

RESEARCH ARTICLE

Open Access

Evolutionary genomics revealed interkingdom distribution of Tcn1-like chromodomain-containing Gypsy LTR retrotransposons among fungi and plants

Olga Novikova*¹, Georgiy Smyshlyaev² and Alexander Blinov¹

Abstract

Background: Chromodomain-containing Gypsy LTR retrotransposons or chromoviruses are widely distributed among eukaryotes and have been found in plants, fungi and vertebrates. The previous comprehensive survey of chromoviruses from mosses (Bryophyta) suggested that genomes of non-seed plants contain the clade which is closely related to the retrotransposons from fungi. The origin, distribution and evolutionary history of this clade remained unclear mainly due to the absence of information concerning the diversity and distribution of LTR retrotransposons in other groups of non-seed plants as well as in fungal genomes.

Results: In present study we performed *in silico* analysis of chromodomain-containing LTR retrotransposons in 25 diverse fungi and a number of plant species including spikemoss *Selaginella moellendorffii* (Lycopodiophyta) coupled with an experimental survey of chromodomain-containing Gypsy LTR retrotransposons from diverse non-seed vascular plants (lycophytes, ferns, and horsetails). Our mining of Gypsy LTR retrotransposons in genomic sequences allowed identification of numerous families which have not been described previously in fungi. Two new well-supported clades, Galahad and Mordred, as well as several other previously unknown lineages of chromodomain-containing Gypsy LTR retrotransposons were described based on the results of PCR-mediated survey of LTR retrotransposon fragments from ferns, horsetails and lycophytes. It appeared that one of the clades, namely Tcn1 clade, was present in basidiomycetes and non-seed plants including mosses (Bryophyta) and lycophytes (genus *Selaginella*).

Conclusions: The interkingdom distribution is not typical for chromodomain-containing LTR retrotransposons clades which are usually very specific for a particular taxonomic group. Tcn1-like LTR retrotransposons from fungi and non-seed plants demonstrated high similarity to each other which can be explained by strong selective constraints and the 'retained' genes theory or by horizontal transmission.

Background

Retrotransposons are a class of mobile genetic elements, which use reverse transcription in their transposition. Five orders of retrotransposons are recognized: those having long terminal repeats (LTRs) (LTR retrotransposons); those lacking LTRs (non-LTR retrotransposons); DIRS retrotransposons; Penelope-like retrotransposable elements; and short interspersed nuclear elements (SINEs). According to the modern classification, LTR ret-

rotransposons are divided into several superfamilies: Copia (Pseudoviridae), Gypsy (Metaviridae), Bel-Pao, Retrovirus (Retroviridae), and ERV [1].

Chromodomain-containing LTR retrotransposons or chromoviruses are the most widespread lineage of Gypsy LTR retrotransposons and are present in genomes of fungi as well as in plants and vertebrates [2,3]. The characteristic feature of chromoviruses is the presence of an additional domain - the chromodomain (CHD). CHDs are present in various eukaryotic proteins involved in chromatin remodeling and regulation of gene expression during development [4-6]. CHDs perform a wide range of

* Correspondence: novikova@bionet.nsc.ru

¹ Laboratory of Molecular Genetic Systems, Institute of Cytology and Genetics, Novosibirsk, Russia

Full list of author information is available at the end of the article

diverse functions including chromatin targeting and protein/DNA/RNA interactions [6]. Recently, it has been shown that the CHDs target integration of new LTR retrotransposon copies into heterochromatin by recognizing histone modifications [7].

Our previous comprehensive survey of chromoviruses from mosses (Bryophyta) suggested that the diversity of CHD-containing Gypsy LTR retrotransposons in plant genomes is underestimated [8,9]. There are four well-known CHD-containing Gypsy LTR retrotransposon clades widely distributed among gymnosperms and angiosperms: Tekay, CRM, Galadriel and Reina [2,3]. Four novel clades were found to be present in mosses. Moreover, we showed that representatives from one of the moss-specific clades are more closely related to retrotransposons from fungi than to retrotransposons from plants. Although we proposed that the retrotransposons from this clade could have been 'retained' from the last common ancestor of Fungi/Metazoa lineage and plants, the origin of this clade remains unclear.

The questions addressed in current investigation are as follows: (1) what kind of Gypsy LTR retrotransposons from fungi are closely related to the LTR retrotransposons detected in mosses; (2) how widely those clades which were previously identified in mosses are distributed among other non-seed plants including lycophytes, ferns and horsetails; and (3) what is the origin of the clade which is common for non-seed plants and fungi. The *in silico* analysis of Gypsy LTR retrotransposons in 25 species of fungi, genomes of which available in public databases, along with survey of related LTR retrotransposons from whole genome sequences (WGS) and expressed sequence tags (ESTs) of diverse plants including spike-moss *Selaginella moellendorffii* (Lycopodiophyta) and PCR-based screening of ferns, horsetails and lycophytes showed that a common clade of CHD-containing Gypsy LTR retrotransposons can be found in mosses, lycophytes from genus *Selaginella*, and basidiomycetes. According to the classification of CHD-containing Gypsy LTR retrotransposons proposed by Gorinsek et al. (2004) [3] this clade has name Tcn1. It seems that Tcn1 is a unique clade of chromoviruses which has a wide inter-kingdom distribution. Tcn1-like LTR retrotransposons from fungi and non-seed plants demonstrated higher similarity to each other in comparison with LTR retrotransposons from other clades. This can be explained by strong selective constraints and the 'retained' genes theory or by horizontal transmission.

