

# LTR Retrotransposons in Fungi

Anna Muszewska\*, Marta Hoffman-Sommer, Marcin Grynberg

Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland

## Abstract

Transposable elements with long terminal direct repeats (LTR TEs) are one of the best studied groups of mobile elements. They are ubiquitous elements present in almost all eukaryotic genomes. Their number and state of conservation can be a highlight of genome dynamics. We searched all published fungal genomes for LTR-containing retrotransposons, including both complete, functional elements and remnant copies. We identified a total of over 66,000 elements, all of which belong to the Ty1/*Copia* or Ty3/*Gypsy* superfamilies. Most of the detected *Gypsy* elements represent *Chromoviridae*, i.e. they carry a chromodomain in the *pol* ORF. We analyzed our data from a genome-ecology perspective, looking at the abundance of various types of LTR TEs in individual genomes and at the highest-copy element from each genome. The TE content is very variable among the analyzed genomes. Some genomes are very scarce in LTR TEs (<50 elements), others demonstrate huge expansions (>8000 elements). The data shows that transposon expansions in fungi usually involve an increase both in the copy number of individual elements and in the number of element types. The majority of the highest-copy TEs from all genomes are Ty3/*Gypsy* transposons. Phylogenetic analysis of these elements suggests that TE expansions have appeared independently of each other, in distant genomes and at different taxonomical levels. We also analyzed the evolutionary relationships between protein domains encoded by the transposon *pol* ORF and we found that the protease is the fastest evolving domain whereas reverse transcriptase and RNase H evolve much slower and in correlation with each other.

**Citation:** Muszewska A, Hoffman-Sommer M, Grynberg M (2011) LTR Retrotransposons in Fungi. PLoS ONE 6(12): e29425. doi:10.1371/journal.pone.0029425

**Editor:** Rosemary J. Redfield, University of British Columbia, Canada

**Received:** March 18, 2011; **Accepted:** November 28, 2011; **Published:** December 29, 2011

**Copyright:** © 2011 Muszewska et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: musze@icm.edu.pl

## Introduction

Mobile elements are genome components that are able to move from one genetic locus to another. The best studied group of mobile elements are transposons (transposable elements) – widespread among all living organisms, they constitute a significant part of most analyzed genomes. In vertebrates and plants, transposon-derived content can exceed half of the whole genome [1]. Transposable elements have been shown to play a crucial role in genome shaping via recombination and expansion events, leading to chromosomal rearrangements and new gene neighborhoods [2], and they have also been shown to alter gene expression [3]. Examples are known where transposon fragments have gained new functions through exaptation and/or adaptation processes [4,5].

Transposable elements (TEs) are traditionally divided into two major classes, based on their dispersion mechanisms [6]. Class I elements (retrotransposons) require an RNA intermediate in their transposition cycle. Retrotransposons synthesize a cDNA copy based on the RNA strand using a reverse transcriptase (RT) related to retroviral RT. Class II elements follow only an excision and insertion cycle. Their basic architecture is simpler, but many complex and variable DNA transposon types have been recently reported [7,8]. Both classes encompass autonomous elements as well as non-autonomous elements which can be mobilized in *trans*, by exogenous enzymes.

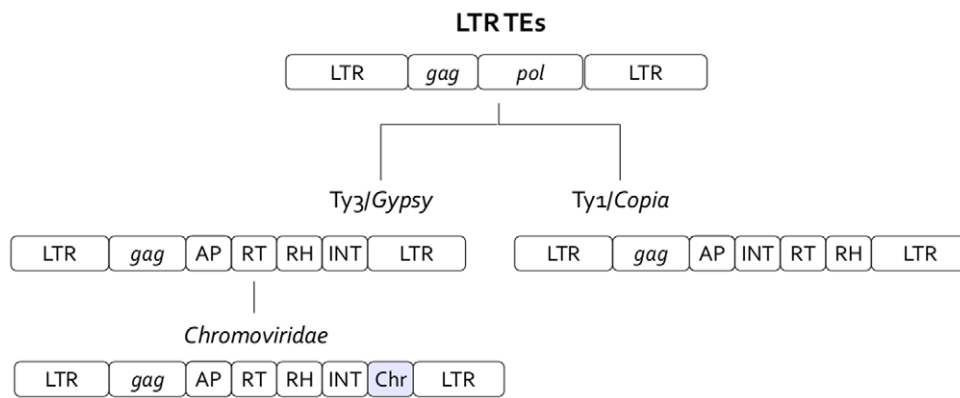
Retrotransposons are further divided into five orders [9]: LTR TEs, characterized by long terminal direct repeats (LTRs) flanking the polyprotein genes, and four different non-LTR orders, named DIRS, PLE, LINE and SINE. Members of the LTR order usually encode two open reading frames (ORFs) (Figure 1). The first ORF

is related to viral structural proteins, termed *gag*, and the latter, termed *pol*, is a polyprotein composed of an aspartic protease (AP) which cleaves the polyprotein, a reverse transcriptase (RT) which produces a cDNA copy of the transposon's RNA, an RNase H (RH) which splits the DNA-RNA hybrid and a DDE integrase (INT) which inserts the cDNA into the host's genome. DDE integrases are endonucleases with a DDE motif and are distantly related to *Mariner* DNA DDE transposases [10]. The order of domains encoded in the *pol* ORF varies between different LTR retrotransposon families, and often additional domains are inserted, e.g. chromodomains [11,12].

LTR retrotransposons are classified into five superfamilies (Figure 1): Ty1/*Copia* (*Pseudoviridae*), Ty3/*Gypsy* (*Metaviridae*), Bel/Pao, retroviruses and ERV (endogenous retroviruses). Up to now, members of the three last superfamilies have been detected only in metazoan hosts, while Ty1/*Copia* and Ty3/*Gypsy* elements have been reported in all eukaryotic lineages. In filamentous fungi, both Ty3/*Gypsy* and Ty1/*Copia* elements have been detected, with *Gypsy* being the most abundant [13]. Most fungal transposons of the Ty3/*Gypsy* superfamily are classified as *Chromoviridae* because of the presence of a chromointegrase (an integrase with a C-terminal chromodomain) [14]. Chromoviruses have been detected in almost all Eukaryotic lineages [15].

As LTR retrotransposons require a multi-compound machinery to be mobile, they easily become non-autonomous. Often one genome harbors both an autonomous element and a related non-autonomous element which acts like a parasite of the functional copy [16].

Genomes tend to fight against the expansion of transposable elements and filamentous fungi have become specialists in this



**Figure 1. A schematic representation of LTR transposable elements present in fungal genomes.**

doi:10.1371/journal.pone.0029425.g001

field: at least three of the main transposon silencing mechanisms – repeat-induced point mutation (RIP), methylation and quelling, as well as the recently discovered sex-induced silencing (SIS) mechanism – have been described in fungi [17,18,19,20].

The presence of transposable elements in fungi has been first reported in yeast in the 1970s, but LTR TE in filamentous fungi were discovered more than 10 years later [21,22]. Although some LTR retroelements have been proven to be functional (e.g. the MAGGY element from *Magnaporthe grisea* [23]), most of the detected mobile elements harbor many stop codons in coding regions which prevents them from being functional.

There has been no whole kingdom analysis of LTR TE in fungi yet; only single genomes, such as those of *Neurospora crassa* or *M. grisea*, have been completely scanned for LTR retroelements. The *M. grisea* genome study has shown an uneven distribution of mobile elements along chromosomes and the presence of novel elements [24]. Here we present the results of a large-scale search for LTR TE in 59 fungal genome sequences. We searched for both full-length and truncated elements and we tried to assign all of them to one of the known LTR TE superfamilies. We attempted to analyze our data from a genome-ecology perspective, i.e. viewing the genome as an environment for (selfish) mobile elements. This approach leads us to questions such as: do high LTR TE counts in fungi result typically from expansions of single clones (many copies of one transposon), single families (one LTR TE type dominates a particular genome), or from a high abundance of all types of elements? Are certain types of genomes (e.g. representing a certain taxonomic group) better environments for mobile elements? Does the lifestyle of the organism correlate somehow with its LTR TE content? Does the presence or absence of known genome defense mechanisms correlate with TE abundance or profile?

