



HAL
open science

Characterization of transposable elements in the ectomycorrhizal fungus *Laccaria bicolor*

Jessy Labbé, Claude Murat, Emmanuelle Morin, Gerald A. Tuskan, François F. Le Tacon, Francis Martin

► **To cite this version:**

Jessy Labbé, Claude Murat, Emmanuelle Morin, Gerald A. Tuskan, François F. Le Tacon, et al.. Characterization of transposable elements in the ectomycorrhizal fungus *Laccaria bicolor*. PLoS ONE, 2012, 7 (8), <10.1371/journal.pone.0040197>. <hal-02650918>

HAL Id: hal-02650918

<https://hal.inrae.fr/hal-02650918v1>

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Characterization of Transposable Elements in the Ectomycorrhizal Fungus *Laccaria bicolor*

Jessy Labbé^{1,2*}, Claude Murat¹, Emmanuelle Morin¹, Gerald A. Tuskan², François Le Tacon¹, Francis Martin¹

¹ Unité Mixte de Recherche de l'Institut National de la Recherche Agronomique – Lorraine Université 'Interactions Arbres/Microorganismes', Centre de Nancy – Champenoux, France, ² BioSciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States of America

Abstract

Background: The publicly available *Laccaria bicolor* genome sequence has provided a considerable genomic resource allowing systematic identification of transposable elements (TEs) in this symbiotic ectomycorrhizal fungus. Using a TE-specific annotation pipeline we have characterized and analyzed TEs in the *L. bicolor* S238N-H82 genome.

Methodology/Principal Findings: TEs occupy 24% of the 60 Mb *L. bicolor* genome and represent 25,787 full-length and partial copy elements distributed within 171 families. The most abundant elements were the *Copia*-like. TEs are not randomly distributed across the genome, but are tightly nested or clustered. The majority of TEs exhibits signs of ancient transposition except some intact copies of terminal inverted repeats (TIRS), long terminal repeats (LTRs) and a large retrotransposon derivative (*LARD*) element. There were three main periods of TE expansion in *L. bicolor*: the first from 57 to 10 Mya, the second from 5 to 1 Mya and the most recent from 0.5 Mya ago until now. LTR retrotransposons are closely related to retrotransposons found in another basidiomycete, *Coprinosps cinerea*.

Conclusions: This analysis 1) represents an initial characterization of TEs in the *L. bicolor* genome, 2) contributes to improve genome annotation and a greater understanding of the role TEs played in genome organization and evolution and 3) provides a valuable resource for future research on the genome evolution within the *Laccaria* genus.

Citation: Labbé J, Murat C, Morin E, Tuskan GA, Le Tacon F, et al. (2012) Characterization of Transposable Elements in the Ectomycorrhizal Fungus *Laccaria bicolor*. PLoS ONE 7(8): e40197. doi:10.1371/journal.pone.0040197

Editor: Jason E. Stajich, University of California Riverside, United States of America

Received: December 6, 2011; **Accepted:** June 4, 2012; **Published:** August 3, 2012

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: This work was supported by INRA, the Region Lorraine Research Council and by the Plant-Microbe Interfaces Scientific Focus Area project at Oak Ridge National Laboratory (ORNL) sponsored by Office of Biological and Environmental Research at the United States Department of Energy Office of Science. ORNL is managed by UT-Battelle, LLC, under contract DE-AC05-00OR22725 for the United States Department of Energy. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: labbejj@ornl.gov

Introduction

The basidiomycete *Laccaria bicolor* (Maire) P.D. Orton is an ectomycorrhizal fungus living in symbiosis with various woody host plants. Symbiotic plant-fungus associations play a fundamental role in biology and ecology of forest trees, affecting growth, water and nutrient absorption, and providing protection from root diseases [1]. *L. bicolor* is an early stage ectomycorrhizal symbiont able to establish itself on seedlings, but it also colonizes roots of adult conifer and hardwood trees. *L. bicolor* is characterized by its ubiquitous geographical distribution and lack of strict host specificity. This habit has been associated with its relatively large genome, containing the largest fungal gene repertoire and an abundance of multiple gene families relative to other ectomycorrhizal fungi, which in turn facilitates its adaptation to and interaction with various hosts [2]. The whole-genome sequence (WGS) of *L. bicolor* S238N-H82 provided the first ectomycorrhizal symbiont blueprint, fostering new insights into the nature of the genome of one of the largest groups of fungi [2]. At 60 Mb, the draft genome of *L. bicolor* is larger than that of previously published fungal genomes. The size is partly explained by unprecedented

transposon diversity – more types than any other fungi studied to date.

Transposable elements (TEs) are short, mobile, conserved segments of DNA that can replicate and randomly insert copies within genomes. They have had an important influence on the evolution of eukaryote genomes [3]. These elements often constitute a large proportion of eukaryotic genomes, e.g., ~45% of the human genome [4] and 50 to 80% of some grass genomes [5], and yet alternatively only 9.7% of the fungal *Magnaporthe grisea* genome [6]. Although TEs were first identified in fungi in the yeast *Saccharomyces cerevisiae* [7], conventional genetic studies with *Ascobolus immersus* mutants established their existence in filamentous fungi [8]. *Ty1-Copia* retroelements have been reported in three ectomycorrhizal basidiomycetes, *L. bicolor*, *Pisolithus macrocarpus* and *Pisolithus* MH155 by Diez *et al.* [9].

TEs are also known to affect genome organization; recombination between elements at different sites can lead to large-scale chromosomal rearrangements. Some TE insertions may become fixed and play a role in gene function [10], and indeed, TE insertions next to genes can modify gene expression patterns while insertions within genes can negatively impact transcription and

gene function. In general, proliferation of TEs induces potentially deleterious effects. To avoid such effects eukaryotes have developed mechanisms that limit TE activity. Paradoxically, TE existence depends on the survival of eukaryote genome. It has been reported that TEs have developed silencing mechanisms directing their integration to specific parts of the genome with minimal damage [11]. Thus, there is no doubt that TEs have evolved in close association with eukaryote genomic DNA.