Results

Gypsy LTR retrotransposons survey from fungal genomes

The Gypsy LTR retrotransposons mining from fungal genomes was initiated in an attempt to identify retrotransposons closely related to those which were found

in mosses (Bryophyta) [8]. The survey was performed using genome sequence data for the 25 fungal species listed in Table 1. The hemiascomycetous yeasts were not included in the present investigation since a comprehensive survey of LTR retrotransposons from this group of ascomycetes was recently published [10]. First, reverse transcriptase (RT) and integrase (Int) coding regions of Gypsy LTR retrotransposons were detected in genomic sequences using algorithm based on hidden Markov model implemented in uGENE software <http://ugene.unipro.ru/>. The transposable elements thus identified were then classified into families based on RT and Int domains sequence similarity. Members of the same family shared high amino acid identity (90-100%) but had very little similarity to elements from other families. Our survey has identified more than 150 novel Gypsy LTR retrotransposon families which have not been described previously (Additional file 1).

The newly identified Gypsy LTR retrotransposons fall into two major, distinct lineages according to phylogenetic analysis based on RT and partial Int domains: chromoviruses (or CHD-containing Gypsy LTR retrotransposons) and Ylt1-like LTR retrotransposons. Two retrotransposons (LacBicTy3-15 from *Laccaria bicolor* and CopConTy3-14 from *Coprinus cinereus*) formed their own lineage closely related to Ylt1-like LTR retrotransposons. This lineage cannot be attributed to Ylt1 because of low bootstrap support (only 64%; Figure 1) and was named SN_1006. Ylt1-like LTR retrotransposons were found in a number of fungal species (Figure 1). Previously, Ylt1-like LTR retrotransposons were reported for *Yarrowia lipolytica* (original Ylt1 retrotransposon) [11], *Candida albicans* (Tca3 element) [12], and basidiomycete *Cryptococcus neoformans* [13]. In the present study, LTR retrotransposons belonging to the Ylt1 lineage have been found in both ascomycete and basidiomycete fungi. They fall into several clearly separated groups: the branch formed by transposable elements from Basidiomycota, the group of retrotransposons from ascomycetes, and the branch of the original Ylt1 LTR retrotransposon (Figure 1).

Twenty monophyletic clades can be recognized in the phylogenetic tree of fungal CHD-containing Gypsy LTR retrotransposons, nine of which have been previously reported [3]. Thirteen clades are specific for ascomycetes (Nessie, Pyret, Maggy, Pyggy, MGLR3, Yeti, Coccy1, Coccy2, Polly, Afut1, Tf1, Ty3, and Afut4), six have been found only in genomes of basidiomycetes (MarY1, Laccy1, Laccy2, Tcn2, Puccy1, and Puccy2) and one clade (Tcn1) is present in both basidiomycetes and chytridiomycetes (*Batrachochytrium dendrobatidis* JEL423) (Figure 2).

For each of the newly identified retrotransposon families, we attempted to isolate a full-length representative

Table 1: List of fungal species, genomes of which were analyzed *in silico* in present study

Phylum/Subphylum	Class	Species and strain	genome size (Mb)	LTRa
Ascomycota/ Pezizomycotina	Sordariomycetes	<i>Chaetomium globosum</i> CBS 148.51	36	23
		<i>Fusarium oxysporum</i> 4286 FGSC	60	39
		<i>Fusarium verticillioides</i> 7600	46	4
		<i>Nectria haematococca</i> MPVI	40	65
		<i>Podospora anserina</i> S mat+	37	11
		<i>Trichoderma reesei</i> QM6a	33	5
		<i>Trichoderma virens</i> Gv29-8	38	5
	Eurotiomycetes	<i>Aspergillus clavatus</i> NRRL 1	35	29
		<i>Aspergillus niger</i> ATCC1015	37	1
		<i>Aspergillus terreus</i> NIH2624	35	3
		<i>Coccidioides immitis</i> RS	29	236
		<i>Histoplasma capsulatum</i> NAM1	28	74
	Leotiomycetes	<i>Uncinocarpus reesii</i> 1704	30	64
		<i>Sclerotinia sclerotiorum</i> 1980	38	25
		<i>Botrytis cinerea</i> B05.10	38	25
	Dothideomycetes	<i>Alternaria brassicicola</i> ATCC 96866	30	108

Table 1: List of fungal species, genomes of which were analyzed *in silico* in present study (Continued)

		<i>Pyrenophora tritici-repentis</i> Pt-1C-BFP	37.8	118
		<i>Stagonospora nodorum</i> SN15	37	12
Basidiomycota/ Agaricomycotina	Agaricomycetes (Homobasidiomycetes)	<i>Amanita bisporigera</i>	58	7
		<i>Coprinus cinereus</i> Okayama7#130	38	160
		<i>Laccaria bicolor</i> S238N	61	114
		<i>Postia placenta</i> MAD-698	90	921
Basidiomycota/ Pucciniomycotina	Urediniomycetes	<i>Sporobolomyces roseus</i>	21	6
		<i>Puccinia graminis f. sp. tritici</i>	81.5	534
Chytridiomycota	Chytridiomycetes	<i>Batrachochytrium dendrobatidis</i> JEL423	24	1

