented using "standalone BLAST"; ftp://ftp.ncbi.nlm.nih.gov/blast/executables/). The most significant hits (maximum *E* value of 10^{-6}) for each family of TE in the TE database (e.g., *Copia*-like, *Gypsy*-like, etc.) were recovered from the fungal genome sequences.

Sets of highly similar TE-like repeat sequences were obtained from within each fungal genome by conducting a nucleotide sequence similarity search against the genome

assembly with the BLASTn program of the NCBI, using as search queries the TE-like sequences previously recovered from those genomes. BLASTn retrieves sequences that are similar at the DNA level and therefore are similar enough to be detected by a RIP process. Genome sequence fragments longer than 1,000 bp and with greater than 98% similarity to one another were aligned using T-COFFEE software in Expresso mode (Notredame et al. 2000).

The genomic hallmark of RIP is the presence of C-to-T transition mutations in repeated sequences. Therefore, transition mutation frequencies were quantified for each alignment of TE-like sequences relative to the consensus sequence of the alignment, which served as a model of the intact progenitor element. The effects of 5′ and 3′ flanking bases on cytosine transition mutation frequencies were assessed by computing the frequency of C-to-T mutations in every dinucleotide sequence context containing a C or G residue (e.g., CpA, TpC) and every trinucleotide sequence context containing a C or G residue in the second position (e.g., TpCpG; as in Hood et al. 2005). Mutations on both strands were counted on the assumption that G-to-A transitions on the sequenced strand represent C-to-T transitions on the complementary strand. G-to-A changes were therefore analyzed as C-to-T transitions in the context of the reverse complement of the surrounding nucleotides. For each di- or trinucleotide at a given position in the consensus sequence of an alignment, the mutation frequency was computed as the number of sequence fragments with a transition mutation at that position divided by the total number of aligned sequences. Nucleotide combinations contained within or spanning the boundaries of insertions that were present in only one of the aligned sequences were excluded from the analysis.

The intensity of mutation of a cytosine residue in a particular adjacent nucleotide context was assessed by computing the proportion of those particular di- or trinucleotide combinations in the consensus sequence that exceeded a mutation frequency threshold based upon the overall mutation rate μ for the alignment. Overall mutation rate μ was calculated for each alignment as the total number of mutations relative to the consensus sequence divided by the total number of nucleotides in the alignment. The mutation frequency threshold is x/n, where n is the number of sequences in the alignment and x is the smallest integer such that x/n is greater than or equal to μ. A nucleotide site with mutation frequency greater than the threshold therefore suffers more mutation than expected based upon the overall mutation rate; this method allows one to distinguish RIP-like hypermutation patterns from random mutation.

To determine which particular di- or trinucleotide contexts had excessive mutation rates, the proportions of particular nucleotide combinations that exceeded the mutation threshold were subjected to a univariate outlier analysis using $|z\text{-score}| > 2.5$ as the criterion to identify outliers