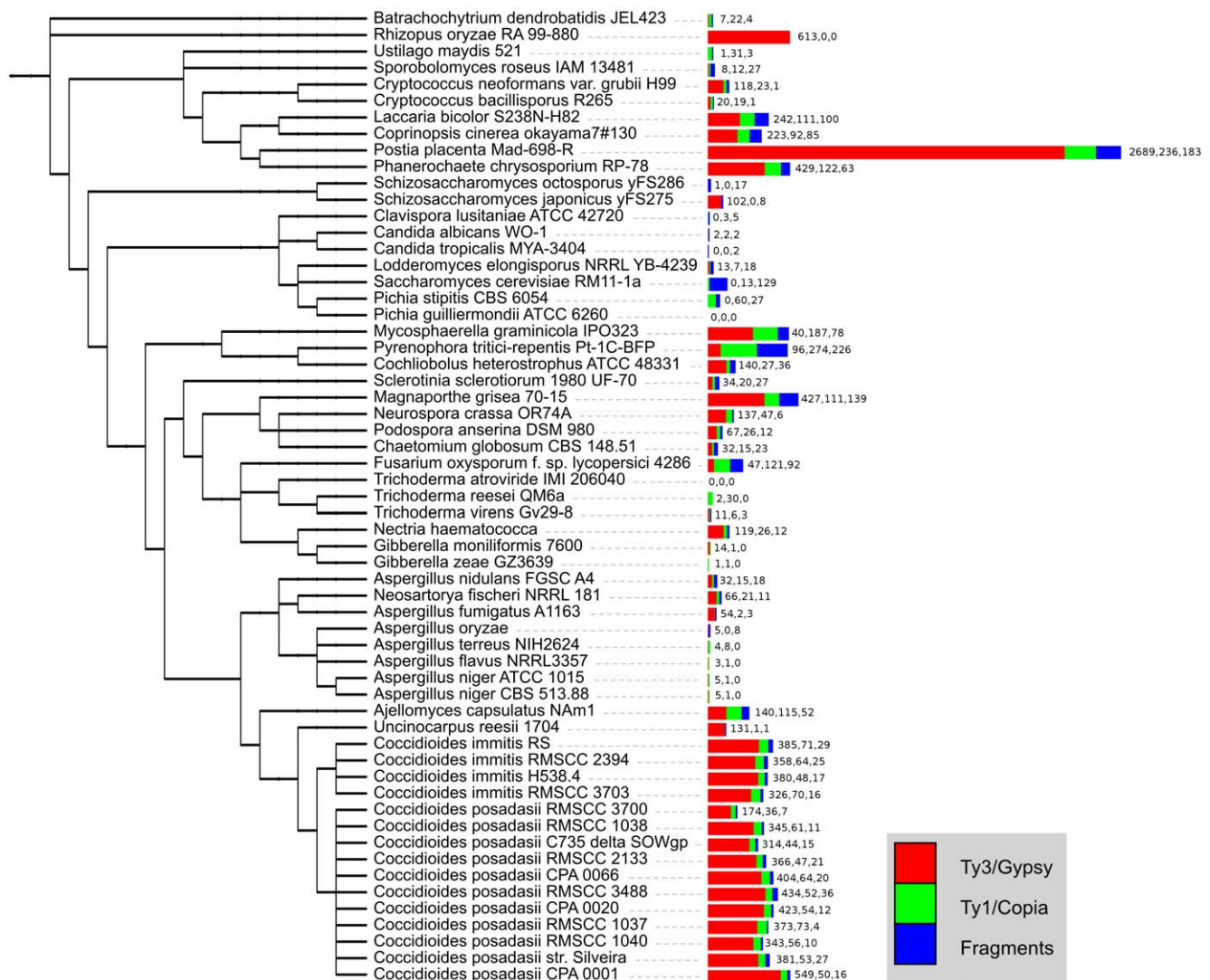
In higher eukaryotes the TE content has been shown to be directly related to the effective population size of the host organism [25]. Studies in *Oxytricha trifallax* have proven that TEs can influence the adaptive capabilities of cells [26]. Our preliminary analysis in fungi shows a great diversity in the abundance of LTR TE among the analyzed genomes, also between closely related species. At the same time, the variability of the identified elements is not that strong: only two LTR TE families (Ty1/*Copia* and Ty3/*Gypsy*) are represented in fungi, and the majority of the detected Ty3/*Gypsy* representatives belong to *Chromoviridae*. An analysis of the most successful element in each genome identifies *Chromoviridae* as the dominating group in fungal genomes. Our data shows that transposon expansions in fungi usually involve both an increase in copy number of individual elements and an increase in the number of different elements.

## Results

### Identification of LTR transposable elements in fungal genome sequences

LTR retrotransposons were detected in 58 out of the 59 analyzed fungal genomes – only *Trichoderma atroviride* lacks any LTR elements. Because most tools designed for the detection of mobile elements have been developed for genomes other than fungal, multiple programs have been applied in this study. Two programs dedicated to LTR retrotransposons were used: LTR Finder and LTR harvest, and one universal tool: a combination of RepeatModeler and RepeatMasker. Both full-length elements and remnant copies were considered in the statistical analysis, as our goal was to explore evolutionary tendencies, not only the current abundance of functional LTR retrotransposons. However, the presence of integrase, transposase, aspartic protease or RNase H was considered a sign of recent activity. More than 66,500 representatives have been identified, the majority of them being short and truncated (see Table S1; for raw data see Table S2). ORF searches revealed that 16,289 elements still carry at least one of the analyzed proteins (Gag, INT, RT, AP or RH) and are therefore considered functional. Genomes abundant in intact elements are also rich in remnant copies. Figure 2 shows the number of elements from each superfamily that are probably still functional per genome. The average LTR retrotransposon number is 1129 per genome and the median is 796. The median is a better measure of the most often encountered state for data with uneven distribution. LTR retrotransposons carrying at least one of the above listed ORFs appear with an average of 276 per genome and with a median of 133. In the following analysis we concentrate on these potentially functional elements.

As stated before [13], the best studied genomes of filamentous fungi are poor in mobile elements. Ascomycota, the best sampled phylum in our study (49 genomes), have genomes with a very variable number of LTR retrotransposons. Most model fungi, belonging to Eurotiales and Saccharomycetes, have very low numbers of LTR retrotransposons: 65 for *Aspergillus nidulans*, 21 for *Saccharomyces cerevisiae*. These species are not representative for the whole phylum. In other ascomycetous groups, such as Dothideomycetes (*Mycosphaerella graminicola* (605)) and Onygenales (*Ajellomyces capsulatus* (307), *Coccidioides immitis* (412–485), *C. posadasii* (217–615)), many genomes are abundant in potentially functional mobile elements. It should be mentioned here that the large differences in LTR TE abundance between different strains of the two analyzed *Coccidioides* species may result from differences in coverage of the genome sequences, since the number of identified



**Figure 2. The number of LTR transposable elements per genome in the dataset.** The length of the bar is proportional to the number of detected LTR retrotransposons of the indicated types; the numbers are also depicted next to every bar. The tree image was prepared with iTol. doi:10.1371/journal.pone.0029425.g002

elements correlates well with the length of the genomic sequence assemblies (data not shown). Sordariales genomes, with the model fungi *Neurospora crassa* (190) and *Podospora anserina* (105), and with the opportunistic *Chaetomium globosum* (70), are not abundant in LTR retroelements, consistent with the well studied mechanisms for genome defense against duplicated content which operate in these organisms [17,27,28]. Hypocreales, represented by Trichoderma (0–35 elements) and Nectriaceae (Giberella 2–15, *Nectria haematococca* 157 LTR TEs), have genomes with moderate and low LTR retrotransposon content. *Magnaporthe grisea*, the only sequenced Magnaporthales representative, outnumbers all other analyzed Sordariaceae, showing a very high LTR retrotransposon content (677 elements) which corroborates results reported in the genome publication [24]. The pattern of LTR retrotransposon distribution in Ascomycota is not clear; some orders have diverse representatives, like Onygenales: the saprophytic *Uncinocarpus reesei* has a low LTR retrotransposon number (133) whereas the systemic infection-causing species has an elevated LTR retrotransposon content (Coccidioides 217–615, *Ajellomyces capsulatus* 307). On the other hand, genomes classified to the subphyla

Taphrinomycotina and Saccharomycotina have LTR retrotransposon numbers below the average for the whole phylum.