^a total copy number of detected Gypsy LTR retrotransposons

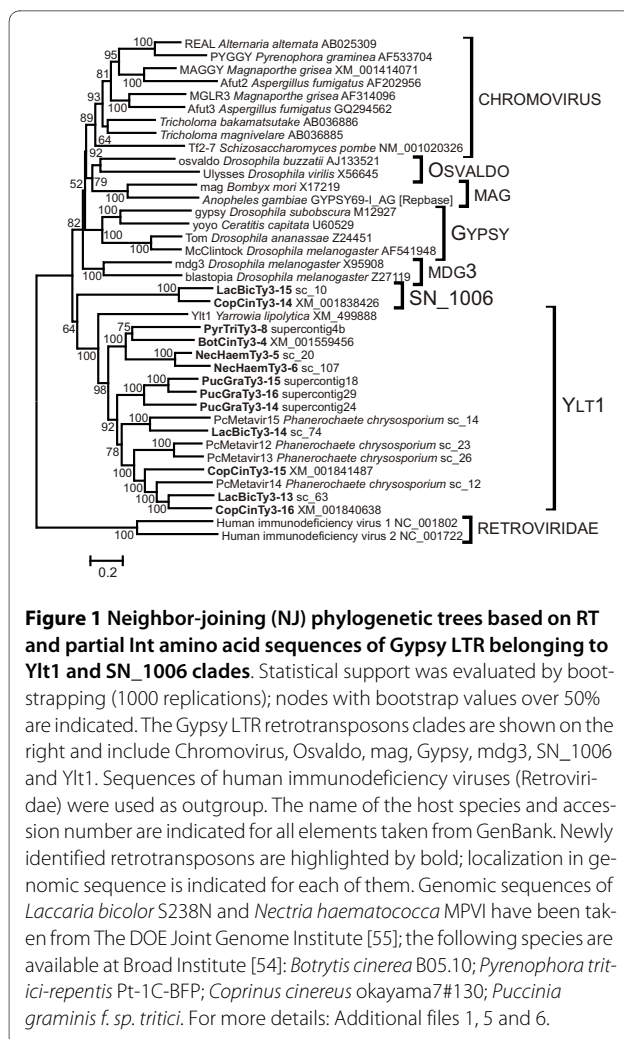
or to reconstruct it using overlaps between partial sequences. The lengths of the elements thus identified varied greatly from approximately 4.4 kb to more than 13 kb. The structural features for each family are listed in Additional Table S1 (Additional file 1). The majority of full-length LTR retrotransposons had either a single open reading frame (ORF) encoding a fused Gag-Pol polyprotein or two ORFs encoding separate proteins (Figure 3). The Gag protein sequences differed greatly between families. Nevertheless, cysteine motifs characterized by the amino acid sequence C-X₂-C-X₄-H-X₄-C (CCHC) were found in Gag for some of the identified Gypsy LTR retrotransposons (see Additional file 1). The Pol polyproteins sequences were more conserved than Gag, especially with the RT and Int domains. RT, Int, PR (proteinase) and chromodomains (CHDs, in chromoviruses) were detected. Characteristic motifs were found throughout the Gag and Pol sequences of all of the putative intact element copies.

In addition to the abovelisted enzymatic domains, a deoxyuridine triphosphatase domain (dUTPase) has been found in several LTR retrotransposons from the basidiomycete *Postia placenta* MAD-698. The location of this

domain varied among diverse families of PosPlaTy3 elements. It can be found either at amino-terminus (PosPlaTy3-3) and carboxyl-terminus of Pol (PosPlaTy3-4) or between PR and RT domains (PosPlaTy3-5) (Figure 3). The presence of dUTPase in LTR retrotransposon sequences has been described earlier for the elements from a basidiomycete *Phanerochaete chrysosporium* and an ascomycete *Tuber melanosporum* [14,15]. The role and origin of this domain remained unclear. Moreover, it seems that the described LTR retrotransposons acquire this domain independently from different sources (Additional file 2) [14]. It was assumed that the presence of dUTPase allows viruses that contain this domain to replicate in non-dividing cells, in which cellular dUTPase activity is absent because replication of DNA does not occur [16].

Tnc1 clade is found in non-seed plants

Further phylogenetic analysis revealed that previously described CHD-containing LTR retrotransposons from mosses including PpatensLTR retrotransposons isolated from genomic sequence of moss *Physcomitrella patens* formed a common branch with Tcn1-like LTR retrotransposons from fungi (Figure 4) [8]. In an attempt to deter-



mine the distribution of the Tcn1 clade among plants, we used public databases for further survey of LTR retrotransposons including genomic databases for red and green algae, spikemoss *Selaginella moellendorffii*, and seed plants (see Materials and Methods section). A Tcn1-like LTR retrotransposon search was implemented with BLAST (blastp and blastx). Amino acid sequences of RT and Int domains of known Tcn1-like (Tcn1 from *C. neoformans*, Cchromovir1 and Cchromovir2 from *C. cinereus*, PcMetavir6 from *Phanerochaete chrysosporium*, and PpatensLTRs from *P. patens*) and newly identified retrotransposons (SpoRosTy3-4 and BatDenTy3-1) were used as the queries. The Tcn1-like LTR retrotransposons were identified only in the whole genomic sequence of *Selaginella moellendorffii* (SM-Tcn1, Figure 4); none of the tested algae or seed plant genomes contained LTR retrotransposons from this clade. It seems that the Tcn1 clade can be found in basidiomycetes and chytridiomycetes fungi as well as non-seed plants (Bryophyta and Lycopodiophyta).