Eukaryotic TEs are divided into two classes, depending on their mode of transposition: Class I elements or retroelements or also retrotransposons, which mobilize via a ‘copy-and-paste’ mechanism that uses an ribonucleic acid (RNA) intermediate [12], and class II elements or deoxyribonucleic acid (DNA) transposons, which mobilize via a ‘cut-and-paste’ mechanism that use a DNA intermediate. These two classes are composed of five major types: long terminal repeat (LTR) retrotransposons, non-LTR retrotransposons, cut-and-paste DNA transposons, rolling-circle DNA and self-synthesizing DNA transposons. Each type of TE is composed of a number of superfamilies or clades based on length and target site features, with each superfamily consisting of numerous families. The DNA transposons (class II elements) have terminal inverted repeats (TIRs) or a rolling circle replicon mechanism, e.g., *Helitrons* elements, similar to some known prokaryotic transposition mechanism or self-synthesizing DNA transposons (*Polintons*). The retrotransposons (class I elements) are the most common TE in fungi [13]. As noted above retrotransposons can be classified into two types – LTR retrotransposons and non-LTR retrotransposons (encompassing LINE elements), depending whether they possess or lack long terminal repeats (LTRs) at both ends [13], tyrosine recombinase retroelements (YR; subdivided in three families, *DIRS*, *Ngaro* and

VIPER), Penelope-like retrotransposable elements, and short interspersed nuclear elements (SINEs) [14]. The LTR retrotransposons, which have a long terminal repeat at their extremities, have been divided into superfamilies: vertebrate retroviruses (Retroviridae), hepadnaviruses, caulimoviruses, *Ty1-Copia-like* (Pseudoviridae), *Ty3-Gypsy-like* (Metaviridae) and *Pao-BEL-like* depending on their sequence similarity and the type of gene products they encode. The two main superfamilies of LTR retrotransposons found in fungi [13] are *Gypsy* and *Copia*, which differ in the order of reverse transcriptase (RT), ribonuclease H (RH), and integrase (IN) domains in the virus-like polyprotein (POL; *Gypsy*: PR-RT-RH-INT, *Copia*: PR-INT-RT-RH).

Fungi typically have small genomes in the 10 to 40 Mb range, usually with limited amounts of repetitive DNA [15] and thus the Ascomycota and Basidiomycota appear to have a tendency towards streamlined genomes. The majority of these taxa contain no more than 10–15% repetitive DNA [16,17]. However, the haploid genome of *L. bicolor* comprised a high proportion of repetitive elements relative to other fungal genomes [2].

As TEs are mobile, profuse, variable and dispersed throughout the genome, they can cause serious difficulties in sequence assembly. To assemble and annotate the *L. bicolor* genome it was necessary to determine the abundance and distribution of TEs. Here, we report the automated annotation and manual curation of TEs in *L. bicolor* genome (v2.0). We investigate the diversity and distribution for all TE types and temporal feature for the LTR retrotransposons. In addition, we compare LTR retrotransposons spatial distribution and insertion age of full-length LTR retrotransposons in *L. bicolor*, with the taxonomically related saprophytic Agaricales *Coprinopsis cinerea*.

Table 1. Transposable elements characterized in *Laccaria bicolor* genome.

TE type	Family	Copy number	Full length/incomplete	DNA amount (Mb)	TE composition in genome (%)
Class I					
LTR	78	6611	110/6501	3.99	6.65
<i>ERV</i>	1	7	4/3	0.03	0.05
<i>Copia</i> -like	46	3382	57/3325	2.23	3.72
<i>Gypsy</i> -like	27	3085	47/3038	1.71	2.85
<i>LARD</i>	3	118	1/117	0.05	0.08
<i>DIRS</i>	1	19	1/18	0.01	0.01
Non-LTR	9	1273	23/1250	2.21	3.68
<i>Tad</i>	2	22	17/5	1.6	2.60
<i>Deceiver</i>	1	1	1/0	0.02	0.03
<i>L1</i>	1	1	0/1	0.01	0.02
<i>LINEs</i>	5	1249	5/1244	0.58	0.96
Total Class I	87	7884	133/7751	6.20	10.33
Class II					
TIRs	23	6369	692/5677	2.85	4.75
<i>Helitrons</i>	1	215	2/213	0.08	0.13
<i>MITEs</i>	4	249	103/146	0.06	0.10
Total Class II	28	6833	797/6036	2.99	4.98
TE fragments ^a	56	11070	–	5.33	8.88
Total TE	175	25787	930/13787	14.52	24.2

^aOthers elements truncated, unclassified, unfound matches and without category (named no category).