The second best represented phylum in our dataset is Basidiomycota, with 8 genomes. Interestingly, both measures for the average value of LTR retrotransposon content per genome, the mean 607 (1980 remnants) and the median 278 (961 remnants), are higher than for Ascomycota (mean 218 (969 remnants), median 119 (796 remnants)). However, the distribution is far from even. Four genomes (*Postia placenta* (3108), *Laccaria bicolor* (453), *Coprinopsis cinerea* (400), *Phanerochaete chrysosporium* (614)) have strong mobile element expansions causing whole-genome size growth while the rest (*Ustilago maydis* (35), *Sporobolomyces roseus* (47), *Cryptococcus neoformans grubii* (155), *Cryptococcus bacillisporus* (40)) are scarce in LTR retrotransposon content. Plant-related Basidiomycota (both symbiotic and pathogenic: *P. placenta*, *L. bicolor*, *P. chrysosporium*) encode multiple LTR retrotransposons, with the exception of *U. maydis* which has a very compact genome [29]. The tiny genome of *Sporobolomyces roseus* also has a modest LTR TE number.

*Postia placenta* represents an interesting case. This species exhibits a huge expansion of LTR TEs, outnumbering all other known

fungi. The elements identified in the *P. placenta* genome can be grouped into 262 LTR TE families (where a family is characterized by more than 80% similarity over the whole DNA sequence). The increase in overall LTR TE number in *P. placenta* must have been achieved by the multiplication and variation of many groups of elements, as the closest relative in the analyzed genome set, *Phanerochaete chrysosporium*, has representatives of only 97 families. The abundance of LTR TE in *P. chrysosporium* genome was previously reported [30]. Many of the elements identified in the *P. placenta* genome encode at least one protein domain and thus should have expanded recently, which might indicate a recent stress in the history of this species and/or a recent decrease in the effective population size [25]. This fungus is unique among cellulose degrading microbes in its glycoside hydrolase gene set and reveals many unusual biochemical features [31]. It would be tentative to study the influence of such an outburst of retrotransposons on the adaptation of *P. placenta* to its ecological niche.

The only included representative of early diverged Eumycota – *Batrachochytrium dendrobatidis*, belonging to the phylum Chytridiomycota – has a low number of LTR elements (33). This is in contrast with the only Mucoromycotina in this study, *Rhizopus oryzae*, which encodes as many as 742 LTR retrotransposons [32].

Since we have not found any taxonomic pattern in genomic LTR TE abundance, we looked whether there was any correlation between the number of LTR TEs per genome and the organism's lifestyle or ecological niche (see Table S1). We observed several tendencies. First, genomes expand in plant related fungi. Second, in Ascomycota saprophytes are generally less abundant in LTR TEs than in non-saprophytes (Table S1). Since our findings are quite weak, we conclude that other factors, not included in our analysis, may influence LTR TE abundance in fungi.

### Evolutionary analysis of reverse transcriptase (RT), RNase H (RH), integrase (INT) and aspartic protease (AP)

We then tried to investigate the evolutionary history that led to the currently observed LTR TE profiles. We estimated the divergence of LTR retrotransposons basing on sequence similarity and localization of protein domains within the transposon sequence. For this purpose we used several domains that are the milestones of phylogenetic studies: RT, RH, INT and AP. For each of these domains, CD-HIT and CLANS clustering were performed. The first step reduced the number of analyzed sequences filtering out the highly similar variants of each element. The second step produced a graphical overview of the variability of the protein domain of interest. In all cases three clades could be identified, corresponding to Ty3/*Gypsy* (with the dominance of *Chromoviridae* elements, but including also some representatives with similarity to the Ylt1 element from *Yarrowia lipolytica* [33]; shown in Figures 3, 4, 5 in dark blue), Ty1/*Copia* (green), and an outlier group (shown in light blue; discussed below). As expected, the closer to each other the domains are located within the *pol* ORF, the more similar the clustering images are. The clustering images of reverse transcriptase and RNase H are almost identical, corroborating the opinion that both domains are indispensable for full catalytic functionality of LTR elements. The RT and RH domains are transmitted together and form one protein when expressed [34].

### Reverse transcriptase

The reverse transcriptase (RNA polymerase) is responsible for DNA synthesis using RNA as a template. The RT domain in the Pfam database is represented by two separate profiles: RVT\_1 (Pfam:PF00078) and RVT\_2 (Pfam:PF07727). Both are described as present in a variety of mobile elements. In our sequence

searches the RVT\_1 profile associated with Ty3/*Gypsy*, whereas RVT\_2 associated with Ty1/*Copia* LTR retrotransposons. Three groups are clearly distinguishable in the CLANS clustering image (Figure 3). Ty3/*Gypsy* form the biggest clade visible on the right side of the image (dark blue). Most filamentous fungi carry more Ty3/*Gypsy* than Ty1/*Copia* elements. In the left part of the image Ty1/*Copia* elements are grouped. Both Ty3/*Gypsy* and Ty1/*Copia* elements were identified in genomes of all taxonomic groups in the analyzed dataset. Some sequences diverge from these major clans, forming a third group (light blue). This group is comprised of degenerated sequences scattered with stop codons and of complex transposons which result from multiple insertions into one genomic locus. These complex transposons contain many additional domains inserted into the structure of typical Ty1 or Ty3 elements (to illustrate this phenomenon, we present the domain architecture of 4 such complex transposons in Figure S1).

### RNase H

In retrotransposons RNase H is responsible for the degradation of the RNA template in the DNA-RNA hybrid. The Pfam RNaseH profile (PF00075) is not sensitive enough to detect transposon-related RHs. Sequence searches conducted with this profile rendered less than 10% of RH domains present in the analyzed dataset and in reference sequences known from literature. This phenomenon can be explained by the divergence of RH domains in LTR retrotransposons from the canonical RNase H described in *E. coli* and viruses [34]. In order to find specific RHs a new “LTR-TE-oriented” profile was built. For this purpose, sequences of the *pol* polyprotein of fungal LTR retrotransposons known from literature were collected from the NCBI protein database. Using this profile, we found three separate clans. The obtained CLANS image (Figure 4) is almost identical to that for the RT clustering, corroborating their direct neighborhood in the mature reverse transcriptase machinery.

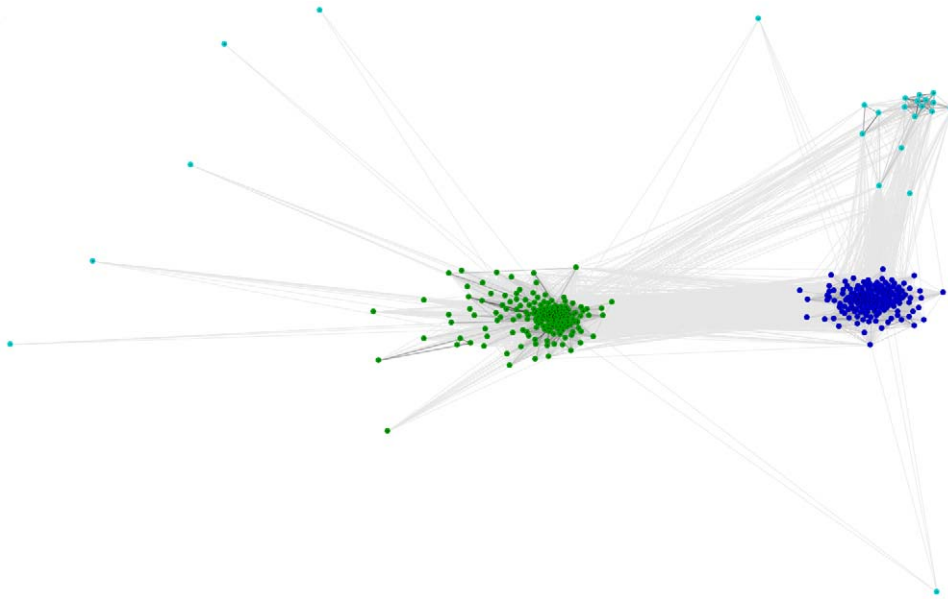
### Integrase

The integrase catalyzes insertion of the retrotransposon cDNA into the genome of a host cell. It consists of three protease-resistant domains, but only rve (PF00665) – the integrase core domain – is well conserved among a variety of organisms. It belongs to the Pfam RNase H clan. This domain was extracted and clustered. The result shows two major clades as well as several outlier sequences (Figure 5). The differences between Ty3/*Gypsy* and Ty1/*Copia* are less sharp here than in the images for RT and RH, resulting in a more compact relationship between the clusters. The outlying sequences visible in the CLANS output are in most cases non-functional copies, with many stop codon mutations. The integrase domain is localized differently in the *pol* gene in Ty3/*Gypsy* and Ty1/*Copia* retrotransposons. In Ty3/*Gypsy* elements integrase is usually the last element in the *pol* gene whereas in Ty1/*Copia* it is located between the sequences coding the AP and the RT. The clustering analysis shows also that integrases are less diverged between Ty3/*Gypsy* and Ty1/*Copia* than the RT and RH domains.