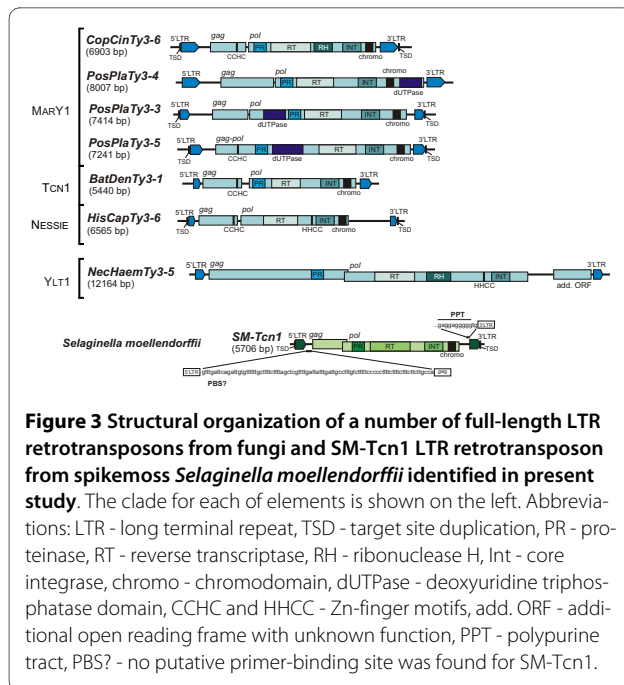
The whole sequence of SM-Tcn1 LTR retrotransposon was obtained from WGS. SM-Tcn1 is 5704 bp in length and carries two putative ORFs. ORF1 or *gag* (969 bp in length) encodes a 323 amino acid (aa) protein with strong similarity to retroviral Gag proteins (pfam03732). ORF2 or *pol* (3714 bp) encodes a 1238 aa polyprotein, with characteristic retroviral aspartyl protease (PR), RT, Int, and CHD domains (Figure 3; Additional file 3). SM-Tcn1 possesses 440 bp LTRs with conserved features, including the dinucleotide end sequences (TG...CA). Target site duplication (AACAC...AACAC) was also detected for the described copy of SM-Tcn1. The LTRs contained TATA and CAAT boxes. No putative primer-binding site (PBS) was found. The PBS is necessary for initiation of reverse transcription and synthesis of the first strand complementary 5'LTR sequence [17]. Typical PBS is located near the 5'LTR and complementary to the 3' terminal nucleotides of the primer tRNA used for initiation. Other known mechanism for initiation of the first strand synthesis is self-priming, in this case a sequence derived from LTR is located just downstream of the 5'LTR [18,19]. However, evidences were found neither for tRNA priming nor for self-priming of SM-Tcn1 LTR retrotransposon. The sequence presented between 5' LTR and *gag* is conservative and thymine-rich (Figure 3). The possible mechanism for initiation of reverse transcription of SM-Tcn1 remained unclear. A polypurine tract (PPT) was detected immediately upstream of the 3'LTR. The PPT sequence is involved in second-strand DNA synthesis. The BLAST search (blastn) of full-length SM-Tcn1 retrotransposon indicated the presence of more than 200 hits in the *S. moellendorffii* genome. The close examination of identified copies showed that they have in average 91.7% nucleotide identity with the original SM-Tcn1 sequence. A BLAST search (tblastn) using SM-Tcn1 putative Pol protein as a query yielded more than 500 hits. The size of *S. moellendorffii* genome is only ~100 Mbp [20], thus SM-Tcn1 retrotransposons comprise ca. 1.5% of genomic sequence.

Gypsy LTR retrotransposons from non-seed vascular plants

The bioinformatic survey of CHD-containing Gypsy LTR retrotransposons, which allowed us to identify Tcn1-like retrotransposons in a few basidiomycetes, *P. patens* and *S. moellendorffii*, did not provide a satisfactory answer to the question concerning distribution of this clade among non-seed vascular plants. Therefore, we used PCR with degenerate primers to investigate the distribution of CHD-containing Gypsy LTR retrotransposons in 26 ferns and horsetails (monilophytes) belonging to three classes: Psilotopsida, Polypodiopsida and Equisetopsida, and in 10 lycophytes from two classes: Isoetopsida and Lycopodiopsida (Table 2).