doi:10.1371/journal.pone.0040197.t001

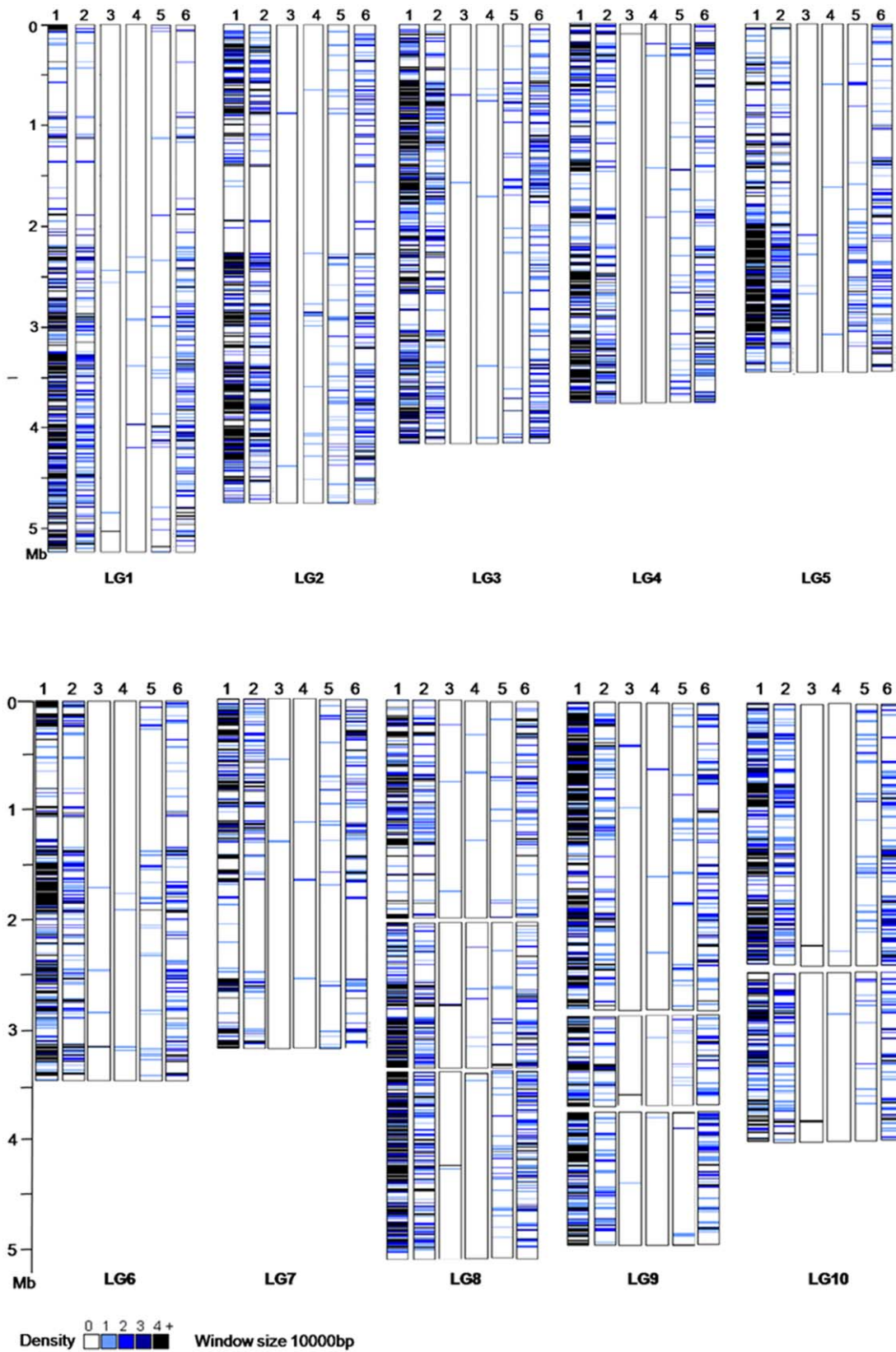


Figure 1. Density of transposable elements (TE) along the linkage groups of *L. bicolor*. The density is graphically represented as the number of TE copy found in a window of 10 kb. Lane 1 represents the density of all TEs; Lane 2 represents the density of *Gypsy* and *Copia*; Lane 3 represents the density of *Helitron*; Lane 4 represents the density of *LARD*; Lane 5 represents the density of *LINE*; Lane 6 represents the density of *TIRs*. doi:10.1371/journal.pone.0040197.g001

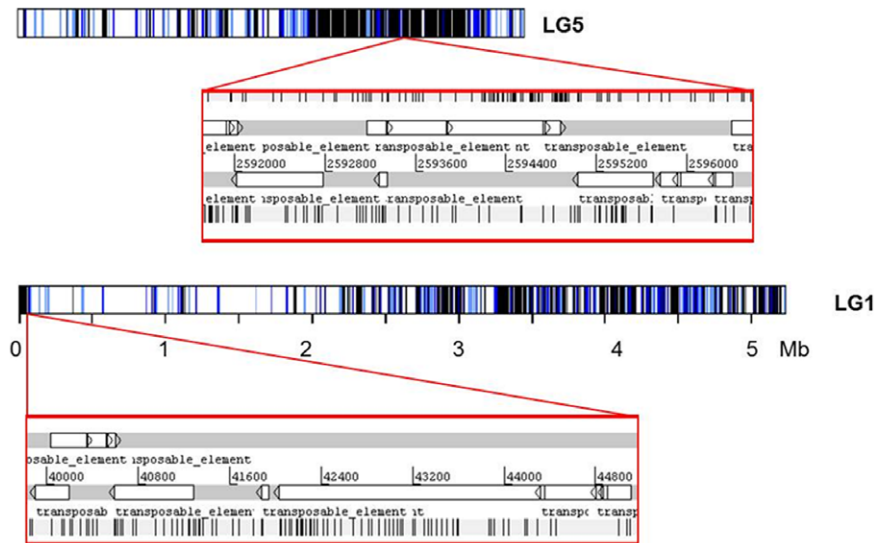


Figure 2. Artemis v11 depiction showing two samples of the nested and clustered distribution of transposable elements (TE) on the *L. bicolor* LG 5 and LG 1.

doi:10.1371/journal.pone.0040197.g002

Materials and Methods

Sequences analyses were performed using the whole-genome sequence (WGS) version 2 of the homokaryotic (haploid) strain S238N-H82 from *L. bicolor* (Agaricomycetes, Agaricales, Hydnangiaceae) downloaded from the Joint Genome Institute (<http://genome.jgi-psf.org/Lacbi2/Lacbi2.download.html>). The WGS of *Coprinopsis cinerea* (Agaricomycetes, Agaricales, Psathyrellaceae) was downloaded from the Broad Institute (<http://www.broadinstitute.org/>). The known TE reference set was constructed from 1) RepeatMasker repeat library (<http://www.repeatmasker.org>), 2) RepBase on the Genetic Information Research Institute website (GIRI) (<http://www.girinst.org/replib/index.html>) and 3) sequences from NCBI (<http://www.ncbi.nlm.nih.gov/>).

Sequence annotation

TEs were annotated in the *L. bicolor* genome using the RMBLR procedure from the TE annotation pipeline described by Quesneville *et al.* [18]. Consecutive fragments on both the genome and the reference TE were automatically joined if they were separated by a sequence of which more than 80% consisted of other TE insertions. Simple repeats were found using the Tandem Repeat Finder program [19] and used to filter out spurious hits. All TE annotations that were less than 20 bp, after removing any regions that overlapped simple repeat regions, were eliminated. Finally, 171 consensus sequences (complete and incomplete) belonging to various classes or types of TEs were obtained. LTR retrotransposons sequences were confirmed using a second identification procedure based on the program LTR_STRUC [20]. To facilitate manual curation we promoted the 171 consensus sequences identified via RMBLR to a candidate annotation set (defined as a set of one or more joined fragments) and by using sequence alignments with CLUSTALW (<http://www.ebi.ac.uk/clustalw>) that were then validated or modified by the curator in Artemis v11. The resulting database of the annotated TEs of *L. bicolor* is available at <http://mycor.nancy.inra.fr/IMGC/LaccariaGenome/download.php?select=anno>.

TE copy number estimation and distribution

The copy number for each TE type was calculated on the basis of the elements obtained from the TE annotation pipeline, TBLASTN searches and RepeatMasker analysis used in the previous pipeline [2]. The TE distribution in *L. bicolor* assembly was determined via RepeatMasker analysis using our curated and annotated *L. bicolor* TE database. Python scripts were used for the mapping of TE density on the assembly anchored on the updated genetic linkage groups (LG) from the genetic map of *L. bicolor* [21]. Multiple sequence alignments were constructed using CLUSTALX [22] to generate nucleotide sequence similarity between LTR sequences of LTR retrotransposon full-length copies. The sequence divergence (in %) was then used to estimate the relative age of TEs.

