Closely-Related Smut Species vary Widely in TE Content

We recently reported on the generation and analysis of the genome sequence of the barley-infecting basidiomycete smut fungus, *Ustilago hordei*.1 In comparing the genome of this fungus to those of two closely-related fungi, the corn-infecting smuts *U. maydis*2 and *Sporisorium reilianum*,3 we revealed several striking differences. The genome of *U. hordei* contained 5- to 12-times more transposable elements (TEs) and repeats than those of *U. maydis* and *S. reilianum*, respectively. LTR (long-terminal repeat) retrotransposon and LINE (long-interspersed element) sequences were responsible for the bulk of the difference. We also found in the *U. hordei* genome clear examples of RIP (repeat-induced point)-like mutations, a genome defense mechanism originally described for the ascomycete fungus *Neurospora crassa*4 and later in basidiomycetes, including *Microbotryum violaceum*,5,6 which introduces mutations in repeats (often TEs) to inactivate them. Coincidentally, mating behavior differs between these smuts. Earlier work on smut fungi helped define genes involved in mating and revealed two loci as crucial determinants: the homeodomain-containing, transcription factor-coding *b*-gene complex locus and the pheromone and pheromone receptor-coding *a*-gene complex locus (reviewed in 7). These gene complexes reside on separate chromosomes in *U. maydis*2 and *S. reilianum*8 and therefore genetically constitute what has been called a tetrapolar mating system: two genetic loci, one, the *a* locus, with two specificities (for *U. maydis*), and the *b* locus, having at least 30 specificities in nature (for *U. maydis*). *U. hordei*, in contrast, possesses a bipolar mating mechanism represented by one mating-type locus with only two specificities in nature (*MAT-1* and *MAT-2*). Years ago, we showed that this locus had similar mating complexes as found in *U. maydis* but that in *MAT-1*, these were physically linked on the largest chromosome, 527 kb apart. This large chromosomal region, suppressed in recombination with its estimated 430 kb *MAT-2* counterpart and having over 50% of its DNA sequence coding for TEs and repeats, resembled a nascent sex chromosome.9-11 In the recent comparative genome paper, we additionally concluded that a single chromosome fusion event led to the bipolar condition in *U. hordei*.1 There was another intriguing finding: comparison to the *U. hordei* genome showed that *U. maydis* lacked several genes

*Correspondence to: Guus Bakkeren; Email: guus.bakkeren@agr.gc.ca
Submitted: 12/17/12; Revised: 01/31/13; Accepted: 02/01/13
http://dx.doi.org/10.4161/psb.23853
Citation: Laurie JD, Linning R, Wong P, Bakkeren G. Do TE activity and counteracting genome defenses, RNAi and methylation, shape the sex lives of smut fungi? Plant Signal Behav 2013; 8: e23853
involved in genome defense such as components of the RNAi-silencing pathway (dicer, argonaught and three RdRps) and DNA methylation (chromodomain-containing proteins and a cytosine 5-specific methyltransferase). Even more astounding was the fact that these genes seemed to be cleanly deleted from the *U. maydis* genome, leaving the flanking genes in conserved synteny compared with the other two genomes. Remnants of TEs and small footprints reminiscent of TE activity were found immediately flanking these deletion sites. The emerging data tempted us to speculate further about possible correlations between TEs and repeats, silencing and genome defense, and their possible effect on mating type evolution.