Figure 2 Neighbor-joining (NJ) phylogenetic trees based on RT and partial Int amino acid sequences of Gypsy LTR retrotransposons including newly described fungal chromodomain-containing LTR retrotransposons. Statistical support was evaluated by bootstrapping (1000 replications); nodes with bootstrap values over 50% are indicated. The clades are shown on the right. The name of the host species and accession number are indicated for all elements taken from GenBank. Newly identified retrotransposons are highlighted in bold; localization in genomic sequence is indicated for each of them. Genomic sequences of *Trichoderma reesei* QM6a, *Trichoderma vires* Gv29-8, *Nectria haematococca* MPV1, *Aspergillus niger* ATCC1015, *Alternaria brassicicola* ATCC 96866, *Stagonospora nodorum* SN15, *Laccaria bicolor* S238N, *Postia placenta* MAD-698, and *Sporobolomyces roseus* have been taken from The DOE Joint Genome Institute [55]; the following species are available at Broad Institute [54]: *Chaetomium globosum* CBS 148.51; *Fusarium oxysporum* 4286 FGSC; *Fusarium verticillioides* 7600; *Aspergillus clavatus* NRRL 1; *Aspergillus terreus* NIH2624; *Coccidioides immitis* RS; *Histoplasma capsulatum* NAM1; *Uncinocarpus reesii* 1704; *Sclerotinia sclerotiorum* 1980; *Botrytis cinerea* B05.10; *Pyrenophora tritici-repentis* Pt-1C-BFP; *Coprinus cinereus* okayama7#130; *Puccinia graminis* f. sp. *tritici*; *Batrachochytrium dendrobatidis* JEL423. The possible horizontal transmission (HT) is marked. For more details: Additional files 1, 5 and 6.



The estimated diversity of monilophytes (= Infradivision Moniliformopses) is about 9000 species and includes horsetails, whisk ferns, and all eusporangiate and leptosporangiate ferns [19]. Most of the species examined in the present study were leptosporangiate ferns from the order Polypodiales, class Polypodiopsida. This order covers more than 80% of current known diversity of ferns [21]. Additionally, one representative of heterosporous ferns, *Salvinia natans* (Polypodiopsida, Salviniaceae), two ophioglossoid ferns (Psilotopsida, Ophioglossales) and three horsetails (Equisetopsida, Equisetales) were included [22]. Lycophytes are much less diverse in comparison with monilophytes and comprise less than 1% of extant land plants (around 1200 living species). Three major lineages are distinguished among lycophytes: clubmosses and firmosses (Lycopodiaceae), spikemosses (Selaginellaceae), and quillworts (Isoetaceae) [23]. Among lycophytes included in the present study are *Isoetes* and *Huperzia* species (Isoetaceae) as well as two *Selaginella* species (Selaginellaceae), which belong to the class Isoetopsida, and seven diverse species from Lycopodiaceae (Lycopodiales, Lycopodiopsida).

The presence of CHD-containing Gypsy LTR retroelements among the listed plants was tested by amplifying genomic DNA with previously developed degenerate oligonucleotide primers [8,24]. Consistent with the spacing of reverse transcriptase (RT) domains, the amplified PCR products were approximately 320 bp in length. In total, 98 clones with sequence similarity to known RT sequences were isolated, of which 76 were from monilophytes and 22 from lycophytes. The preliminary blastp search

revealed that 10 clones were not from CHD-containing LTR retrotransposons but were from Athila-like Gypsy elements (Additional file 4). Many representatives of this clade possess not only classical gag and pol sequences, but also an additional open reading frame that might encode an env-like protein [25,26].

The phylogenetic relationships among obtained clones and known CHD-containing Gypsy LTR retrotransposons, extracted from databases, were reconstructed using neighbor-joining (NJ) analysis based on the multiple alignment of nucleotide sequences of RT fragment (Figure 4). The Gypsy LTR retrotransposons from *Drosophila melanogaster* were used as an outgroup. The newly identified LTR retrotransposon grouped into four large clusters on the phylogenetic tree. The group of clones from diverse monilophytes and one LTR retrotransposon from lycophytes *Lycopodium alpinum* (LycAlpTy3-1 clone) form a common group with previously described retroelements from mosses *Tetraphis pellucida* and *Vesicularia dubyana* [8]. Although the bootstrap support is below 50% (data not shown), this new clade (named "Galahad") seems to be a sister group to the Galadriel clade. Galahad appears to be one of the oldest widely distributed clades of CHD-containing LTR retrotransposons from plants. Since Galahad clade was found in all non-seed plants including mosses, the probable age of this clade would be in the range of 400-700 Myr, which is estimated time divergence of liverworts and mosses from vascular plants [27].

The second group is formed by LTR retrotransposons from both monilophytes and lycophytes. The phylogenetic analysis did not provide support for a monophyletic origin of this cluster. Moreover, the relationships inside the cluster remained unclear, with the exception of several lineages. One of the lineages ('a' in Figure 4), contained members from lycophytes in the family Lycopodiaceae: *Lycopodium clavatum* (LycClavGty3 clones), *L. japonicum* (LycJapGty3-1 clone), and *Diphasiastrium complanatum* (DiphComGty3 clones). Another lineage seems to have had a long-term association with fern genomes, since it appears to be widely distributed among leptosporangiate ferns and can be found in Dennstaedtiaceae (*Pteridium aquilinum*), Pteridaceae (*Adiantum pedatum*), Aspleniaceae (*Asplenium viride*), Woodsiaceae (*Athyrium distentifolium* and *Cystopteris fragilis*), and Dryopteridaceae (*Dryopteris crassirhizoma* and *Polystichum tripterum*) (lineage 'b' on Figure 4). Additionally, five satellite lineages, represented mostly by single clones can be found on the phylogenetic tree.