Sequence alignments and comparison

To study the phylogenetic relationship between the major elements form in *L. bicolor* and *C. cinerea*, the reverse transcriptase (RT) and the ribonuclease H (RH) amino acid sequences were aligned for all full-length, manually curated LTR elements. Multiple sequence alignments were constructed with CLUSTALX [22]. Phylogenetic trees were constructed on the basis of the neighbor-joining method using PHYLIP with default parameters [23,24]. Bootstrap values were estimated for each tree from 1000 replicates. The phylogram graphics were created using FigTree v1.2.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

The insertion age of full-length LTR retrotransposons was determined by comparing their 5' and 3' LTR sequences and based on the assumption that the two LTRs of a single element are identical at the time of the insertion. The two LTRs were aligned with CLUSTALX software [22] and the number of transition and transversion mutation were counted. The insertion times were dated using the Kimura two-parameter method (emboss distmat, <http://150.185.138.86/cgi-bin/emboss/distmat>). The time (T) since element insertion for each family of retroelements (e.g., LTR *Gypsy* family 1 vs. LTR *Gypsy* family 2 found in *L. bicolor*) was estimated using the formula $T = K2P/2r$, where K = evolutionary distance per site, P = probability of transition and r = probability of identity at homologous sites [25]. For the conversion of

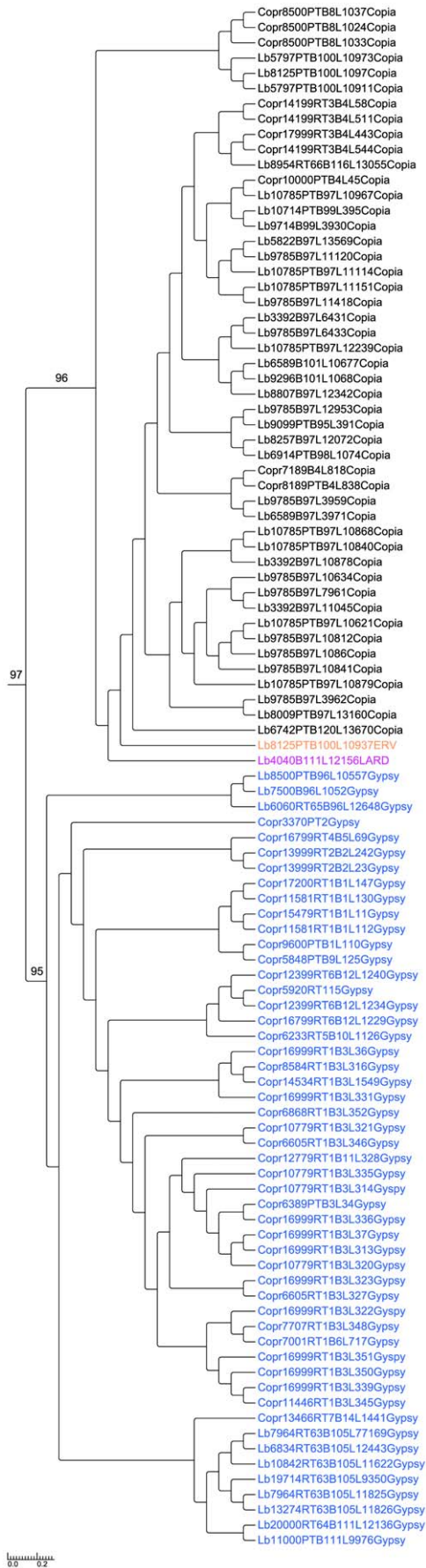


Figure 3. Relationships between LTR retrotransposons of *Laccaria bicolor* and *Coprinopsis cinerea*. This phylogenetic tree is based on reverse transcriptase (RT) and ribonuclease H (RH) amino-acid sequences of all the LTR retrotransposons identified by LTR_STRUC and after manual curation. In black: clade 1, in blue: clade 2. *Gypsy* elements are designated by blue font, *Copia* by black, *LARD* by magenta and *ERV* by orange.
doi:10.1371/journal.pone.0040197.g003

sequence distance to insertion age, an average substitution rate of $1.3E^{-8}$ mutations per site per year was used based on estimates by Ma and Bennetzen [26] and considered close enough to the molecular clock of LTR retrotransposons in fungi [27].

Results

TE abundance, diversity and distribution

Approximately 15 Mb (24%) of the 60 Mb of *L. bicolor* genomic assembly sequence is comprised of TEs (<http://mycor.nancy.inra.fr/fr/index.html>). Among the 15 Mb, TEs of class I represent 10% of the genome (~6 Mb) and TEs of class II represent 5% (~3 Mb). The remaining 5 Mb (9% of the genome) were truncated or did not match classified TEs. Consensus sequences obtained by the TE RMBLR annotation pipeline, used as queries in BLASTN searches of *L. bicolor* genomic sequence, did not contain any additional repeats, indicating that the vast majority of TEs had been identified. The results of these analyses are summarized in Table 1 and presented in detail in Table S1.

From the 171 consensus sequences obtained by the RMBLR procedure, excluding 56 TE fragments (each fragment representing a potential TE family; 11,070 copies), we identified 115 families of TEs in *L. bicolor* containing 930 full-length copies and 13,787 incomplete copies. The major were in the class I (87 families) cluster, in which the *Copia* families were the most numerous (46 members), followed by *Gypsy* (27 members). The third most abundant group was represented by the TIRs (class II) with 23 members. We also identified two *Tad*, one *Deceiver*, one *LI*, five *LINEs*, four *MITEs*, three large retrotransposon derivatives (*LARD*), one *ERV*, one *DIRS*, and one *Helitron*. The *Helitron*, as well as the *MITEs* and *LARD*, were primarily found in gene-rich regions coding for signal transducers, nucleotide binding proteins, proteins G and RNA binding proteins (data not shown). An intact copy of *LARD* was also located less than 1 kb from the 3' end of the *B* mating type locus following by an intact *MITE* [28].

The identification of thousands of TEs in the *L. bicolor* genome provided the informatics resource to address questions about TE diversity and temporal aspect of TE amplification. We compared all TE types and families previously identified by multiple alignments using CLUSTALX [22]. According to sequence diversity metrics, most of the TE types and families originate in old transposition events except for the intact copies of four TIRs, two *MITEs*, three LTR retrotransposons and one *LARD*, which represent more recent events and in total encompass 5% of the genome (Table S1). These elements contained nearly identical members (threshold >95% identity), suggesting they are potentially still active.