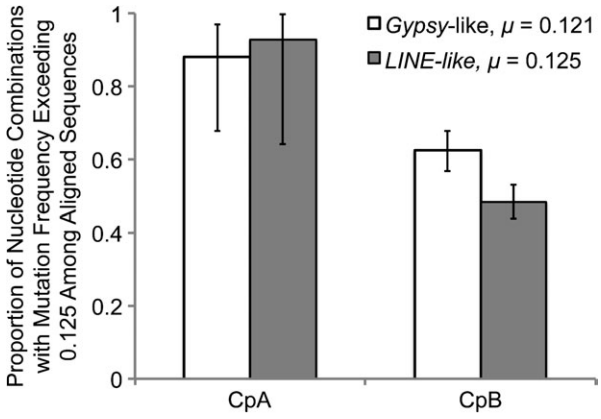

Fig. 1.—Proportion of various dinucleotide sequences with C-to-T mutations in two alignments of TE-like repeat sequences from the ascomycete fungus *Neurospora crassa*. Numbers and lengths of aligned sequences are given in table 2. Overall mutation rate μ of each alignment was calculated as the total number of mutations relative to the consensus sequence divided by the total number of base pairs in the alignment. Mutation rates of cytosine residues in particular sequence contexts are shown as the proportion of dinucleotide combinations in the consensus sequence with C-to-T mutation frequency among sequences in the alignment exceeding the threshold defined by the mutation rate μ for that particular alignment, as described in the text. Standard International Union of Biochemistry (IUB) codes are used for incompletely specified nucleotides: B indicates nucleotides other than A. Bars represent the 95% confidence intervals associated with the proportion of nucleotide combinations with mutation frequency among sequences exceeding the mutation threshold.

(Rousseeuw and Hubert 2011); di- or trinucleotide contexts were analyzed separately, and the proportional data were arcsin square root transformed prior to analysis. Nucleotide contexts in which C-to-T transitions occurred with excessive frequency were then compared with the combined data for alternative nucleotide contexts for graphical representation.

These methods were used to analyze the mutational patterns of TE-like repeat sequences from the genomes of fungal species from each major subphylum within the basidiomycete fungi (table 1).

Results

To establish that our approach is capable of detecting RIP-like mutational patterns, the method was first applied to the genome sequence of *N. crassa*, which is known to have a RIP process that targets the dinucleotide combination CpA (Selker 1990). Our search revealed a RIP-like pattern of hypermutation at CpA dinucleotide sites in two alignments of TE-like sequences, thereby serving as a positive control for the informatics and computational procedures (fig 1; supplementary fig. 1, Supplementary Material online). Furthermore, the consensus of each alignment had significantly fewer CpA dinucleotides than expected from a random distribution of bases (P value $< 10^{-9}$ estimated based on

a binomial distribution), which is consistent with the long-term activity of RIP removing CpA sites in this species (Hane and Oliver 2008).

Further validation of our approach was obtained by analysis of the genome sequences of *M. lychnidis-dioicae*, which is known to have hypermutations in TpCpG trinucleotides (Hood et al. 2005). Hypermutation was detected in four alignments of TE-like sequences from *M. lychnidis-dioicae* with the TpCpG trinucleotide identified as an outlier for having excessive C-to-T transitions (table 2). In contrast, partial matches to TpCpG sequences that differed immediately 3′ or 5′ to the cytosine residue did not show elevated C-to-T mutation frequencies in comparison with other cytosine residues (fig. 2; supplementary fig. 2, Supplementary Material online).

To characterize systematic variation in hypermutation patterns among basidiomycetes, we analyzed the genome sequences of eight species from the subphyla Pucciniomycotina, Ustilaginomycotina, and Agaricomycotina. From the Pucciniomycotina subphylum that includes *M. lychnidis-dioicae*, three additional species, *Puccinia graminis*, *Melampsora laricis-populina*, and *Rhodotorula graminis*, showed evidence of hypermutation patterns at the trinucleotide combination TpCpG (table 2; fig. 2; supplementary fig. 2, Supplementary Material online). Evidence for the TpCpG hypermutation site was provided when multiple alignments of TE-like sequences were available from the genomes and exceeded the *z*-score criterion, with only two of four alignments from *R. graminis* failing to exceed the *z*-score criterion (table 2).

In contrast, repeat sequences from *Ustilago maydis*, in the subphylum Ustilaginomycotina, did not provide evidence of hypermutation at particular nucleotide combinations (supplementary fig. 3, Supplementary Material online). Two other available ustilaginomycetes, *Sporosorium reilianum* and *Malassezia globosa*, yielded poor alignments of TE-like repeated sequences and could not be included in the analysis. TE-like repeat sequences from three species in the Agaricomycotina, *Tremella mesenterica*, *Postia placenta*, and *Coprinopsis cinerea*, also provided no clear evidence of RIP-like hypermutation (table 2; supplementary figs 4–6, Supplementary Material online). One alignment from *P. placenta* showed a *z*-score approaching the outlier criterion, also for the TpCpG trinucleotide.