The largest group is represented by 37 LTR retrotransposons. Three clearly separated clusters can be found inside this group (marked as 'd', 'f', and Mordred on Figure 4). One of these clusters is formed by LTR retrotransposons from Ophioglossaceae (*Botrychium multifidum*),

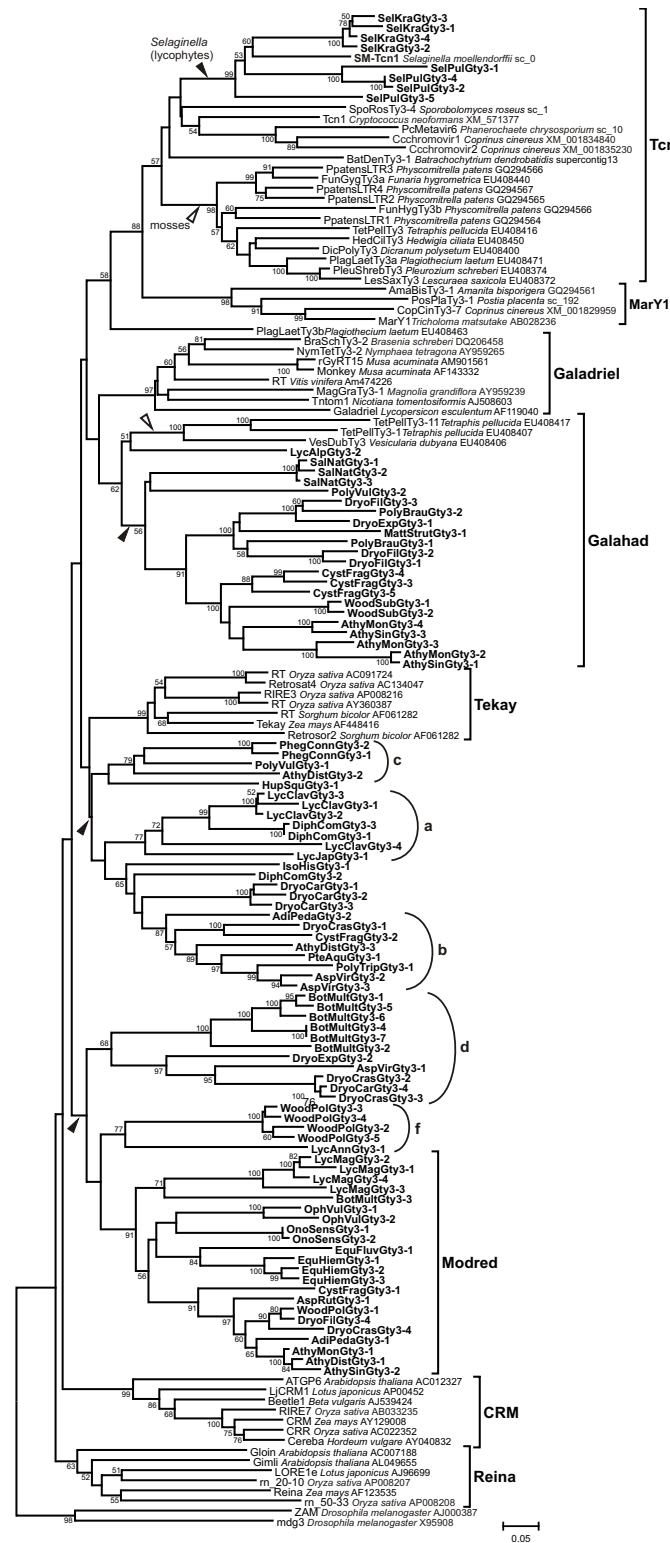


Figure 4 Neighbor-joining (NJ) phylogenetic tree based on RT nucleotide sequences of CHD-containing Gypsy LTR retrotransposons including newly described elements from monilophytes and lycophytes plants (highlighted in bold). Statistical support was evaluated by bootstrapping (1000 replications); nodes with bootstrap values over 50% are indicated. The name of the host species and accession number are indicated for LTR retrotransposons taken from GenBank. Four diverse clusters of LTR retrotransposons from mosses, monilophytes and lycophytes are shown by arrows. The group of Tcn1-like LTR retrotransposons from mosses (Bryophyta) is also indicated. Previously known clades, clades described in this study, and unclassified lineages (a-f) are shown on the right.