Along the 10 *L. bicolor* pseudochromosomes (designated LG1 to 10), the average density of TEs across the whole-genome was 430 copies per Mb. TE density was not correlated with the chromosome size ($r = 0.2$), though TE density was higher in the unmapped small scaffolds, likely reflecting the difficulty of unambiguously aligning and assembling contigs with high repeat content. To determine whether TEs might display non-uniform accumulation across the genome, we analyzed their distribution within a sliding 10 kb intervals along the anchored genetic-

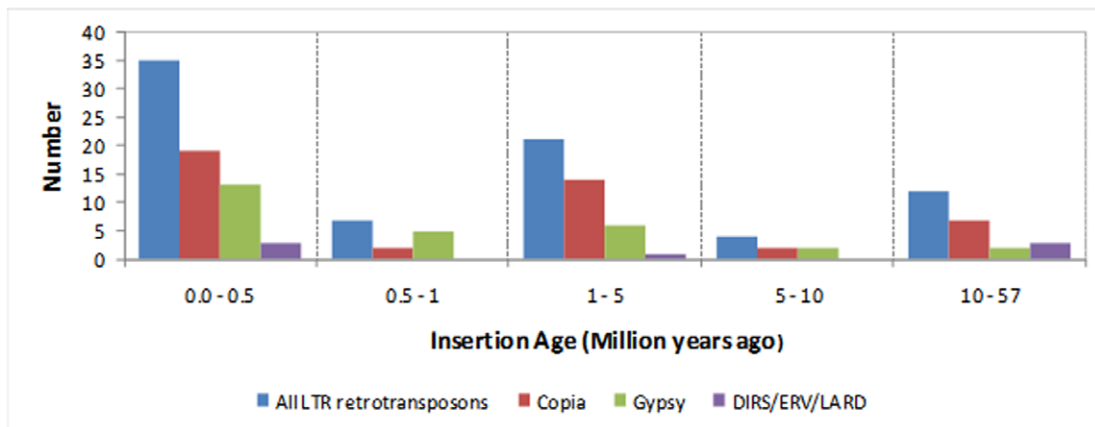
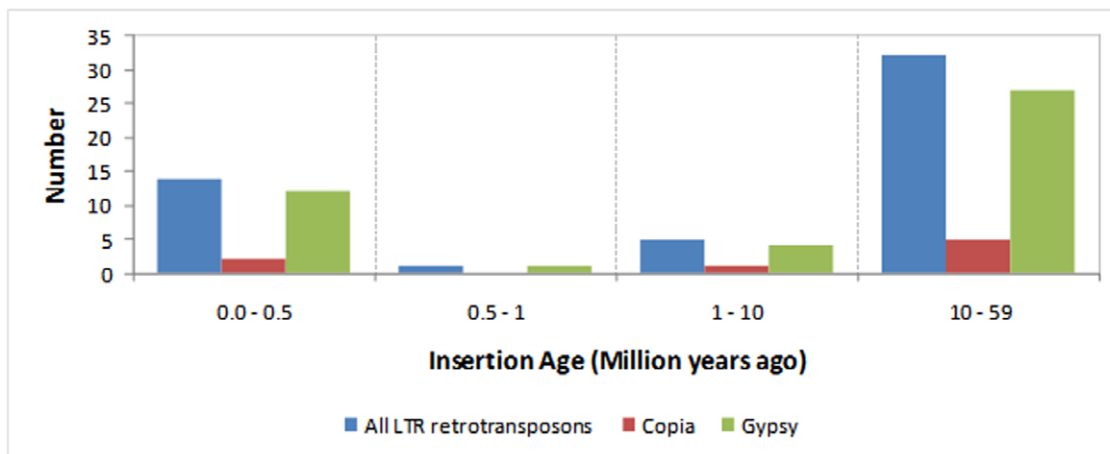
(a) the *Laccaria bicolor* genome(b) the *Coprinopsis cinerea* genome

Figure 4. Number of LTR retrotransposon families relative to their insertion age expressed in million years.
doi:10.1371/journal.pone.0040197.g004

physical LGs of *L. bicolor* representing 47.7 Mb of the total 60 Mb. Due to their small length, the unmapped scaffolds were not included in this analysis. As shown in Figure 1, TEs were not uniformly distributed within or among LGs. Similar results were obtained within 100 kb intervals (data not shown). Each LG contained regions exhibiting high TE density, mainly *Copia*, *Gypsy* and TIR elements. In general, these regions corresponded to the ends of the LGs and were found most abundantly in the telomeric regions of LG2, LG4, LG6 and LG9.

By convention [29,30], TEs inserted within another transposable element are defined as nests, and groups of TEs located within 10 kb of each other are defined as clusters. We found 382 nests and/or clusters of TEs containing 138 families (e.g., Figure 2), suggesting that about 80% of TEs of *L. bicolor* were either inserted into another element and/or were positioned adjacently to another element. Among the 382 nests or clusters, 153 were within the terminal 1 Mb region of the LGs. Most of TEs belonging to nests or clusters were incomplete. LTR retrotransposons *Gypsy* and *Copia* (34%) were nested or clustered more often than either TIRs (25%) or LINE (16%), which is roughly proportional to their relative abundance within the genome. As previously reported [2], this nested structure was predominant in

weakly recombinant telomeric regions, where TE insertion may produce rearrangements without deleterious effects.

Phylogenetic analysis of the LTR retrotransposons

Given that LTR retrotransposons are the most frequent TEs in the *L. bicolor* genome (*Copia* > *Gypsy*), we analyzed the relationships between these elements and those of *C. cinerea* where LTR retrotransposons are also over-represented (*Gypsy* > *Copia*). We constructed phylogenetic trees based on multiple alignments of reverse transcriptase (RT) and ribonuclease H (RH) domains including all identified families of LTR retrotransposons on both genome sequences using LTR_STRUC (Figure 3) [20]. *L. bicolor* and *C. cinerea* TEs generally exhibited the same phylogenetic structure each containing two independent monophyletic clades each supported with bootstrap values >90. In *L. bicolor*, as well as in *C. cinerea*, clade 1 is mainly comprised of *Copia* elements, while clade 2 is comprised of *Gypsy* (Figure 3). The RT and RH sequences of most of *Gypsy* and *Copia* families appeared to be closely related to each other in both genomes (Figure 3). Interestingly, the *ERV* element appears in separate clade, suggesting independent origins of this TE (Figure 3).

The insertion age of LTR retrotransposons

As shown by Figure 4A and 4B, the oldest detected LTR retrotransposon insertion occurred in the genome of the ancestor of *L. bicolor* and *C. cinerea* 57 and 59 million years ago (Mya). In *L. bicolor*, there were three main periods of expansion of TE families: the first between 57 and 10 Mya, the second between 5 and 1 Mya and the last from 500,000 years ago until present. In *C. cinerea*, there were only two main periods of TE expansion: the first between 59 and 10 Mya and a second from 0.5 Mya ago until now. In *L. bicolor*, during the first two periods *Copia* expansion was more abundant than *Gypsy*, but between 0.5 to 1 Mya *Gypsy* expansion was more abundant than *Copia*. In contrast, in *C. cinerea* during the two main periods of TE expansion, *Gypsy* expansion was consistently more abundant than that for *Copia*.