### Aspartic protease

The aspartic protease is responsible for processing the large transposon transcripts into smaller protein products. Four different aspartic protease profiles from the Peptidase\_AA clan (CL0129) were used in our study. Two of them are dominant and unevenly distributed: RVP (PF00077) is found in Ty1/*Copia* and RVP\_2 (PF08284) in Ty3/*Gypsy* LTR TEs. The Peptidase\_A2B (PF12384) and Peptidase\_A2E (PF12382) domains are found only in several copies. The high variability of the AP resulted in a very low





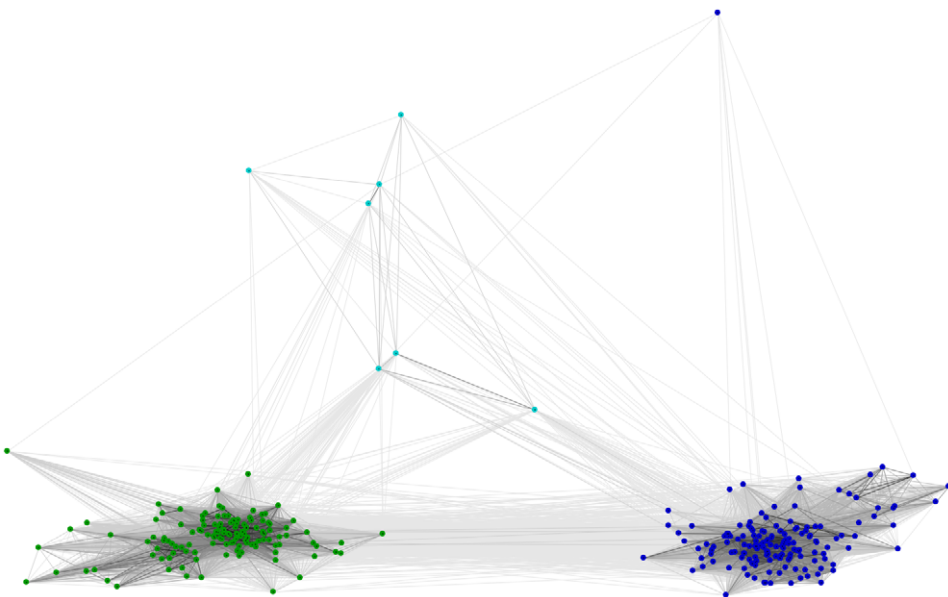
**Figure 3. CLANS clustering of RT domains.** Dark blue – Ty3/Gypsy, green – Ty1/Copia, light blue – fragmented or complex transposons. The intensity of the connecting lines reflects the level of sequence similarity. doi:10.1371/journal.pone.0029425.g003

detection rate in LTR TE sequences when compared to the RT, RH and INT domains. There are neither clades specific for fungal taxonomic groups nor clades related to a specific type of LTR TE (Figure 6). The aspartic proteases are said to be difficult to analyze due to their low similarity and different evolution rates. This subject is being studied in detail by the authors of GyDb [35].

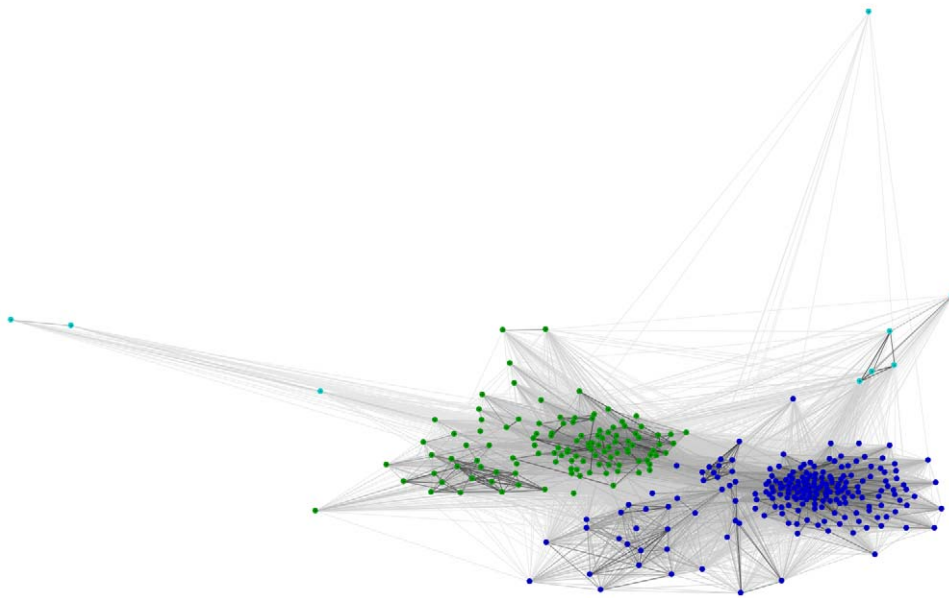
Taken together, the presented CLANS results show that all analyzed *pol*-encoded domains display a similar clustering pattern – with Ty1/Copia and Ty3/Gypsy elements forming two major clusters – which remains in concordance with the currently accepted view of LTR TE evolutionary history.

### Most successful element analysis

In order to see which kinds of transposable elements are most successful in which genomes, we searched for the highest copy number LTR TE in every analyzed genome. The identified elements were then subjected to a phylogenetic analysis. The *pol* gene (encoding the RT, RH, INT and AP proteins) was chosen for this analysis because of its conservation. The only other ORF present in most LTR retrotransposons – *gag* – is much more variable and it is difficult to identify using sequence profiles. Not all of the most successful elements encode a complete set of *pol*-encoded protein domains suggesting their non-autonomous



**Figure 4. CLANS clustering of RH domains.** Dark blue – Ty3/Gypsy, green – Ty1/Copia, light blue – fragmented or complex transposons. doi:10.1371/journal.pone.0029425.g004



**Figure 5. CLANS clustering of INT domains.** Dark blue – Ty3/*Gypsy*, green – Ty1/*Copia*, light blue – fragmented or complex transposons.  
doi:10.1371/journal.pone.0029425.g005

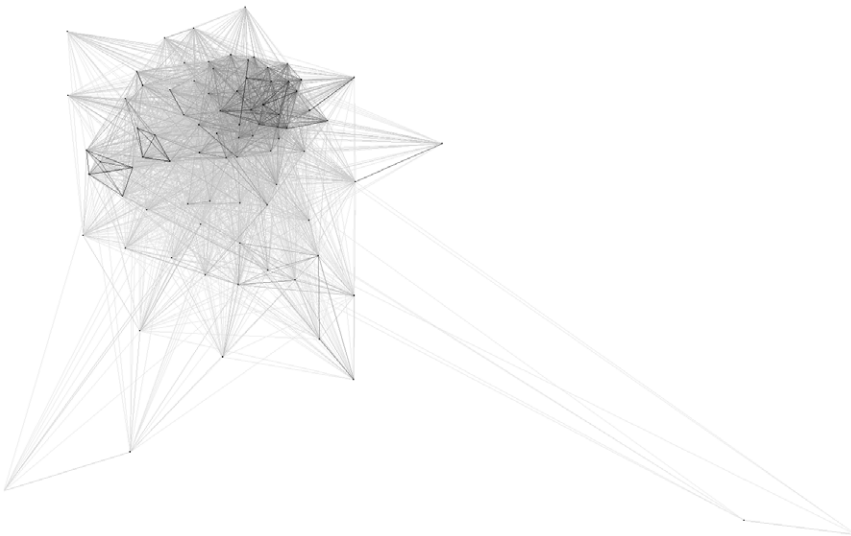
character or a non-functional state. This distribution depicts the battle between genome defense mechanisms and the expansion of mobile elements in fungi. The presence of multiple stop codons in the majority of the characterized mobile elements – which most likely renders them non-functional – can be regarded as an emanation of this phenomenon. Most of the elements have incomplete *pol* genes lacking one or more of the core components (INT, RT, RH, AP).