Table 2: List of monilophytes and lycophytes, which were analyzed experimentally in present study

Division	Class	Family	Species	Chr (Athila)a
Moniliformopses	Psilotopsida	Ophioglossaceae	<i>Ophioglossum vulgatum</i> L.	2
			<i>Botrychium multifidum</i> (Gmelin) Rupr.	6
	Polypodiopsida	Dennstaedtiaceae	<i>Pteridium aquilinum</i> (L.) Kuhn	1 (1)
			<i>Adiantum pedatum</i> L.	2
			<i>Asplenium viride</i> Huds.	3 (1)
			<i>Asplenium ruta-muraria</i> L.	1
			<i>Athyrium sinense</i> Rupr.	3 (3)
			<i>Athyrium distentifolium</i> Tausch ex Opiz	3 (1)
			<i>Athyrium monomachii</i> (Kom.) Kom.	4 (1)
			<i>Cystopteris fragilis</i> (L.) Bern.	5
			<i>Woodsia polystichoides</i> DC Eaton	5
			<i>Woodsia subcordata</i> Turcz.	2
			<i>Phegopteris connectilis</i> (Michx.) Watt.	2
			<i>Matteuccia struthiopteris</i> (L.) Tod.	1 (1)
			<i>Onoclea sensibilis</i> L.	2
<i>Dryopteris expansa</i> (Presl) Fraser-Jenk. & Jermy	2			
<i>Dryopteris filix-mas</i> (L.) Schott	4			
<i>Dryopteris crassirhizoma</i> Nak.	4			

Table 2: List of monilophytes and lycophytes, which were analyzed experimentally in present study (Continued)

			<i>Dryopteris carthusiana</i> (Vill.) Fuchs	4
			<i>Polystichum braunii</i> (Spenner) Fée	2
			<i>Polystichum tripterum</i> (Kunze) Presl	1
		Polypodiaceae	<i>Polypodium vulgare</i> L.	2
			<i>Pyrrosia lingua</i> (Thunb.) Farw.	-- (1)
		Salviniaceae	<i>Salvinia natans</i> (L.) All.	3
	Equisetopsida	Equisetaceae	<i>Equisetum hiemale</i> L.	3
			<i>Equisetum fluviatile</i> L.	1
Lycopodiophyta	Isoetopsida	Isoetaceae	<i>Isoetes histrix</i> Bory	1
		Selaginellaceae	<i>Selaginella kraussiana</i> (Kunze) A. Braun	4
			<i>Selaginella pulvinata</i> (Hook. & Grev.) Maxim.	4
	Lycopodiopsida	Lycopodiaceae	<i>Diphasiastrum complanatum</i> (L.) Holub	3
			<i>Huperzia squarrosa</i> (Forst.) Trevis.	1
			<i>Lycopodium alpinum</i> L.	1
			<i>Lycopodium clavatum</i> L.	4
			<i>Lycopodium magellanicum</i> Hert. ex Nessel	4
			<i>Lycopodium japonicum</i> Thunb. ex Murr.	1
			<i>Lycopodium annotinum</i> L.	1 (1)

^a number of unique sequences for CHD-containing LTR retrotransposons (Chr) and Athila-like LTR retrotransposons obtained in present study

Aspleniaceae (*A. viride*) and Dryopteridaceae (*Dryopteris expansa*, *D. carthusiana*, and *D. crassirhizoma*). The second cluster has a bootstrap support of 77% and is represented by clones isolated from *Lycopodium annotinum* (Lycopodiaceae) and *Woodsia polystichoides* (Woodsiaceae). The last monophyletic cluster, clade Mordred, is the largest, well-supported clade which is widely distributed among representatives of all investigated classes except Isoetopsida. It was found to be present in fern genomes from families Ophioglossaceae (*Ophioglossum vulgatum* and *Botrychium multifidum*), Pteridaceae (*Adiantum pedatum*), Woodsiaceae (*Athyrium sinense*, *A. distentifolium*, *A. monomachii*, *C. fragilis*, and *W. polystichoides*), Onocleaceae (*Onoclea sensibilis*), and Dryopteridaceae (*Dryopteris filix-mas* and *D. crassirhizoma*); in horsetails *Equisetum hiemale* and *E. fluviatile* (Equisetaceae); and lycophyte *Lycopodium magellanicum* (Lycopodiaceae).

The Tcn1-like LTR retrotransposons were detected only in *Selaginella* species, *S. kraussiana* (SelKraGty3 clones) and *S. pulvinata* (SelPulGty3 clones) in addition to the previously described LTR retrotransposons from mosses and SM-Tcn1 LTR retrotransposon from *Selaginella moellendorffii* (Figure 4) [8]. The absence of Tcn1-like clones isolated from other lycophytes, ferns, and horsetails can be explained by failed PCR amplification due to the high divergence of Tcn1-like elements in genomes of these species or, more likely, by lack of these elements from their genomes.

Tcn1-like LTR retrotransposons: 'retained' or horizontally transmitted?

As a rule chromovirus clades are specific for a particular group of eukaryotic organisms such as Ascomycota fungi (Nessie, Pyret, Maggy, Pyggy, MGLR3, Yeti, Coccy1, Coccy2, Polly, Afut1, Tf1, Ty3, and Afut4), Basidiomycota fungi (MarY1, Laccy1, Laccy2, Tcn2, Puccy1, and Puccy2), or plants (Reina, CRM, Tekay, Galadriel, and Chlamyvir as well as additional less investigated clades from mosses) [3,8]. In the light of such specificity, it was unexpected to find a clade containing elements from basidiomycetes and non-seed plants. Nevertheless, it seems that Tcn1 clade has an interkingdom distribution and can be found in a number of fungi, diverse mosses (Bryophyta) as well as in lycophytes (genus *Selaginella*). Such a wide distribution makes the Tcn1 clade unique among the CHD-containing Gypsy LTR retrotransposons. The interkingdom distribution of Tcn1 clade could be the result of horizontal transmission (HT) of LTR retrotransposons among fungi and plants; otherwise, Tcn1-like LTR retrotransposons could have been 'retained' by mosses and lycophytes from the most recent common ancestor of plants and Fungi/Metazoa lineage of eukaryotes [8,28].