**TEs as Modifiers of Genomic Information**

TEs are classified into two broad categories depending on their mode of replication. Type I TEs or retro-elements replicate via an RNA intermediate and require a reverse transcriptase to form a DNA replicate. Type II TEs, however, require no RNA intermediate and replicate directly from DNA. Hypotheses have been put forward that TEs (retroelements and noncoding RNAs) are former viral agents derived from an ancient RNA world that have become neutralized (balanced after “endogenization”) in a “persistent symbiotic lifestyle” in the genome and are “natural genome editors” and “agents of (beneficial) change” by playing important roles in the regulation of (all) cellular processes in the host organism (reviewed in 13-16). Most eukaryotes have genomes with a significant proportion of TEs that consequently have hampered assembly efforts following whole genome sequencing. TEs, their remnants and other repetitive DNA sequences are often found in a special form of chromatin called heterochromatin that is repressed in both transcription and recombination. Additionally, heterochromatin is replicated separately during the cell cycle from the more active euchromatin, where the majority of genes are housed. Having large amounts of repetitive DNA that is actively processed into discrete chromatin structures and replicates separately, likely comes at a cost for cells. Indeed, TEs are traditionally known for having negative impacts on the genomes of their hosts. These include in addition to complications in chromosome replication and energy costs, gene deletions or inactivation through TE insertions and ectopic recombination leading to gene loss or more severely to chromosome instability. On the other hand, and more in tune with modern views emanating from the genomics era, TEs are more-and-more seen as “genetic settlers” that bring about genetic and genomic innovations such as altered gene expression through their influence on neighboring genes, post-transcriptional RNA processing, translation control, epigenetic control, etc., that are beneficial to organisms by providing adaptations on which selective forces can act. A non-exhaustive list of 30 documented examples by which a particular retroelement in humans can exert functional activity is given by von Sternberg and Shapiro. Adaptations can be on a relatively short time scale such as in pathogens having to overcome host defense mechanisms as an immediate selection pressure (see below). It has been argued that under such specific conditions and for organisms with large genomes the cost of housing potential detrimental TEs is outweighed by their beneficial contributions.

**Control of TE Function and Activities**

Unbridled proliferation of TEs in an organism though, is detrimental and is necessarily balanced by host immune mechanisms evolved to defend the genome and remove, inactivate or control competing TEs (and their modulating functions). Mating systems have been described affecting the spread and persistence of TE load in genomes. One of the main factors believed to have favored sexual over asexual reproduction during evolution besides the generation of reshuffled alleles on which natural selection can act, is the removal of detrimental DNA, i.e., deleterious mutations including TEs, from a population (Fisher-Müller hypothesis). At the same time, sex can accelerate dispersal of TEs in a population. Means to modulate activities (inactivate) TEs in organisms exist as part of their genome defense capabilities, include silencing mechanisms (see below). An alternative perspective is that TEs evolve and can accumulate in genomes because of silencing mechanisms. This perspective implies that the more variable and complex silencing strategies an organism has, the more likely it is to house numerous TEs. Consequently, TEs evolve together with genome silencing mechanisms, allowing genomes to expand. However, true to the modern view that repeat elements (including TEs) are integral functional components of the genome and not merely passive parasites and a dead weight to organisms, it is becoming clear that TEs can counteract host genome defense mechanisms, thus balancing the negative and positive consequences of their propagation. For example, in a fashion similar to viruses interfering with their silencing by their host, recent exciting work in rice demonstrates that miRNAs produced by TEs can interfere with the host epigenetic TE methylation/silencing mechanism. It is likely the “balancing act,” a deliberate (evolved) “sloppiness” of the self-nonself recognition mechanism (of introduced or transposed DNA/RNA molecules), that allows TEs to modify genomes and bring about (temporary) changes for selection forces to act on.