Discussion

Hypermutation patterns indicative of a RIP-like genome defense are present in repetitive DNA throughout the subphylum Pucciniomycotina. Furthermore, the identified trinucleotide target sequence, TpCpG, as previously shown in *M. lychnidis-dioicae* (Hood et al. 2005), appears to be highly conserved within this group (fig. 3) We did not detect evidence for RIP-like context-specific hypermutation patterns in other Basidiomycete subphyla.

The present study provides the first evidence that a RIP-like pattern of hypermutation occurs in diverse lineages within the Pucciniomycotina subphylum, including the important model organisms *P. graminis*, *M. laricis-populina*, and *R. graminis*. We have shown that TE-like repeated elements from these species suffer an elevated rate of mutation at cytosine residues in the context of TpCpG trinucleotides relative to those in other nucleotide contexts. This observation does not simply reflect a difference in the frequency of particular di- or trinucleotides (i.e., the possibility that TpCpG is more common than VpCpG and therefore there is a greater absolute number of TpCpG mutations) because our approach considers the rate of mutation per occurrence of the di- or trinucleotide combination. Moreover, the observed mutation patterns are not accounted for by known patterns of DNA methylation in fungi. CpG dinucleotides are a target of methylation as part of epigenetic controls over chromosome condensation or gene expression (Jaenisch and Bird 2003). Passive deamination of methylcytosine is therefore expected to cause a higher frequency of transition mutations at CpG dinucleotides than other dinucleotides (Cooper and Krawczak 1990). Evidence for this effect in fungi can be seen in the higher overall transition mutation rates of cytosines followed by guanines (e.g., higher mutation rates of VpCpG over VpCpH trinucleotides in fig. 2). However, there is no evidence that the targeting of regulatory DNA methylation is influenced by the nucleotide that is 5′ to the CpG dinucleotide (e.g., preferential methylation of TpCpG trinucleotides relative to VpCpG trinucleotides).

Thus, our findings suggest that fungi in the Pucciniomycotina exhibit RIP-like hypermutation of specific trinucleotides in repeated sequences or that there exists a characteristic of regulatory cytosine methylation in these fungi that is previously unseen among Eukaryotes. We note that this study has not determined whether excess mutation at TpCpG trinucleotides is present only in repetitive DNA sequences. However, a prior study on *M. lychnidis-dioicae* provided evidence that the distribution of TpCpG sites in repetitive sequences contrasts that in single copy genes, consistent with the long-term influence of hypermutation in repetitive elements (Hood et al. 2005). Our methodology could be applied to the analysis of sequences of single copy genes from multiple strains or from sister species of each fungus to confirm that RIP-like patterns of hypermutation occur only in repetitive DNA. Further studies involving the experimental generation of multiple-copy genomic sequences are required to confirm the presence of an active RIP-like process targeting only repetitive DNA and its target site among the Pucciniomycotina.

In contrast to the Pucciniomycotina, there was not strong evidence of RIP-like patterns of hypermutation in repetitive sequences from species of the subphyla Agaricomycotina and Ustilaginomycotina. Although we observed an overrepresentation of transition mutations in CpG dinucleotides, this pattern can be explained by the spontaneous

Table 2

Lengths and Numbers of Sequences in the Alignments of TE-Like Sequences from Fungal Genomes, Classification of TEs, Mutational Patterns Observed in This Study, and Statistical Support for Those Patterns