Some most successful elements have been excluded from the phylogenetic analysis: the biggest LTR TE families from *Gibberella zea*, *Candida tropicalis* MYA-3404, *Candida albicans* WO1, *Schizosaccharomyces octosporus* yFS286, *Aspergillus flavus* and both *Aspergillus niger* genomes have only one member and in consequence cannot be considered successful in expanding along the genome, *Candida guilliermondii* and *Trichoderma atroviride* have no LTR TE with a

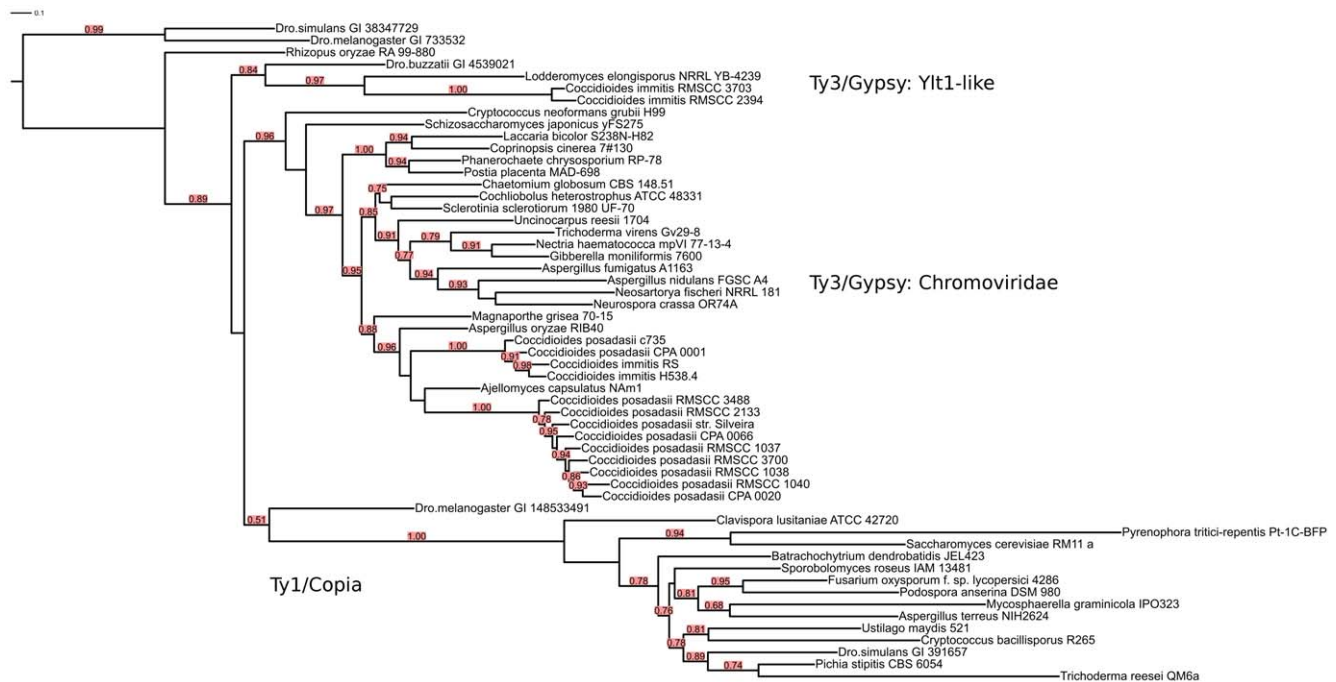
detectable *pol* gene at all. Some of these genomes have already been described as scarce in repetitive content [36,37].

The obtained phylogenetic tree (Figure 7) depicts the relationships between the most successful elements from all analyzed genomes. It shows clearly that *Chromoviridae* (belonging to the Ty3/*Gypsy* superfamily) are almost twice as successful as Ty1/*Copia* in dominating fungal genomes. Elements similar to *Yarrowia lipolytica* Ylt1 dominate in 3 genomes.

In almost every taxonomical range the history of genome invasion was different which is reflected by the differences and conservation of the most successful elements among taxa. The genomes of *Coccidioides immitis* and *C. posadasii* were invaded by at least two very dynamically dispersing elements, all belonging to Ty3/*Gypsy*. Two *C. immitis* strains have Ylt1-like elements related to those in the *Lodderomyces elongisporus* genome. The *Chromoviridae*



**Figure 6. CLANS clustering of AP domains.** These sequences do not form any well-defined groups.  
doi:10.1371/journal.pone.0029425.g006



**Figure 7. Maximum-likelihood phylogenetic tree of the most successful LTR transposable elements in the analyzed fungi.** Concatenated amino acid sequences of RT, INT, RH and AP protein domains were used as the dataset. The phylogenetic analysis was performed with PhyML. Approximate likelihood ratio test SH-like branch supports above 50% are shown. The tree image was prepared with iTol. doi:10.1371/journal.pone.0029425.g007

elements dominating in most *Coccidioides* and in the *Ajellomyces capsulatus* genome are closely related. Four different Agaricomycotina genomes were conquered by a very similar element. Ty3/Gypsy elements insert preferentially into heterochromatin regions whereas Ty1/Copia retrotransposons have been reported to integrate into transcriptionally active regions of the genome [14]. Recently, heterochromatin has been considered a crucial player in speciation [38]. If so, the role of abundant TE elements in this process needs elucidation. The activity of mobile elements could serve as an indicator of a strong need for adaptation to new environmental conditions.

The phylogenetic tree shown in Figure 7 does not overlap with the taxonomic classification of the host genomes, indicating that there are no general rules as to which types of LTR TEs are most successful in a particular taxon. In the Ascomycota division the frequency of TEs differs between subphyla (Saccharomycotina, Taphrinomycotina and Eurotiomycotina). *Saccharomyces cerevisiae* has more Copia elements than Gypsy, and in our study also the most successful element in this species is a Ty1/Copia type element, which is consistent with previous studies [39]. Some elements seem to be common to a taxonomical group, e.g. the most successful Ty3/Gypsy element in three Hypocreales species: *Nectria haematococca*, *Gibberella moniliformis* and *Trichoderma virens*. In the same order, a Copia element was the most successful in the genome of *T. reesei*. The highest-copy elements in most Eurotiales (*Aspergillus nidulans*, *Neosartorya fischeri* and *A. fumigatus*) display similarity to that of *Neurospora crassa*, a member of Sordariales.

Basidiomycota genomes were ‘invaded’ by different elements. The most successful elements in two *Cryptococcus* genomes belong to two separate LTR TE superfamilies. Surprisingly, four other Basidiomycota genomes: *Phanerochaete chrysosporium*, *Coprinopsis cinerea*, *Laccaria bicolor* and *Postia placenta*, were invaded by Ty3/Gypsy elements which cluster together on the phylogenetic tree. These elements might descend from some common ancestor.

The most successful element in the genome of the orphan representative of Mucoraceae, *Rhizopus oryzae*, is a Ty3/Gypsy LTR TE. This element does not cluster together with any other sequence, possibly as a consequence of its early divergence. The only chytrid *Batrachochytrium dendrobatidis* was invaded by a Ty1/Copia.