**TEs in Fungi**

Fungi represent good models for studying TEs because a large variability is found among sequenced genomes that either contain large amounts of TEs and repeats (e.g., 28,29), a moderate number, or are nearly devoid of TEs (e.g., 2,3,30). With short generation times, both sexual and asexual representatives, and often easy handling as (molecular genetic) experimental systems, consequences of TE activity are therefore easier to follow at both the individual and the species level in an evolutionary context. Studying TEs in relation to evolutionary time is important since effects on genome diversity become more apparent. From such a perspective, a role is seen for TEs in creating variability upon which selection can act. When species encounter changing or hostile environments, stress often leads to TE activation implying a special role for TEs under such conditions. For example, recent
Fungi can possess both type I and type II TEs that can contribute differently to total genome size. Certain fungi have become model organisms for genetics and molecular biology and in particular, fungi have been useful in furthering our understanding of how TEs are controlled. Pioneering work on quelling in *N. crassa* helped support parallel work in RNA silencing in plants and animals. Key genes, including *dicer*, *argonaute* and *RdRp*, were identified showing involvement in a conserved eukaryotic defense mechanism acting to repress expression of duplicated sequences, including TEs (reviewed in 38). The RIP mutation mechanism has already been mentioned and an important aspect of RIP is that it works on duplicated sequences during the dikaryotic stage, which is after mating in *Neurospora*, but before nuclear fusion and meiosis, leading during sexual reproduction to the accumulation of more-and-more mutations in repetitive loci at each pass. Also in *N. crassa*, components of the RNA silencing machinery were shown to act in silencing unpaired DNA during meiosis (MSUD). Recent work in *Cryptococcus* showed that RNA silencing plays an important role in controlling TEs during sexual reproduction but clearly also controls duplicated genes during vegetative growth. Experiments in fission yeast have provided valuable information on how TEs are packaged into heterochromatin and how this packaging is important for proper centromere functioning; centromeres often accumulate TEs (e.g., 44). Most notable is how heterochromatin is transcribed during the S-phase of the cell cycle and acted upon by the RNA silencing machinery to recruit repressive chromatin complexes including histone H3 lysine 9 methylation and HP1-like chromodomain proteins. For many species, including many fungi, DNA methylation is also an important means to regulate gene expression including controlling TEs (reviewed in 47). Methylation of cytosine residues is often directed to repetitive DNA including TEs, especially in the CpG context, in a variety of organisms including fungi. Interestingly, in several fungi from diverse taxa, CpG dinucleotides are also targeted for mutation in repetitive DNA, such as the C-to-T transitions seen in RIP mutations. The consequence of C-to-T transitions is a general depletion of C/G nucleotides and this hallmark RIP mutation signature has now been seen in diverse taxa from both ascomycete and basidiomycete fungi, including in the *U. hordaei* genome. An essential component of the RIP mutation machinery is RID, a protein with homology to DNA methyltransferases. It has been suggested that the RIP mutation mechanism may have evolved from a preexisting DNA methylation pathway that recognizes CpG dinucleotides to actively mutate TEs upon their replication. Collectively, the many studies on model fungi point to a need for mutational inactivation and control of TEs in genome defense during sexual reproduction.

**Fungal Mating Systems, TEs and Recombination**

Fungi are also excellent models for studying various modes of reproduction, representing both asexuals and sexuals. It should be noted that species typed as asexual may have a cryptic sexual cycle, not yet discovered. Indeed, in some species, recent sequenced genomes have revealed complete mating-related genes suggesting they are capable of sex (for example in the dandruff-causing fungus, *Malassezia globosa*). The variation in sexual reproduction is astounding, encompassing homo- and heterothallic, selfing and outcrossing species with variable degrees of inbreeding (among both haploids or diploids) or outcrossing among them; the various modes are sometimes even found among closely-related species. For sexual species, mating compatibility is governed by specific genetic loci and in its simplest form involved a transcription factor essential for regulating mating. Repressed recombination coinciding with a TE eventually led to sequence divergence and the formation of mating types; in the basidiomycete lineages, recruiting a separate self-nonself recognition system as a second genetic locus occurred (reviewed in 25,58,59). Interesting experimental data from the yeast, *Kluyveromyces lactis*, corroborates the notion that a TE could have evolved into or have become part of a functional control module to promote host sexual reproduction. In a recent study, TE activity was implicated in the mating-type rearrangements that occurred during evolution in *Neurospora* species and which led in several independent occasions to the generation of homothallism (‘same-clone mating system’) from likely ancestral heterothallism (which prevents ‘same-clone mating’). Similarly, in the basidiomycete lineage, rearrangements of mating-type loci to form large nascent sex chromosomes, which leads to increased selfing potential, have been described for multiple species and in all cases, repeats and (LTR) TEs were implicated. TEs, therefore, are associated with steps in the evolution of sexual types.