Phylum	Subphylum	Species	TE Classification[a]	Number of Aligned Sequences	Length of Consensus Sequence (base pairs)	Overall Mutation Rate among Aligned Sequences (mutations per base pair)[b]	Nucleotide Context with Maximum Hypermutation Pattern[c]	Maximum *z*-Score[d]
Ascomycota	Pezizomycotina	*Neurospora crassa*	*Gypsy*-like	24	4,346	0.121	[C]pA	1.67
			LINE-like	18	1,887	0.125	[C]pA	1.92
Basidiomycota	Pucciniomycotina	*Microbotryum lychnidis-dioicae*	*Copia*-like	9	4,438	0.025	**Tp[C]pG**	**3.02**
			Copia-like	8	3,029	0.024	**Tp[C]pG**	**3.03**
			Copia-like	8	2,214	0.015	**Tp[C]pG**	**3.14**
			Gypsy-like	8	1,458	0.017	**Tp[C]pG**	**2.54**
		Puccinia graminis	*Gypsy*-like	10	5,282	0.021	**Tp[C]pG**	**2.70**
		Melampsora laricis-populina	*Gypsy*-like	21	6,001	0.002	**Tp[C]pG**	**2.88**
			Gypsy-like	8	3,998	0.002	**Tp[C]pG**	**2.89**
		Rhodotorula graminis	*Gypsy*-like	65	1,761	0.032	Tp[C]pG	2.34
			Gypsy-like	34	1,506	0.030	**Tp[C]pG**	**2.58**
			Gypsy-like	15	1,482	0.029	**Tp[C]pG**	**3.02**
			LINE-like	10	4,334	0.036	Tp/Cp[C]pG	2.00
	Ustilaginomycotina	*Ustilago maydis*	*Copia*-like	12	5,923	0.003	Cp[C]pG	1.31
			Copia-like	12	5,230	0.003	Cp[C]pC	1.34
			Copia-like	10	5,617	0.002	Tp[C]pG	1.56
	Agaricomycotina	*Tremella mesenterica*	*Gypsy*-like	58	2,903	0.005	Gp[C]pC	1.97
			Gypsy-like	36	5,078	0.006	Cp[C]pC	1.20
			Copia-like	20	1,480	0.007	Gp[C]pG	1.41
		Postia placenta	*Gypsy*-like	15	2,230	0.022	Tp[C]pG	2.38
			Gypsy-like	13	2,158	0.020	Ap[C]pG	1.92
		Coprinopsis cinerea	*Gypsy*-like	14	3,147	0.046	Cp[C]pG	1.74
			Copia-like	14	2,305	0.029	Gp[C]pG	1.86
			Gypsy-like	11	6,001	0.011	Tp[C]pG	1.69
			Copia-like	10	2,620	0.011	Tp[C]pG	1.93
			Copia-like	10	2,371	0.010	Cp[C]pG	1.85
			Gypsy-like	7	5,875	0.012	Tp[C]pG	1.69

[a] The TE order or superfamily, as described in Wicker et al. (2007), of each set of TE-like sequences determined using tBLASTx of the REPBASE database of fungal TE sequences. In some genomes, TE from multiple alignments belonged to the same order or superfamily of TE but could not be combined because they were too divergent at the DNA level to be aligned.

[b] Overall mutation rate was calculated as the total number of mutations relative to the consensus sequence divided by the total number of base pairs in the alignment.

[c] Within each nucleotide context shown, the cytosine nucleotide mutated is indicated within brackets.

[d] *z*-scores shown in bold text identify outliers with respect to C-to-T mutation frequency for the nucleotide context in the previous column by univariate outlier analysis using $|z\text{-score}| > 2.5$ as the criterion.