Discussion

The *L. bicolor* genome is rich in TEs (24% of the genome), which partially contributes to its relative large genome size compared to other Basidiomycota. Indeed a clear connection is emerging from genome comparisons: the larger the genome, the higher the number of TEs [31]. Analogous considerations can be extended from the 60 Mb of *L. bicolor* to the 140 Mb of the Ascomycota *Tuber melanosporum* (58% of TEs) [32].

The largest class of TEs highlighted in this study was the class I elements, with a high proportion of LTR retrotransposons (*Copia* > *Gypsy*), which confirms Muszewska *et al.* observations [33]. Moreover, we identified additional TE types such as *LINE*, *ERV* and *LARD* and we established the presence of *Tad*, *Deceiver* and *L1* non-LTR elements previously suggested in *L. bicolor* by Novikova *et al.* [34]. A large proportion of class II elements also were found, mainly *TIRs*, *Helitron* and *MITEs*. In *L. bicolor* the single *Helitron* and its related copies account for about 0.13% of the genome and are mainly found in comparatively gene-rich regions, as has been reported in maize [31], while conversely they appear to be preferentially localized in the gene-poor region in *Arabidopsis* and rice genomes [35,36]. In either case, the characteristics of *Helitrons* make them one of the primary sources of DNA variation [35]. Despite the small size and low proportion of *Helitrons* in the *L. bicolor* genome, its presence in gene-rich regions suggests this element may be involved in introducing sequence variation within coding regions in this genome.

MITEs, among the smallest TEs, and *LARDs* have been more commonly associated with genes than are other types of TEs in plants [37,38,39,40]. In *L. bicolor* *MITEs* and *LARDs* represent 0.10% and 0.08% of the genome, respectively, and were found among gene-rich regions. This gene-TE association may have implications for gene regulation, mRNA stability or conservation of expression of loci in large multigenic families [12], and as has been related to R genes in rice [41]. The occurrence of intact, i.e., recent, *LARD* and *MITE* in the 3' end of the *L. bicolor* *B*-mating type locus suggests a possible role of this element in the rapid evolution of the *B*-mating type containing region. Interestingly, Niculita-Hirzel *et al.* [28] observed a similar distribution of the transposon insertion around the *B*-containing region and not around the *A*-mating type.

Among the 171 TE families, 43 individual elements exhibited intact structure, suggesting that these elements are still active. This hypothesis is supported by expression evidence corresponding to their transcripts detected in free-living mycelium, fruiting bodies and ectomycorrhizas [2]. However, it has been proposed that fungal cells may control transposition at a post-transcriptional level.

Phylogenetic analysis of the LTR retrotransposons revealed that these elements are closely related amongst themselves and to retrotransposons found in another Agaricales, i.e., *C. cinerea*, as Muszewska *et al.* [33] has shown. These elements might descend from a common ancestor of *L. bicolor* and *C. cinerea* that split from their common ancestor ~90 Mya ago (D. Hibbett, personal communication). Our estimation of the insertion age of LTR retrotransposons suggests that *L. bicolor* ancestors experienced three main bursts of LTR retrotransposon expansion, including a recent one. *C. cinerea* ancestor had only two main periods of expansion, which could explain the higher frequency of LTRs observed in *L. bicolor*. Estimation of insertion age, as well as phylogenetic tree, revealed a recent expansion of *Copia* elements in *L. bicolor*, while the *Gypsy* elements expansion appeared to be more ancient. The first main period of TE in *L. bicolor* ancestor dates from 57 Mya. According to Ryberg and Matheny [42], Hydnangiaceae appeared before 40 Mya, much before *L. bicolor* evolved. The *L. bicolor* and *L. nobilis* species appears to have evolved from *L. laccata* 12–14 Mya [42,43], supporting our contention that the second LTR retrotransposon expansion might have coincided with the species divergence.

TEs are not randomly distributed across the *L. bicolor* genome, but are tightly nested or clustered. The analysis of the TE spatial distribution on the LGs revealed that the nests and/or clusters are mostly positioned at the proximal ends in the telomeric regions. The majority of TEs belonging to nests or clusters are mainly LTR retrotransposons (*Copia* and *Gypsy*), and to a lesser degree, *TIRs* and *LINE*. There are several mechanisms that may explain this type of distribution. The ectopic exchange model states that repetitive DNA promotes deleterious ectopic recombination events [44]. Since selection pressure would be reduced in regions of low recombination, repetitive DNA should accumulate in such regions. Under the insertion model, TE insertions generally have a deleterious effect on fitness and are thus likely to be lost from the population [12]. That is, in chromosomal segments that have low recombination rates, deleterious mutations are more likely to be genetically linked to neutral genes, decreasing their negative fitness effect. The insertion of TE may trigger a runaway process, where there would be a cascade displacement of TEs, as each TE would provide a target into which other elements could be insert without deleterious consequences [45].

The insertion model is supported by the observation that TE occurrence was highest in regions of the genome with high microsatellite density [46]. It has been suggested that TE-associated microsatellites (SSRs) are either components of active TEs spreading throughout the genome, acting as a 'landing pad' for TE insertion or they arise after the integration of an extended and polyadenylated retro-transcript into the genome [47,48]. Arcot *et al.* [49], Ellegren [50] and Zhu *et al.* [51] have shown that *LINEs*, *LTRs* and *MITEs* were able to generate and propagate microsatellites. Thus, the distribution of TEs in clusters could result from potential insertion sites formed by microsatellites flanking these TEs that may subsequently lead to accumulation of chromosome deletions through unequal homologous recombination or interchromosomal recombination [52,53].

As in all plant, animal and fungal genomes characterized to date, the *L. bicolor* and *C. cinerea* genomes have a distinctive TE composition and comparative evolutionary history. The high proportion, the large diversity, and the nested and clustered distribution of TE in *L. bicolor* suggest rearrangements promoted by their occurrence. This nested and clustered distribution is associated with SSR occurrence and could be the result of potential insertion sites formed by the SSRs flanking the TEs [54,47]. The great diversity of TEs in *L. bicolor* genome can be

partially explained by the several periods of expansion of TEs but may also be the result of potential horizontal transfer. Horizontal transfer is the exchange of genetic information between reproductively isolated organisms and has been shown to be of significance in the evolution of some functional fungal genes [55,56,57]. Diversity of TEs and their copy number depends on the evolutionary history of a particular species or a cluster of closely related species, their population structure and ecological features. There are several main processes which could affect the copy number and diversity of TEs in fungal genomes: stochastic loss of elements, burst of transposition, the limitation of copy number increase by natural selection which removes deleterious insertions, passive and active inactivation of repetitive sequences, and self-regulation of transposition (decrease of the transposition rate when the copy number increases) [58]. The population structure and dynamics, as well as mating mode, also play an important role in the transposable elements evolution. The inactivation of repeated sequences is also a very important factor, which leads to the shifts in diversity and copy number of TEs, especially in fungi. Sequencing of additional strains of *L. bicolor* and *Laccaria* species should provide the needed information to investigate the mechanisms in action in *Laccaria*.