## Discussion

LTR TEs are ubiquitous elements present in almost all analyzed genomes. A large fraction of the identified LTR TEs is degraded but full copies can also be found. LTR TE variability can be observed even between very closely related species (e.g. between the two *Coccidioides* species), showing that otherwise highly similar organisms can be in some cases identified by their LTR TE content, whereas in other cases metabolically distant strains can harbour almost identical (presumably ancestral) LTR TEs (*Aspergillus niger*). These two examples suggest that using LTR TEs for strain differentiation could be useful for analyzing recent changes, although it is clearly not suitable for the analysis of old historical events. Also, gradual changes in LTR TE content should be considered for highly “transpositionally” active and dynamically changing genomes; even strains of a single species can display a certain percentage of variability in their mobilome during cultivation, as an answer to different environmental conditions.

In this work we searched only for LTR-containing retrotransposons and we were able to identify elements representing both of the LTR TE superfamilies known to reside in fungal genomes: Ty3/Gypsy and Ty1/Copia. A preference for Copia over Gypsy elements can be noticed in some fungi (e.g. *Saccharomyces cerevisiae*) but these are exceptions – in most genomes both groups are present in high numbers and Gypsy elements prevail. Of the Ty3/Gypsy superfamily, *Chromoviridae* representatives were most frequently found in the analyzed genomes. We observe that in

genomes rich in LTR TE the highest copy element is usually more numerous than in genomes scarce in LTR TEs.

More sequencing data will be necessary for general patterns in the distribution of LTR transposons in fungal genomes to become visible. Of the Ascomycota, *Ajellomyces capsulatus* and *Coccidioides* – both representatives of pathogenic orders classified to Onygenales – have many LTR TEs while *Uncinocarpus reesii* – a closely related non-pathogen – has very few elements. Also among the Pezizomycotina, the non-pathogenic *Neurospora crassa* shows a low TE content whereas the plant pathogen *Magnaporthe grisea* displays a high TE content. However, the simplifying conclusion that the presence of multiple LTR TEs is directly related to a pathogenic niche should not be drawn on the basis of these scarce examples.

Basidiomycota are strongly underrepresented in our dataset, which prevents us from identifying any patterns of LTR TE distribution in this taxon. More genomes of Basidiomycota are currently being sequenced so a broader analysis will be possible in the near future.

Among the basal fungi, *Rhizopus oryzae*, representing Mucoraceae in our study, has a genome abundant in LTR TEs, but another basal fungus, *Batrachochytrium dendrobatidis* – the only available Chytridiomycota – has a completely different TE profile, with multiple remnant copies and few complete elements. It has been suggested that basal fungi should have genomes more abundant in Chromoviruses (the most numerous class of LTR TE in Fungi) than Asco- and Basidiomycota [15]. One of the possible hypotheses is that *B. dendrobatidis* is an exception among basal fungi in that it carries a very efficient TE elimination system. If so, other basal fungi would still be rich in TEs. The alternative scenario is that the TE content of most Chytridiomycota is more similar to that of the common Metazoa/Fungi ancestors. The typical fungal TE repertoire would then be established later, in separate evolutionary events. TE loss (compared to *Rhizopus oryzae*) could have been favored later in the evolution of *B. dendrobatidis*, similarly as can be observed in the genomes of the Eurotiales as well as of certain Sordariales, Taphrinomycotina and Saccharomycotina. More sequencing data regarding basal fungal lineages would be necessary to decide which of these two scenarios has actually taken place.

TE expansions are popular, but seem to appear independently in distant genomes at different taxonomical levels. Some expansions seem to have occurred in the ancestor of Onygenales (*Coccidioides*, *Ajellomyces*) and in the ancestor of Polyporales (*Phanerochaete*, *Postia*). Others are shared only by strains of closely related species. Most expansions are genome specific and may constitute a response to some specific conditions favoring genome variability instead of stability. Retrotransposon expansions seem to be individual stories in *Coccidioides*, where highly divergent LTR TEs dominate the genomes of closely related strains, in contrast to *Aspergillus niger*, where the TE content is common. In *Coccidioides*, high variability might have been advantageous because of the extreme changes in their ecological niches and the recent speciation between *C. immitis* and *C. posadasii*. As mentioned before, TE content in *Coccidioides* may be related to genome sequencing coverage, not to real numbers of TEs.

In our analysis we could follow the stepwise process of degradation of an LTR TE, with the least conserved domains being proteases and *gag*'s (according to protein definitions and profiles supplied by Pfam). The core integrase domain (rve) is one of the top scoring domains in Pfam showing that this core fold is favored in evolution. We show that the integrase domain is conserved in transposable elements; however, this observation is directly dependent on the Pfam domain definition. *gag* genes are

said to be diversified and the available protein sequence profiles are not applicable for distant elements.

The INT-RH-RT set seems to be inherited together, while the protease (AP) is more variable. This could mean that the protease can be exchanged for some similar “module”, but it is more likely that the protein domain profiles of the various aspartic proteases available in databases are still not sensible enough to detect all the existing variants of this protein.

The obtained data points to a complex phylogeny of LTR TEs in fungi, including expansions, losses and transfers of retrotransposons in almost every species and strain. LTR TE number, distribution and state of conservation can be a valuable source of information about genome dynamics and evolutionary strategies. Further, we still think that there may be a dependency between the quantity and quality of LTR retrotransposons and the host ecological niche, but more sequencing data would be required to investigate this possibility.

## Materials and Methods

### Data Mining

Genome sequences have been obtained from sequencing consortia: Fungal Genome Initiative (BROAD Institute) and the DOE Joint Genome Institute (JGI). All downloads were performed before December 27<sup>th</sup>, 2008, and only genomes publicly available in March 2010 were included into the analysis. To detect all the available sequences corresponding to LTR retrotransposons, three different programs were applied on each genome sequence. There are no fungal-specific transposable element-directed tools thus a combined search was needed to obtain reliable results. Two tools, LTR harvest [40] and LTR Finder [41], are dedicated to this group of mobile elements. Both programs search for LTR TE-specific features like long terminal repeats. The third tool we used is a set composed of RepeatModeler followed by RepeatMasker 3.3.0. RepeatMasker was run with the RepBase library of manually curated mobile element and repetitive DNA sequences as a reference dataset [42]. RepeatModeler was run separately on each genome in order to produce consensus sequences for genome specific repeat classes. All genomic consensus sequences together with the fungal subset of RepBase database were used as the reference library for RepeatMasker searches, carried out on each genome. Results from all three predictions were merged, yielding a set of full length and truncated LTR retrotransposons. Duplicated hits were removed from the cumulative result based on overlaps exceeding 80% of length of the shorter sequence.

### Most successful element analysis

Records for each genome were clustered with a threshold of 80% of nucleotide sequence similarity using cd-hit-est from the CD-HIT package in order to get LTR retrotransposon family count per genome [43]. Genomes harboring only orphan elements were discarded from this analysis. The richest family from each genome was selected and the longest representative from this family was further analyzed. Each element was translated in 6 frames and the coding sequence was extracted. If the longest element had the coding region distributed among many frames or was a complex transposon, the next longest element was chosen. All elements were clustered with CLANS to observe general tendencies. The sequences corresponding to 4 protein domains: RT, INT, AP and RH, were extracted. Protein sequences were aligned with the localpair iterative algorithm implemented in Mafft [44]. Conserved columns from each multiple sequence alignment were chosen with TrimAl [45]. The selected set of columns was concatenated with an in house Python script. The



most suitable model for phylogenetic analysis was selected with ProtTest [46]. According to the AIC criterion, an LG+G+F model with a score of 0.63 was the most suitable for the dataset. Maximum likelihood analysis of the pol region was carried out with PhyML on-line facility [47] with the following settings: LG model of amino acid substitution, 4 categories in gamma model with the shape parameter estimated as 2.48. Instead of bootstrap replicates branch support was calculated using an approximate likelihood ratio test. Trees were visualized in iTOL [48]. *Drosophila simulans* (GI: 38347729) and *D. melanogaster* (GI: 733532) sequences belonging to Bel/Pao LTR retrotransposons were set as an outgroup. *D. simulans* (GI: 391645) was used as a type representative of *Copia* elements. *D. buzzatii* (GI:4539021) *Oswaldo* transposon was used as a *Gypsy* outgroup and *D. melanogaster* (GI:148553491) as a *Gypsy* type specimen.