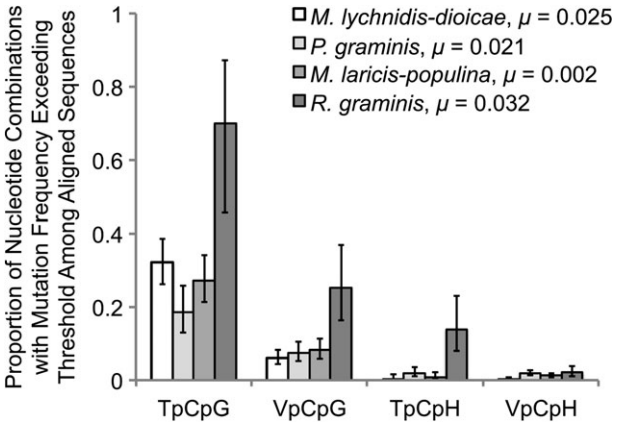

Fig. 2.—Proportion of various trinucleotide sequences with C-to-T mutations in representative alignments of TE-like repeat sequences from species of the fungal basidiomycete subphylum Pucciniomycotina. Alignment of TE-like sequences from *Microbotryum lychnidis-dioicae* is composed of 9 sequences of a *Copia*-like element; *Puccinia graminis*, 10 sequences of a *Gypsy*-like element; *Melampsora laricis-populina*, 21 sequences of a *Gypsy*-like element; *Rhodotorula graminis*, 65 sequences of a *Gypsy*-like element; analysis of other alignments is presented in table 2. Overall mutation rate μ of each alignment was calculated as the total number of mutations relative to the consensus sequence divided by the total number of base pairs in the alignment. Mutation rates of cytosine residues in particular sequence contexts are shown as the proportion of trinucleotide combinations in the consensus sequence with C-to-T mutation frequency among sequences in the alignment exceeding the threshold defined by the mutation rate μ for that particular alignment, as described in the text. Standard IUB codes are used for incompletely specified nucleotides: V indicates nucleotides other than T; H indicates nucleotides other than G. Bars represent the 95% confidence intervals associated with the proportion of nucleotide combinations with mutation frequency among sequences exceeding the mutation threshold.

deamination of methylated cytosines, as has been observed in many genes (Krawczak et al. 1998; Zhao and Boerwinkle 2002; Jiang and Zhao 2006). There was no evidence of significantly elevated mutation rates at cytosine residues in other dinucleotide contexts or in specific trinucleotide contexts. Although these observations are consistent with the absence of targeted hypermutation, we cannot exclude the possibility that a RIP-like process exists in these species but was not detected possibly due to the limited number and quality of genome assemblies available or due to statistical limitations of the analysis of alignments that have low overall mutation rate. In particular, regarding the Ustilaginomycotina, we note that only one species was analyzed and that TE sequences from that species had low overall mutation rate, which limits the statistical power to detect mutation patterns. Other available DNA sequence resources from this subphylum were of poor quality or did not yield enough repetitive elements to construct usable alignments. Also, the TpCpG trinucleotide in *P. placenta* showed a *z*-score approaching the identification criterion. As additional and higher coverage genome assemblies become available, we expect that further analysis using computational approaches will reveal whether RIP-like hypermutation patterns exist in the Ustilaginomycotina and Agaricomycotina.

The results of this study suggest that the coevolutionary dynamics between TEs and a process of targeted hypermutation are long-term interactions. There is evidence that TE lineages in *M. lychnidis-dioicae* have a general paucity of TpCpG trinucleotides in positions where C-to-T transition would cause a nonsynonymous substitution, thereby possibly imposing selection on the host genome to modify the RIP-like process to target a different nucleotide combination (Hood et al. 2005). Conservation of the trinucleotide target site among members of the subphylum Pucciniomycotina implies that counteradaptation by TE lineages has not imposed sufficient selective pressure to modify the target site specificity. However, among ascomycete species with RIP, there is considerable target site variation (Galagan and Selker 2004; Clutterbuck 2011). A RIP-like process may operate in a completely different manner in basidiomycetes than ascomycetes. The *rid* gene in *N. crassa* is the putative DNA methyltransferase that is necessary for RIP (Freitag et al. 2002). Notably, sequences with strong similarity to *rid* can be found in other ascomycetes, but we could not find similar cytosine methyltransferases in the Pucciniomycotina, suggesting that the *rid* homologue may have diverged during the substantial evolutionary time separating ascomycetes and basidiomycetes or that different enzymes or processes are responsible for the hypermutation patterns observed in the Pucciniomycotina.