*...in U. hordaei*

Among the sequenced smut genomes, there was a difference in TE load and diversity and this seemed to be correlated with the bipolar and tetrapolar mating systems: both tetrapolar species *U. maydis* and *S. reilianum* were relatively devoid of TEs (2% and 0.8%, respectively), while the bipolar species *U. hordaei* had significantly larger amounts of various types of both DNA and retroelements (up to 10%, with ~50% in the 527-kb *MAT*-region suggested to be caused by recombination suppression). In addition, the TEs were mostly species-specific with a few very large ‘families’ of related elements indicating recent invasion and proliferation. Of the retroelements, two groups of highly similar elements (Ty1/Copia-like and Ty3/Gypsy-like) were found scattered throughout the genome of only *U. hordaei*. We cannot determine a genome sequences from plant-pathogenic oomycetes (fungus-like organisms) have revealed a role for TE activity in creating variability among effector proteins which are crucial during plant colonization. Similarly, in plant pathogenic fungi, TE activity has been involved in overcoming evolutionary and environmental (host) pressures imposed by resistance genes. It could be argued that evolutionary pressures are different in pathogen-host systems, thereby skewing existing theories and population-genetic models.
In the smuts, the observed differences in TE load between bipolar and tetrapolar species is likely resulting from a combination of mating system and genome defense strategies. In *U. hordei*, recombination was suppressed, impeding purging them, or both. An interesting consequence of the formation of the MAT locus is that the resulting bipolar mating system yields progeny with a higher probability of diploid selfing than progeny produced through the tetrapolar system. In bipolars, one out of two random progeny can mate with any other random progeny from a cross whereas one out of four can do this in tetrapolars; within the population at large, due to the bi-allelic pheromone-receptor locus, the success rate of compatibility should theoretically be 50% in both bipolar and tetrapolar species, thus not promoting outcrossing by one system over the other. Of course, the extent of outcrossing in general depends on the availability of unrelated teliospores germinating in close proximity, which depends on disease severity. In any event, more frequent selfing in bipolars would lead to increased homozygosity and reduced ectopic recombination between non-syntenic repetitive elements. This would have been beneficial to *U. hordei* through an increase in genome stability and would have protected Gypsy-like elements from the purging process. Formation of the bipolar MAT locus therefore would have maintained more adaptive potential such as the Gypsy-like elements constituting a selective advantage over the tetrapolar mating system.

**Selective Pressure from a Host Environment**

Even though increased selfing might be advantageous in the short-term for populations occupying successfully a certain niche, it usually drives populations under (strong) selection pressure slowly to extinction likely because it prevents rapid adaptation. This is especially true in (fast) co-evolving pathogen-host populations under higher selection pressures and experiments aimed at supporting the “Red Queen Hypothesis” (which posits that in an evolutionary system, continuing adaptation is needed in order for a species to maintain its relative fitness among the systems co-evolving) do not favor self-fertilizing mating systems. For *U. hordei*, selective forces would have favored the presence of a compensating mechanism to create the diversity needed to adapt to the ever-changing, hostile host environment and this could be increased activity (and maintenance) of TEs. Indeed, in population genetic studies, when the effective population size decreases (e.g., due to mating restrictions or other bottleneck), TE insertions are found in higher frequencies. In *Ustilago*, and other obligate parasites, the hostile host environment can be considered a bottleneck, leading to smaller effective population sizes.

**RIP and Genome Defense in *U. hordei***

In the smuts, the observed differences in TE load between bipolar and tetrapolar species is likely resulting from a combination of mating system and genome defense strategies. In *U. hordei*, likely origin of these elements. For example, a BLASTn search at the NCBI nr database of a representative Ty1/Copia-type element sequence did not return significant hits, although a BLASTx sequence did not return significant hits, although a BLASTxSearch matched gag-pol elements from Arabidopsis (e-150) to rice (e-131), but all having just only 30–35% amino acid identities. Obviously, gag-pol ORFs (proteins), originating from common ancestral viral elements, are widely maintained and relatively conserved in nature. However, a Ty3/Gypsy class representative sequence did not match any sequences, either as BLASTx or BLASTn. Since many of these *U. hordei* elements seem to be unique in the currently-available sequenced smut fungal genomes, many more sequenced genomes of related species need to become available (possibly identifying an ancestral genome) to solve phylogenies. These species-specific elements likely contain lineage-specific, self/nonself differentiation competencies allowing modulating effects only within the species. 