TEs are distributed throughout the *L. bicolor* genome and could be major contributors to the genesis of new genes or to the adaptation of existing genes, notably via mechanisms such as molecular domestication, ectopic recombination and gene retrotransposition. Molecular domestication, also known as the process of TE recruitment by the host genome, is the coopted use by the organism of a function carried by a TE. Because TEs encode proteins that can for example bind, copy, break, join or degrade nucleic acids, they have been repeatedly domesticated during eukaryotic evolution [59]. As another mechanism, retrotransposon-mediated ectopic recombination results from the physical occurrence of retrotransposon insertions at particular sites in the genome and can imply various genomic rearrangements, such as duplications, deletions and translocations. Gene retrotransposition is also another mechanism that can rearrange genes. Gene retrotransposition operates during the retrotransposition process itself and only duplicates gene sequences but no retrotransposon sequence. Thus, movement of TEs increasing their density in the *L. bicolor* genome would support the theory that the higher than normal expansion of certain gene families in this genome. Notably, this may have created paralogous genes such as hydrophobin-coding-genes [60], important in host-fungal interactions and giving then a selective advantage to this mutualistic fungus in colonizing a wide range of host plants.

References

- Harley JL, Smith SE (1983) Mycorrhizal Symbiosis. Academic Press, London.
- Martin F, Aerts A, Ahrén D, Brun A, Duchaussoy F, et al. (2008) Symbiosis insights from the genome of the mycorrhizal basidiomycete *Laccaria bicolor*. *Nature* 452: 88–92.
- Kidwell MG, Lisch DR (2000) Transposable elements and host genome evolution. *Trends Ecol. Evolut.* 15: 95–99.
- Lander EM, Linton LM, Birren B, Nusbaum C, Zody MC (2001) Initial sequencing and analysis of human genome. *Nature* 409: 860–921.
- Meyers BC, Tingey SV, Morgante M (2001) Abundance, distribution and transcriptional activity of repetitive elements in the maize genome. *Genome Res.* 11: 1660–1676.
- Deane RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchel TK (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434: 980–986.
- Boeke JD (1989) Transposable elements in *Saccharomyces cerevisiae*. In: Berg DE and M Howe, editors. *Mobile DNA*. ASM, Washington DC. 335–374.
- Decaris B, Francou F, Lefort C, Rizet G (1978) Unstable aeciospore color mutants of *Ascobolus immersus*. *Mol. Gen. Genet.* 162: 69–81.
- Diez J, Beguristain T, Le Tacon F, Casacuberta JM, Tagu D (2003) Identification of Ty1- *copia* retrotransposons in three ectomycorrhizal basidiomycetes: evolutionary relationships and use as molecular markers. *Curr. Genet.* 43: 34–44.
- Kidwell MG (2002) Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115: 49–63.
- Sandmeyer S (1998) Targeting transposition: at home in the genome. *Genome Res.* 8: 416–418.
- Feschotte C, Jiang N, Wessler S (2002) Plant transposable elements: where genetics meets genomics. *Nature Rev. Genet.* 3: 329–341.
- Boeke JD, Stoye JP (1997) Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In: retroviruses, coffins JM, Hughes SH, Varmus HE, editors. Cold Spring Harbor Laboratory Press, New York. 343–435.
- Deninger PL (1989) SINEs: Short interspersed repeated DNA elements in higher eukaryotes. *ASM* 19: 619–637.
- Baker SE, Thykaer J, Adney WS, Brettin T, Brockman FJ, et al. (2008) Fungal genome sequencing and bioenergy. *Fungal Biology Reviews* 22:1–5.
- Kempken F, Jacobsen S, Kück U (1998) Distribution of the Fungal Transposon *Restless*: Full-Length and Truncated Copies in Closely Related Strains. *FGB* 25(2): 110–118.
- Wöstemeyer J, Kreibich A (2002) Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. *Curr. Genet.* 41: 189–198.

Finally, the high density of TEs in *L. bicolor* genome could be associated with its host plasticity and its evolution from a saprophytic to a symbiotic habit or with TE proliferation after the split between the saprotrophic ancestor and ectomycorrhizal ancestor of *Laccaria*. Each of the above molecular mechanisms appears to have driven the genome expansion in *L. bicolor*, conferring increased genome reorganization and possibly leading to genome divergence among related strains. Genome plasticity and high variability among strains suggest that a single reference genome for each species will be insufficient to completely describe and understand the genetic complement of a particular species [54]. Characterizing interspecific differences in TEs among various *Laccaria* species and isolates will be aided by the present study. A *Laccaria* Pan-Genome project is underway which will sequence the genomes of 12 *L. bicolor* strains from geographically diverse origins as well as four related *Laccaria* species. The assignment of TE activity (*Helitrons* insertions, and recombination events) inside the intergenic and intronic regions in these samples will clarify the role of TEs in the evolution and genome organization of ectomycorrhizal genus. The resource developed here will contribute to the extraction of information and/or confirmation of hypotheses concerning genome reorganization and adaptive mutations, as well as act as a reference for shared sequences and differences among genomes within and among related strains.

Supporting Information

Table S1 Table showing the percentage of identity of multiple alignment of TE copies corresponding to 175 consensus TE families identified by the RMBLR procedure.

(DOCX)

Acknowledgments

We would like to thank B. Hiselberger, Dr. E. Tisserant, Y. Cheng, Dr. P. Rouzé, Dr. J. Wuyst and Dr. H. Quesneville for their assistance and helpful discussions.

Author Contributions

Conceived and designed the experiments: JL FL FM. Performed the experiments: JL CM EM. Analyzed the data: JL CM. Contributed reagents/materials/analysis tools: JL CM EM. Wrote the paper: JL. Edited manuscript: JL FL FM GAT.