We have introduced a bioinformatic approach for the detection of RIP-like patterns of hypermutation in genome sequences (see also the alignment-based tool applied to ascomycetes by Hane and Oliver 2008). Our framework is modular in nature and can be readily integrated with other methods for the detection and alignment of TEs in genomic sequences. However, our method suffers several limitations. First, mutational profiles have lower statistical support in genomes that have low overall mutation rate. In the present study, for example, the mutational profile of *U. maydis* has low statistical support because TE-like repeat sequences from this genome have few mutations overall. We note that this may also limit detection of RIP in TE sequences that have recently proliferated and therefore have not sustained many mutations due to RIP and other processes. Second, the depletion of RIP target sites from TE sequences can reduce the level of statistical support for observed mutational patterns. In this study, two alignments of repeated elements from *R. graminis* had excess mutation at cytosines in the context of TpCpG, but this pattern was not well-supported statistically because of the paucity of those trinucleotides (table 2). Third, a genome in which RIP is strongly effective would display little detectable evidence of RIP. Such a genome would have few highly degraded TEs that are difficult to detect

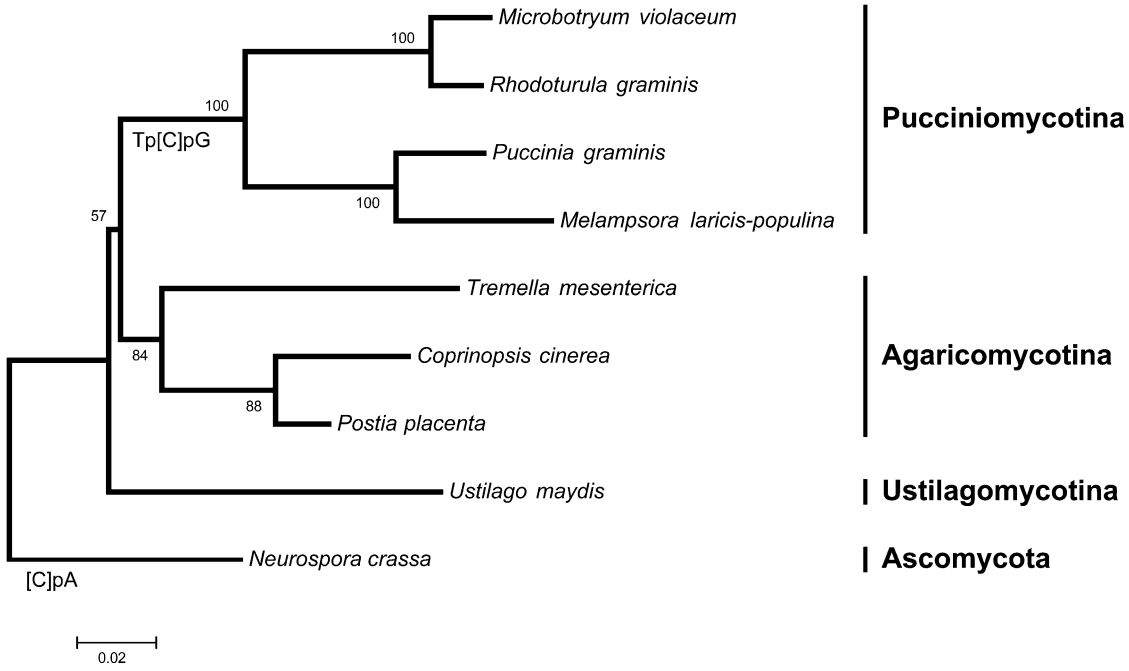

Fig. 3.—Molecular phylogenetic analysis of the taxa used in the analysis based on the partial sequence of the 18S ribosomal RNA gene. The evolutionary history was inferred by using the maximum likelihood method based on the Kimura 2-parameter model (Kimura 1980). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches (Felsenstein 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved nine nucleotide sequences. There were a total of 1367 positions in the final data set. Evolutionary analyses were conducted in MEGA5 (Tamura et al. 2011). Hypermutation patterns detected in this study are indicated below the branch at the base of the clade in which the pattern is found in all taxa examined.

and align. Furthermore, alignments of TE sequences that have been completely mutated at RIP target sites would show no evidence of RIP activity because comparison of aligned sequences would reveal no variation from the RIP-mutated consensus sequence. In such species, RIP activity must be inferred from the dearth of intact RIP target sites or overrepresentation of the nucleotide combinations generated by RIP, if they are known. For example, the depletion of CpA dinucleotides in TE-like sequences from *N. crassa* observed in the present study and previously (Hane and Oliver 2008) likely reflects the activity of a RIP process that mutates those dinucleotides.