Perhaps the most interesting of the *U. hordei* retroelements was the Gypsy-like class that showed the highest density in the MAT-1 locus (Fig. 1, row d). Gypsy-like elements are special in that they contain a chromodomain in the C-terminal region of the polyprotein that they encode. It is thought that the chromodomain helps the Gypsy-like elements recognize heterochromatic DNA prior to insertion; the chromodomain interacts with H3K9 methylated histones, which are recruited to and are hallmarks of heterochromatin. *U. hordei* Gypsy polyproteins contain a chromodomain at the C-terminal end (Fig. 2), but no such Gypsy-like elements or chromodomains were identified in the *S. reilianum* or *U. maydis* genomes (some weak similarity was found with the integrase domain in an *U. maydis* HOB5 element). The fact that Gypsy-like elements are enriched at the *U. hordei* MAT-1 locus and that a low recombination frequency has been observed at this locus, suggests that the locus is largely heterochromatic. A key question for the relationship of the Gypsy-like elements to the *U. hordei* MAT-1 locus is whether the elements played a role in the original rearrangement(s)/translocation events that formed the nascent sex chromosome, or accumulated subsequently as
bipolarity favors selfing which would generally increase homozygosity. However, outcrossing is feasible and could introduce new elements, which could be purged through ectopic recombination in subsequent sexual cycles. The Gypsy-like elements may help enforce this bipolar mating system by helping the large MATA locus to remain heterochromatic and suppressed in recombination. TEs abundant within the MATA locus and elsewhere in the genome are then controlled by various means. A functional RNA silencing pathway likely plays an important role in controlling the abundant TEs and may work to enforce heterochromatic silencing during the S-phase of the mitotic cell cycle as in fission yeast or during sexual reproduction as in C. neoformans. As well, cytosine DNA methylation may also help control the TEs in U. hordei.

Our genome study identified a RIP-like mutation mechanism and a resultant observed depletion in C/G nucleotides, only in U. hordei. When all genes and TEs were analyzed, it was evident that the CG depletion was found primarily in TE sequences (see the Supplemental Figure 7 in ref. 1). Here we report on an expanded genome analysis. Calculations yielded a CG LogRatio observed/expected < -1.0 for 1,883 out of the 4,910 identified TEs, but a slight decrease in C/G nucleotides could also be observed in a subset of the U. hordei gene models. Further analysis of these revealed that 495 out of the predicted 7,043 U. hordei genes had a CG LogRatio observed/expected < -1.0 indicating a depletion whereas this was found for only 13 S. reilianum and none of the U. maydis genes (Fig. 3). When comparing these 495 U. hordei genes by BLASTn to each other, at least 96 genes appeared to match duplicated sequences over at least 50% of their length, mostly present in two copies in the genome but also some representing small gene families; these do not match TEs or repeats and code mostly for genes of unknown functions. When comparing these 495 genes to the collection of U. hordei TEs and repeat sequences, 378 genes (76%) matched TE and repeat sequences (at e < 10\textsuperscript{-20}) of which 253 (51%) represent the two largest TE families found in the genome: 183 are related to the Ty1/copia-type elements (Fig. 1, row c) and 70 to the Ty3/Gypsy class elements (Fig. 1, row d). Overall, approximately 200 genes match to TE sequences over at least 25% of their lengths, either at the 5’- or 3’-ends. These are likely chimeric genes and could represent TE insertions. When aligned, clear RIP mutation signatures are observed, including in the duplicated, non-TE genes (as in 1; data not shown).

We mapped these CG-depleted U. hordei genes on the genome and found clear groupings. An extreme example is found on chromosome I where a dense clustering can be seen at the MATA locus. Within the 527 kb MATA region delimited by the a and b mating-type loci and in which 142 gene models have been called,\textsuperscript{11} 73 genes have a calculated CG LogRatio observed/expected < -1.0 (Fig. 1, row g). This suggests that gene duplication is common and that the RIP mutation machinery seems highly active at this locus. Although genes revealing a high degree of RIP mutations are common at the MATA locus, often in stretches of 3 to 5 genes (a stretch next to the b locus has 10), several genes showing no sign of RIP mutations, are frequently interspersed. Interestingly, these genes remained conserved among the smuts, even after the proposed chromosome fusion event in U. hordei, and this finding confirms that RIP-like mutations do not extend beyond duplicated sequences.