18. Quesneville H, Bergam CM, Andrieu O, Autard D, Nouaud D, et al. (2005) Combined evidence annotation of transposable elements in genome sequences. *PLoS Comp. Biol.* 1(2): e22.
19. Benson G (1999) Tandem repeat finder: a program to analyses DNA sequences. *Nucleic Acids Res.* 27(2): 573–580.
20. McCarthy EM, McDonald JF (2003) LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* 19(3): 362–367.
21. Labbé J, Zhang X, Yin T, Schmutz J, Grimwood J, et al. (2008) A genetic linkage map for the ectomycorrhizal fungus *Laccaria bicolor* and its alignment to the whole-genome sequence assemblies. *New Phyt.* 180: 316–328.
22. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876–4882.
23. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406–425.
24. Felsenstein J (1989) PHYLIP – phylogeny inference package (version3.2). *Cladistics* 5: 164–166.
25. 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.
26. Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. *PNAS* 101(34): 12404–12410.
27. Berbee ML, Taylor JW (2010) Dating the molecular clock in fungi – how close are we? *Fung. Biol. Rev.* 24: 1–16
28. Niculita-Hirzel H, Labbé J, Kohler A, Le Tacon F, Martin F, et al. (2008) Gene organization of the mating type regions in the ectomycorrhizal fungus *Laccaria bicolor* reveals distinct evolution between the two mating type loci. *New Phyt.* 180(2): 329–342.
29. Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JL, et al. (1999) Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. *Proc. Natl. Acad. Sci. USA* 96: 7409–7414.
30. Fu H, Park W, Yan X, Zheng Z, Shen B, et al. (2001) The highly recombinogenic *bz* locus lies in an unusually gene-rich region of the maize genome. *Proc. Natl. Acad. Sci USA* 98: 8903–8908.
31. Barabaschi D, Guerra D, Davide L, Katia L, Paolo M, et al. (2011) Emerging Knowledge from Genome Sequencing of Crop Species. *Mol. Biotech.* DOI: 10.1007/s12033-011-9443-1.
32. Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, et al. (2011) Perigord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* 464: 1033–1038.
33. Muszewska A, Hoffman-Sommer M, Grynberg M (2011) LTR Retrotransposons in Fungi. *PLoS ONE* 6(12): e29425.
34. Novikova O, Fet V, Blinov A (2009) Non-LTR retrotransposons in fungi. *Funct. Integr. Genomics* 9(1): 27–42.
35. Yang L, Bennetzen JL (2009) Distribution, diversity, evolution, and survival of *Helitrons* in the maize genome. *Proc. Natl. Acad. Sci. USA* 106: 19922–19927.
36. Morgante M (2006) Plant genome organization and diversity: The year of the junk! *COB* 17: 168–173.
37. Bureau TE, Wessler SR (1994a) Mobile inverted-repeat elements of the Tourist family are associated with the genes of many cereal grasses. *Proc. Natl. Acad. Sci. USA* 91: 1411–1415.
38. Bureau TE, Wessler SR (1994b) Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* 6: 907–916.
39. Jiang N, Wessler SR (2001) Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested element. *Plant Cell* 13: 2553–2564.
40. Iwamoto M, Higo K (2003) Tourist C transposable elements are closely associated with genes expressed in flowers in rice (*Oryza sativa*). *Mol. Genet. Genomics* 268: 771–778.
41. Hu J, Reddy VS, Wessler SR (2000) The rice R gene family: two distinct subfamilies containing several miniature inverted-repeat transposable elements. *Plant Mol. Biol.* 42: 667–678.
42. Ryberg M, Matheny PB (2011) Asynchronous origins of ectomycorrhizal clades of Agaricales. *Proc. R. Soc. B.* doi: 10.1098/rspb.2011.2428.
43. Halling RE (2001) Ectomycorrhizae: co-evolution, significance, and biogeography. *Ann. Mol. Bot. Gard* 88:5–13.
44. Baucom RS, Estill JC, Leebens-Mack J, Bennetzen JL (2009) Natural selection on gene function drives the evolution of LTR retrotransposon families in the rice genome. *Genome Res.* 19:243–254.
45. Walbot V, Petrov DA (2001) Gene galaxies in the maize genome. *Proc. Natl. Acad. Sci. USA* 98: 8163–8164.
46. Labbé J, Murat C, Morin E, Le Tacon F, Martin F (2011) Survey and analysis of simple sequence repeats in the *Laccaria bicolor* genome, with development of microsatellite markers. *Curr. Genet.* 57(2): 75–88.
47. Akagi H, Yokozeki Y, Inagaki A, Mori K, Fujimura T (2001) Micron, a microsatellite-targeting transposable element in the rice genome. *Mol. Genet. Genomics* 266: 471–480.
48. Tay WT, Behere GT, Batterham P, Heckel D (2010) Generation of microsatellite repeat families by RTE retrotransposons in Lepidopteran genomes. *BMC Evol. Biol.* 10: 144–152.
49. Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA (1995) Alu repeats: a source for the genesis of primate microsatellite. *Genomics* 29: 136–144.
50. Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* 5: 435–445.
51. Zhu Y, Strassmann JE, Queller DC (2000) Insertions, substitutions, and the origin of microsatellites. *Genet. Res. Camb.* 76: 227–236.
52. Metzgar D, Bytof J, Wills C (2000) Selection Against Frameshift Mutations Limits Microsatellite Expansion in Coding DNA. *Genome Res* 10: 72–80.
53. Flutre T, Duprat E, Feuillet C, Quesneville H (2011) Considering transposable element diversification in de novo annotation approaches. *PLoS ONE* 6: 1–15.
54. Morgante M, De Paoli E, Radovic S (2007) Transposable elements and the plant pan-genomes. *COPD* 10: 149–155.
55. Wenzl P, Wong L, Kwang won K, Jefferson R (2005) A Functional Screen Identifies Lateral Transfer of BetaGlucuronidase (Gus) from Bacteria to Fungi. *Mol. Biol. Evol.* 22: 308–316.
56. Slot JC, Hibbett DS (2007) Horizontal Transfer of a Nitrate Assimilation Gene Cluster and Ecological Transitions in Fungi: A Phylogenetic Study. *PLoS ONE* 2: 1097–1099.
57. Khaldi N, Collemare J, Lebrun MH, Wolfe KH (2008) Evidence for Horizontal Transfer of Secondary Metabolite Gene Cluster between Fungi. *Genome Biol.* 9: 18–22.
58. Johnson LJ (2007) The genome strikes back: the evolutionary importance of defense against mobile elements. *Evol. Biol.* 34: 121–129.
59. Miller WJ, McDonald JF, Nouaud D, Anxolabéhère D (1999) Molecular domestication: more than a sporadic episode in evolution. *Genetica* 107: 197–207.
60. Plett JM, Gibon J, Kohler A, Duffy K, Hoegger PJ, et al. (2012) Phylogenetic, genomic organization and expression analysis of hydrophobin genes in the ectomycorrhizal basidiomycete *Laccaria bicolor*. *Fungal Genetics and Biology* 49(3): 199–209.