In conclusion, we have presented evidence of RIP-like patterns of hypermutation at TpCpG trinucleotides in TE-like repeat sequences from *P. graminis*, *M. laricis-populina*, and *R. graminis* and confirmed their presence in *M. lychnidis-dioicae*. The process and target site for hypermutation appear to be highly conserved within the subphylum Pucciniomycotina, but these patterns are not currently detectable among other Basidiomycetes. Further investigation into TE dynamics and genome defenses in the Pucciniomycotina should be pursued, particularly through the creation of repeated sequences by transformation as a means to investigate the hypermutation process, which has contributed greatly to our understanding of RIP in ascomycete model species (Selker 2002).

Supplementary Material

Supplementary figures 1–6 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was supported under the award National Science Foundation DEB-0747222 to M.E.H., and the authors are grateful to Tatiana Giraud for helpful comments.

Literature Cited

Braumann I, van den Berg M, Kempken F. 2008. Repeat induced point mutation in two asexual fungi, *Aspergillus niger* and *Penicillium chrysogenum*. Curr Genet. 53:287–297.

Brookfield JFY. 2003. Genome sequencing: the ripping yarn of the frozen genome. Curr Biol. 13:R552–R553.

Buchon N, Vaury C. 2005. RNAi: a defensive RNA-silencing against viruses and transposable elements. Heredity 96:195–202.

Cambareri EB, Aisner R, Carbon J. 1998. Structure of the chromosome VII centromere region in *Neurospora crassa*: degenerate transposons and simple repeats. Mol Cell Biol. 18:5465–5477.

Cambareri EB, Singer MJ, Selker EU. 1991. Recurrence of repeat-induced point mutation (Rip) in *Neurospora Crassa*. Genetics 127:699–710.

Cerutti H, Casas-Mollano J. 2006. On the origin and functions of RNA-mediated silencing: from protists to man. Curr Genet. 50:81–99.