Can TE “Modifiers” Influence Mating Systems?

As mentioned, some of the observed duplications involve gene fragments fused to TE-related sequences. Recently, high expression of Gypsy-like elements was observed during ovule development in sexual but not apomictic (asexual) genotypes of plants.\textsuperscript{68} More specifically, expression of entire chimeric Gypsy-like element-gene fusions occurred and it was noted that several of the gene fragments corresponded to genes known to play a role in apomixis. The authors suggested that an autocatalytic process could be at play in sexual genotypes where gene fragments fused to Gypsy-like elements help suppress apomictic development. Given that RNA silencing plays an important role in controlling TEs during sexual reproduction, one could speculate that TEs may coincidentally be active. Perhaps the Gypsy-like elements in U. hordei, upon which RNA silencing (RIP) could act, help enforce a particular sexual behavior by modulating expression of such TE-gene fusions that are involved in sexual reproduction (at the MATA locus). If the modulation is regulated through the TE part, this could potentially allow the Gypsy-like elements control over the nature of sex of its host. As hypothesized above, the bipolar mating system may constitute a selective advantage over the tetrapolar system and the changes that led to this system undoubtedly involved TEs. It has been shown that beneficial mutations can arise not only randomly but also through directed or adaptive mutagenesis, oftentimes due to stress and
damaging radiation and the efficient homologous recombination point to the existence of a very effective (alternative) genome defense and maintenance mechanism in light of the lack of certain epigenetic control mechanism(s). The mechanisms at play in *S. reilianum* remain to be determined; silencing pathway components were present in its genome but it is currently unknown whether these are functional.3 This tetrapolar smut could very well represent an intermediate between *U. hordei* and *U. maydis* but is different in that multiple pheromone-receptor alleles evolved thereby increasing its outbreeding potential to more than 50%.8 Active elimination of TEs may be achieved through increased outcrossing and recombination geared toward purging of repetitive DNA. The loss of RNA silencing genes in *U. maydis*, possibly through the action of TEs followed by recombination, may have created a selective advantage in this species. In this species, variability (among effectors), thought essential and caused by TEs in some other plant pathogens, may have been achieved through recombination or other means. Indeed, with the emerging view that the TE and repeat content, previously thought of as “selfish junk DNA” in organisms, actually represents integral functional components of the ‘genome’ (defined as the complex, informational ‘organelle’ which includes multilevel epigenetic control14), it is interesting to learn how organisms nearly devoid of them achieve and maintain such functional diversity. In conclusion, the smuts represent excellent models for studying the role of TEs, genome defenses and mating systems in creating adaptive potential. A tantalizing hypothesis that TEs have influenced some of these processes, including sexual behavior, needs testing. We are eagerly awaiting the sequencing of other smut genomes to help delineate the various proposed forces shaping them.

**Conclusions**

The lack of TEs and repeats in certain smuts is likely a consequence of an extremely efficient repetitive DNA purging mechanism. The near-absence of repetitive DNA in *U. maydis* might have resulted from the loss of RNA silencing and possibly the RIP mutation mechanism (if present in a progenitor of *U. maydis*). Indeed, disruption of the RNAi machinery in *S. pombe* negatively affects the repair of double-strand DNA breaks and increases homologous recombination.71 Homologous recombination is highly efficient in *U. maydis* and is likely involved in genome maintenance. As well, *U. maydis* can survive extremely high doses of ionizing radiation that would be lethal to other eukaryotes. It therefore seems that *U. maydis* has an efficient mechanism to repair double-stranded DNA breaks. An interesting feature of the *U. maydis* genome is the high frequency of short, 10-bp repeats that populate intergenic sequences and were proposed to be footprints remaining from the purging of TEs.13 Collectively, the short repeats, the ability to survive DNA environmental pressures on the phenotype, e.g., 69,70. An in-depth analysis of the gene fragments fused to TEs is needed to confirm this possibility.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

The authors thank Drs. T. Giraud and J.W. Kronstad for comments on the manuscript, and an anonymous reviewer for pointing out insightful literature.

50. Kemppainen M, Duplessis S, Martin F, Pardo AG. Plant Signaling & Behavior Volume 8 issue 4