### Protein analysis

Since the order of the encoded proteins and the sequence similarity of the reverse transcriptase form the base for LTR retrotransposon classification, repetitive elements were screened for retrotransposon-related protein domains. Each putative LTR retrotransposon was translated in 6 frames with transeq from the EMBOSS package [49] and searched with HMMsearch (from HMMer3.0 package [50]) against 13 HMM (Hidden Markov Model) profiles corresponding to 8 different protein domains (INT, RT, AP, RNase H, Gag, Chromo, and two RT-related: RVT\_thumb and RVT\_connect) included in the Pfam25 database [51] and one developed specially for this project: the Pfam RNase H profile does not include retrotransposon RNase H sequences which are known to vary from canonical RNase H sequences [34],

so we made our own RNase H profile, using known fungal retrotransposon RNase H sequences together with two PDB [52] non-fungal structures for alignment guiding purposes. Sequence searches were automated with pfam\_scan.pl, a tool available at the Pfam database site [51], which enables the user to create a database of HMM profiles to be searched with a database of protein sequences. Protein sequences corresponding to the mentioned protein domains were clustered together with CD-HIT to remove highly similar sequences. The clan representatives were clustered in CLANS, a tool which enables clustering and visualization of sequence similarities [53].

### Supporting Information

**Figure S1** Domain architecture of four selected complex transposons. 1 and 2 are retrotransposons found in the *Pyrenophora tritici-repentis* genome, 3 was identified in *Rhizopus oryzae* and 4 in the *Chaetomium globosum* genome. (TIF)

**Table S1** Transposon content and selected features of all analyzed fungi. Includes citations to references [54–88]. (XLS)

**Table S2** All mobile elements identified in this work (raw data). (CSV)

### Author Contributions

Conceived and designed the experiments: MG AM. Performed the experiments: AM. Analyzed the data: AM MG MH-S. Wrote the paper: AM MH-S.

### References

- Morse AM, Peterson DG, Islam-Faridi MN, Smith KE, Magbanua Z, et al. (2009) Evolution of genome size and complexity in *Pinus*. PLoS One 4: e4332.
- Rachidi N, Barre P, Blondin B (1999) Multiple Ty-mediated chromosomal translocations lead to karyotype changes in a wine strain of *Saccharomyces cerevisiae*. Mol Gen Genet 261: 841–850.
- Zemajtel T, Kielbasa SM, Arndt PF, Chung HR, Vingron M (2009) Methylation and deamination of CpGs generate p53-binding sites on a genomic scale. Trends Genet 25: 63–66.
- Piriyapongsa J, Rutledge MT, Patel S, Borodovsky M, Jordan IK (2007) Evaluating the protein coding potential of exonized transposable element sequences. Biol Direct 2: 31.
- Sasaki T, Nishihara H, Hirakawa M, Fujimura K, Tanaka M, et al. (2008) Possible involvement of SINEs in mammalian-specific brain formation. Proc Natl Acad Sci U S A 105: 4220–4225.
- Craig NL (2002) Mobile DNA II. Washington, D.C.: ASM Press. pp xviii, 1204 p., [1232] p. of plates p.
- Kapitonov VV, Jurka J (2007) Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet 23: 521–529.
- Pritham EJ, Putliwala T, Feschotte C (2007) Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. Gene 390: 3–17.
- Wicker T, Keller B (2007) Genome-wide comparative analysis of copia retrotransposons in *Triticeae*, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. Genome Res 17: 1072–1081.
- Capy P, Langin T, Higuat D, Maurer P, Bazin C (1997) Do the integrases of LTR-retrotransposons and class II element transposases have a common ancestor? Genetica 100: 63–72.
- Kordis D (2005) A genomic perspective on the chromodomain-containing retrotransposons: Chromoviruses. Gene 347: 161–173.
- Llorens C, Fares MA, Moya A (2008) Relationships of gag-pol diversity between Ty3/Gypsy and Retroviridae LTR retroelements and the three kings hypothesis. BMC Evol Biol 8: 276.
- Daboussi MJ, Capy P (2003) Transposable elements in filamentous fungi. Annu Rev Microbiol 57: 275–299.
- Gao X, Hou Y, Ebina H, Levin HL, Voytas DF (2008) Chromodomains direct integration of retrotransposons to heterochromatin. Genome Res 18: 359–369.
- Gorinsek B, Gubensek F, Kordis D (2004) Evolutionary genomics of chromoviruses in eukaryotes. Mol Biol Evol 21: 781–798.
- Kejnovsky E, Kubat Z, Macas J, Hobza R, Mracek J, et al. (2006) Retand: a novel family of gypsy-like retrotransposons harboring an amplified tandem repeat. Mol Genet Genomics 276: 254–263.
- Chicas A, Cogoni C, Macino G (2004) RNAi-dependent and RNAi-independent mechanisms contribute to the silencing of RIPed sequences in *Neurospora crassa*. Nucleic Acids Res 32: 4237–4243.
- Galagan JE, Selker EU (2004) RIP: the evolutionary cost of genome defense. Trends Genet 20: 417–423.
- Wang X, Hsueh YP, Li W, Floyd A, Skalsky R, et al. (2010) Sex-induced silencing defends the genome of *Cryptococcus neoformans* via RNAi. Genes Dev 24: 2566–2582.
- Windhofer F, Catchside DE, Kempken F (2000) Methylation of the foreign transposon *Restless* in vegetative mycelia of *Neurospora crassa*. Curr Genet 37: 194–199.
- Cameron JR, Loh EY, Davis RW (1979) Evidence for transposition of dispersed repetitive DNA families in yeast. Cell 16: 739–751.
- Dobinson KF, Hamer JE (1993) The ebb and flow of a fungal genome. Trends Microbiol 1: 348–352.
- Nakayashiki H, Kiyotomi K, Tosa Y, Mayama S (1999) Transposition of the retrotransposon MAGGY in heterologous species of filamentous fungi. Genetics 153: 693–703.
- Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, et al. (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 434: 980–986.
- Lynch M, Conery JS (2003) The origins of genome complexity. Science 302: 1401–1404.
- Nowacki M, Higgins BP, Maquilan GM, Swart EC, Doak TG, et al. (2009) A functional role for transposases in a large eukaryotic genome. Science 324: 935–938.
- Clutterbuck JA (2010) Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. Fungal Genet Biol 48: 306–326.
- Espagne E, Lespinet O, Malagnac F, Da Silva C, Jaillon O, et al. (2008) The genome sequence of the model ascomycete fungus *Podospora anserina*. Genome Biol 9: R77.
- Kamper J, Kahmann R, Bolker M, Ma LJ, Brefort T, et al. (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature 444: 97–101.
- Novikova OS (2010) [Diversity and evolution of LTR retrotransposons in the genome of *Phanerochaete chrysosporium* (Fungi: Basidiomycota)]. Genetica 136: 725–733.
- Martinez D, Challacombe J, Morgenstern I, Hibbett D, Schmol M, et al. (2009) Genome, transcriptome, and secretome analysis of wood decay fungus *Postia*