Charlesworth B, Lapid A, Canada D. 1992. The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and distribution. Genet Res. 60:103–114.
Chung W-J, Okamura K, Martin R, Lai EC. 2008. Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. Curr Biol. 18:795–802.
Clutterbuck JA. 2011. Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. Fungal Genet Biol. 48:306–326.
Coleman JJ, et al. 2009. The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. PLoS Genet. 5:e1000618.
Cooper DN, Krawczak M. 1990. The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. Hum Genet. 85:55–74.
Daboussi M-J, Capy P. 2003. Transposable elements in filamentous fungi. Annu Rev Microbiol. 57:275–299.
DiGuistini S, et al. 2011. Genome and transcriptome analyses of the mountain pine beetle-fungal symbiont *Grosmannia clavigera*, a lodgepole pine pathogen. Proc Natl Acad Sci U S A. 108:2504–2509.
Doolittle WF, Sapienza C. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603.
Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.
Freitag M, Williams RL, Kothe GO, Selker EU. 2002. A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. Proc Natl Acad Sci U S A. 99:8802–8807.
Galagan JE, Selker EU. 2004. RIP: the evolutionary cost of genome defense. Trends Genet. 20:417–423.
Graïa F, et al. 2001. Genome quality control: RIP (repeat-induced point mutation) comes to *Podospora*. Mol Microbiol. 40:586–595.
Hane J, Oliver R. 2008. RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. BMC Bioinformatics 9:478.
Hickey DA. 1982. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics 101:519–531.
Hood ME, Katawczik M, Giraud T. 2005. Repeat-induced point mutation and the population structure of transposable elements in *Microbotryum violaceum*. Genetics 170:1081–1089.
Hua-Van A, Héricourt F, Capy P, Daboussi M, Langin T. 1998. Three highly divergent subfamilies of the *impala* transposable element coexist in the genome of the fungus *Fusarium oxysporum*. Mol Gen Genet. 259:354–362.
Idnurm A, Howlett BJ. 2003. Analysis of loss of pathogenicity mutants reveals that repeat-induced point mutations can occur in the Dothideomycete *Leptosphaeria maculans*. Fungal Genet Biol. 39:31–37.
Ikeda K, et al. 2002. Repeat-induced point mutation (RIP) in *Magnaporthe grisea*: implications for its sexual cycle in the natural field context. Mol Microbiol. 45:1355–1364.
Jaenisch R, Bird A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 33:245–254.
Jiang C, Zhao Z. 2006. Directionality of point mutation and 5-methylcytosine deamination rates in the chimpanzee genome. BMC Genomics 7:316.
Jurka J. 1998. Repeats in genomic DNA: mining and meaning. Curr Opin Struct Biol. 8:333–337.
Jurka J. 2000. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. 16:418–420.
Jurka J, et al. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Gen Res. 110:462–467.
Kidwell MG, Lisch DR. 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution 55:1–24.
Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 16:111–120.
Kinsey JA, Garrett-Engele PW, Cambareri EB, Selker EU. 1994. The *Neurospora* transposon *Tad* is sensitive to repeat-induced point mutation (Rip). Genetics 138:657–664.
Krawczak M, Ball EV, Cooper DN. 1998. Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. Am J Hum Genet. 63:474–488.
Miura A, et al. 2001. Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. Nature 411:212–214.
Montiel MD, Lee HA, Archer DB. 2006. Evidence of RIP (repeat-induced point mutation) in transposase sequences of *Aspergillus oryzae*. Fungal Genet Biol. 43:439–445.
Nakayashiki H, Nishimoto N, Ikeda K, Tosa Y, Mayama S. 1999. Degenerate *MAGGY* elements in a subgroup of *Pyricularia grisea*: a possible example of successful capture of a genetic invader by a fungal genome. Mol Gen Genet. 261:958–966.
Neuveglise C, Sarfati J, Latge JP, Paris S. 1996. *Afut1*, a retrotransposon-like element from *Aspergillus fumigatus*. Nucl Acids Res. 24:1428–1434.
Nielsen ML, Hermansen TD, Aleksenko A. 2001. A family of DNA repeats in *Aspergillus nidulans* has assimilated degenerated retrotransposons. Mol Genet Genomics. 265:883–887.
Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol. 302:205–217.
Orgel LE, Crick FH. 1980. Selfish DNA: the ultimate parasite. Nature 284:604–607.
Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. 2003. Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. Mol Biol Evol. 20:880–892.
Rousseeuw PJ, Hubert M. 2011. Robust statistics for outlier detection. WIREs Data Mining Knowl Discov. 1:73–79.
Selker EU. 1990. Premeiotic instability of repeated sequences in *Neurospora crassa*. Annu Rev Genet. 24:579–613.
Selker EU. 2002. Repeat-induced gene silencing in fungi. Adv Genet. 46:439–450.
Selker EU, Cambareri EB, Jensen BC, Haack KR. 1987. Rearrangement of duplicated DNA in specialized cells of *Neurospora*. Cell 51:741–752.
Selker EU, et al. 2003. The methylated component of the *Neurospora crassa* genome. Nature 422:893–897.
Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–2739.
Watters MK, Randall TA, Margolin BS, Selker EU, Stadler DR. 1999. Action of repeat-induced point mutation on both strands of a duplex and on tandem duplications of various sizes in *Neurospora*. Genetics 153:705–714.
Wicker T, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 8:973–982.
Wright SI, Agrawal N, Bureau TE. 2003. Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. Genome Res. 13:1897–1903.
Zhao Z, Boerwinkle E. 2002. Neighboring-nucleotide effects on single nucleotide polymorphisms: a study of 2.6 million polymorphisms across the human genome. Genome Res. 12:1679–1686.

Associate editor: Esther Betran