- placenta* supports unique mechanisms of lignocellulose conversion. Proc Natl Acad Sci U S A 106: 1954–1959.
32. Ma LJ, Ibrahim AS, Skory C, Grabherr MG, Burger G, et al. (2009) Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. PLoS Genet 5: e1000549.
  33. Schmid-Berger N, Schmid B, Barth G (1994) Ylt1, a highly repetitive retrotransposon in the genome of the dimorphic fungus *Yarrowia lipolytica*. J Bacteriol 176: 2477–2482.
  34. Malik HS (2005) Ribonuclease H evolution in retrotransposable elements. Cytogenet Genome Res 110: 392–401.
  35. Llorens C, Futami R, Renaud G, Moya A (2009) Bioinformatic flowchart and database to investigate the origins and diversity of clan AA peptidases. Biol Direct 4: 3.
  36. Cuomo CA, Guldener U, Xu JR, Trail F, Turgeon BG, et al. (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science 317: 1400–1402.
  37. Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, et al. (2011) Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. Genome Biol 12: R40.
  38. Hughes SE, Hawley RS (2009) Heterochromatin: a rapidly evolving species barrier. PLoS Biol 7: e1000233.
  39. Neueglise C, Feldmann H, Bon E, Gaillardin C, Casaregola S (2002) Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. Genome Res 12: 930–943.
  40. Ellinghaus D, Kurtz S, Willhoeft U (2008) LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics 9: 18.
  41. Xu Z, Wang H (2007) LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res 35: W265–268.
  42. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, et al. (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110: 462–467.
  43. Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26: 680–682.
  44. Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33: 511–518.
  45. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25: 1972–1973.
  46. Abascal F, Zardoya R, Posada D (2005) ProfTest: selection of best-fit models of protein evolution. Bioinformatics 21: 2104–2105.
  47. Guindon S, Delsuc F, Dufayard JF, Gascuel O (2009) Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol 537: 113–137.
  48. Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23: 127–128.
  49. Olson SA (2002) EMBOSS opens up sequence analysis. European Molecular Biology Open Software Suite. Brief Bioinform 3: 87–91.
  50. Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. Genome Inform 23: 205–211.
  51. Finn RD, Mistry J, Tate J, Coghill P, Heger A, et al. (2010) The Pfam protein families database. Nucleic Acids Res 38: D211–222.
  52. Dutta S, H MB, W FB (2007) Using the tools and resources of the RCSB protein data bank. Curr Protoc Bioinformatics Chapter 1: Unit1 9.
  53. Frickey T, Lupas A (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics 20: 3702–3704.
  54. Martinez D, Larrondo LF, Putnam N, Gelpke MD, Huang K, et al. (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. Nat Biotechnol 22: 695–700.
  55. Stajich JE, Wilke SK, Ahren D, Au CH, Birren BW, et al. (2010) Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). Proc Natl Acad Sci U S A 107: 11889–11894.
  56. Martin F, Aerts A, Ahren D, Brun A, Danchin EG, et al. (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. Nature 452: 88–92.
  57. Plett JM, Martin F (2011) Blurred boundaries: lifestyle lessons from ectomycorrhizal fungal genomes. Trends Genet 27: 14–22.
  58. Chaturvedi V, Chaturvedi S (2011) *Cryptococcus gattii*: a resurgent fungal pathogen. Trends Microbiol.
  59. Cleveland TE, Yu J, Fedorova N, Bhatnagar D, Payne GA, et al. (2009) Potential of *Aspergillus flavus* genomics for applications in biotechnology. Trends Biotechnol 27: 151–157.
  60. Baker SE (2006) *Aspergillus niger* genomics: past, present and into the future. Med Mycol 44 Suppl 1: S17–21.
  61. Balajee SA (2009) *Aspergillus terreus* complex. Med Mycol 47 Suppl 1: S42–46.
  62. Sharpton TJ, Stajich JE, Rounsley SD, Gardner MJ, Wortman JR, et al. (2009) Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. Genome Res 19: 1722–1731.
  63. Pantou MP, Kouvelis VN, Typas MA (2008) The complete mitochondrial genome of *Fusarium oxysporum*: insights into fungal mitochondrial evolution. Gene 419: 7–15.
  64. Dorn B, Forrer HR, Jenny E, Wettstein FE, Bucheli TD, et al. (2011) *Fusarium* species complex and mycotoxins in grain maize from maize hybrid trials and from grower's fields. J Appl Microbiol 111: 693–706.
  65. Pelizza S, Stenglein S, Cabello M, Dinolfo M, Lange C (2011) First Record of *Fusarium verticillioides* as an Entomopathogenic Fungus of Grasshoppers. J Insect Sci 11: 70.
  66. Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, et al. (2009) The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. PLoS Genet 5: e1000618.
  67. Paoletti M, Saupé SJ (2008) The genome sequence of *Podospora anserina*, a classic model fungus. Genome Biol 9: 223.
  68. Anselem J, Cuomo CA, van Kan JA, Viaud M, Benito EP, et al. (2011) Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS Genet 7: e1002230.
  69. Reissingner A, Winter S, Steckelbroeck S, Hartung W, Sikora RA (2003) Infection of barley roots by *Chaetomium globosum*: evidence for a protective role of the exodermis. Mycol Res 107: 1094–1102.
  70. Yu J, Yang S, Zhao Y, Li R (2006) A case of subcutaneous phaeohiphomycosis caused by *Chaetomium globosum* and the sequences analysis of *C. globosum*. Med Mycol 44: 541–545.
  71. Oliver RP, Solomon PS (2010) New developments in pathogenicity and virulence of necrotrophs. Curr Opin Plant Biol.
  72. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, et al. (2008) Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. PLoS Genet 4: e1000046.
  73. Rokas A (2009) The effect of domestication on the fungal proteome. Trends Genet 25: 60–63.
  74. Turgeon BG, Baker SE (2007) Genetic and genomic dissection of the *Cochliobolus heterostrophus* Tox1 locus controlling biosynthesis of the polyketide virulence factor T-toxin. Adv Genet 57: 219–261.
  75. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, et al. (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat Biotechnol 26: 553–560.
  76. Goodwin SB, M'Barek SB, Dhillon B, Wittenberg AH, Crane CF, et al. (2011) Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensable structure, chromosome plasticity, and stealth pathogenesis. PLoS Genet 7: e1002070.
  77. Tian C, Beeson WT, Iavarone AT, Sun J, Marletta MA, et al. (2009) Systems analysis of plant cell wall degradation by the model filamentous fungus *Neurospora crassa*. Proc Natl Acad Sci U S A 106: 22157–22162.
  78. Desnos-Ollivier M, Ragon M, Robert V, Raoux D, Gantier JC, et al. (2008) *Debaryomyces hansenii* (*Candida famata*), a rare human fungal pathogen often misidentified as *Pichia guilliermondii* (*Candida guilliermondii*). J Clin Microbiol 46: 3237–3242.
  79. Blandin G, Ozier-Kalogeropoulos O, Wincker P, Artiguenave F, Dujon B (2000) Genomic exploration of the hemiascomycetous yeasts: 16. *Candida tropicalis*. FEBS Lett 487: 91–94.
  80. Favel A, Michel-Nguyen A, Peyron F, Martin C, Thomachot L, et al. (2003) Colony morphology switching of *Candida lusitanae* and acquisition of multidrug resistance during treatment of a renal infection in a newborn: case report and review of the literature. Diagn Microbiol Infect Dis 47: 331–339.
  81. Lockhart SR, Messer SA, Pfaller MA, Diekema DJ (2008) *Lodderomyces elongisporus* masquerading as *Candida parapsilosis* as a cause of bloodstream infections. J Clin Microbiol 46: 374–376.
  82. Jeffries TW, Grigoriev IV, Grimwood J, Laplaza JM, Aerts A, et al. (2007) Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. Nat Biotechnol 25: 319–326.
  83. Bruno VM, Wang Z, Marjani SL, Euskirchen GM, Martin J, et al. (2010) Comprehensive annotation of the transcriptome of the human fungal pathogen *Candida albicans* using RNA-seq. Genome Res 20: 1451–1458.
  84. Bruckner S, Mosch HU (2011) Choosing the right lifestyle: adhesion and development in *Saccharomyces cerevisiae*. FEMS Microbiol Rev.
  85. Replansky T, Koufopanou V, Greig D, Bell G (2008) *Saccharomyces sensu stricto* as a model system for evolution and ecology. Trends Ecol Evol 23: 494–501.
  86. Yukawa M, Maki T (1931) *Schizosaccharomyces japonicus* nov. spec. La Bul. Sci. Falkultato Terkultura. Kjusu Imp Univ, Fukuoka, Japan 4: 218–226.
  87. Northam BE, Norris FW (1951) Growth requirements of *Schizosaccharomyces octosporus*, a yeast exacting towards adenine. J Gen Microbiol 5: 502–507.
  88. Lannoo MJ, Petersen C, Lovich RE, Nanjappa P, Phillips C, et al. (2011) Do frogs get their kicks on Route 66? Continental U.S. transect reveals spatial and temporal patterns of *Batrachochytrium dendrobatidis* infection. PLoS One 6: e22211.