The hypothesis of 'retained' genes is based on the observation that EST data of *Physcomitrella* contained a fraction of transcripts derived from putative genes ('retained' genes), which are not present in seed plants but can be found in other kingdoms including fungi. It was proposed that such retained genes along with *Physcomitrella*-specific (or moss-specific) genes encode functions that make mosses unique in terms of physiology and metabolism [28]. We used these data and compared the levels of similarity for RT-Int fragments from Tcn1-like LTR retrotransposons and two putatively retained genes from *Physcomitrella*, which showed a high similarity with functional genes from fungi: uric acid-xanthine permease (uapA, TIGR00801) and inorganic phosphate transporter (Pho88, pfam10032).

The pairwise comparisons between hypothetical Pho88 proteins from basidiomycetes *Coprinus cinereus* Okayama7#130, *Phanerochaete chrysosporium* and *Cryptococcus neoformans* revealed 51.3% to 64.9% similarity whereas only 23.0% identical amino acid residues was found on average in pairwise comparisons between fungal proteins and putative Pho88 from *Physcomitrella* (Table 3). The most closely related homolog for putative Pho88 from *P. patens* was found in *Schizosaccharomyces pombe* (26.8% of similarity). The similarity between uapA from *Physcomitrella* and *Cryptococcus* (36.9%) was almost the same as between proteins from *Cryptococcus* and *Coprinus* (39.6%) or *Cryptococcus* and *Phanerochaete* (43.2%). More than 67% of amino acid residues are identical in permeases from *Coprinus* and *Phanerochaete*. Predicted uapA from *Ashbya gossypii* and *Physcomitrella* share 40.7% of amino acid residues.

It seems to be that the hypothesis of 'retained' genes cannot be implemented as explanation for Tcn1 clade distribution since investigated RT-Int fragments of Tcn1-like LTR retrotransposons from fungi and plants have higher similarity to each other than functional proteins which were proposed to be 'retained' [28]. RT-Int fragments from Tcn1-like LTR retrotransposons have average similarity 49%. Moreover, evolutionary rates estimated for Tcn1 LTR retrotransposons appeared to be less than evolutionary rates for 'retained' genes or other LTR retrotransposons (Table 3).

Discussion

Despite a number of whole genome sequence studies, the distribution and diversity of CHD-containing Gypsy LTR retrotransposons is still poorly understood. Current knowledge of distribution and evolution of this group of mobile elements has been mainly obtained from diverse model organisms [3]. The quickly generated massive data sets (for example, WGS and EST databases) provide a great opportunity to perform detailed analysis for non-model organisms. Experimental data accumulation also

Table 3: Amino acid divergences of proteins and RT-Int fragments of CHD-containing Gypsy LTR retrotransposons from Tcn1, Pyggy and Pyret clades

Genes or LTR retrotransposons	Length	Amino acid identity (%)	Evolutionary rate (10 ⁻⁹) ^b
inorganic phosphate transporter (Pho88) ^a			
<i>Physcomitrella patens</i> (XM_001783642)/ <i>Schizosaccharomyces pombe</i> (NM_001019478)	154 aa	26.8	0.420
<i>P. patens</i> (XM_001783642)/ <i>Coprinus cinereus</i> (XM_001836748)	151 aa	23.1	0.445
<i>P. patens</i> / <i>Phanerochaete chrysosporium</i> (JGI: scaffold_7 [325337..326089])	151 aa	25.0	0.428
<i>P. patens</i> / <i>Cryptococcus neoformans</i> (XM_569621)	152 aa	20.1	0.521
<i>C. cinereus</i> / <i>Ph. chrysosporium</i>	151 aa	64.9	ND
<i>C. cinereus</i> / <i>C. neoformans</i>	151 aa	51.3	0.472
<i>Ph. chrysosporium</i> / <i>C. neoformans</i>	151 aa	58.6	ND
uric acid-xanthine permease (uapA) ^a			
<i>P. patens</i> (XM_001784081)/ <i>Ashbya gossypii</i> (NM_212305)	466 aa	40.7	0.266
<i>P. patens</i> / <i>C. cinereus</i> (XM_001839036)	456 aa	38.4	0.288
<i>P. patens</i> / <i>Ph. chrysosporium</i> (JGI: scaffold_22 [347275..348820])	391 aa	43.6	0.263
<i>P. patens</i> / <i>C. neoformans</i> (AF542528)	472 aa	36.9	0.318
<i>C. cinereus</i> / <i>Ph. chrysosporium</i>	391 aa	67.9	ND
<i>C. cinereus</i> / <i>C. neoformans</i>	456 aa	39.6	0.560

Table 3: Amino acid divergences of proteins and RT-Int fragments of CHD-containing Gypsy LTR retrotransposons from Tcn1, Pyggy and Pyret clades (Continued)

<i>Ph. chrysosporium</i> / <i>C. neoformans</i>	391 aa	43.2	ND
Tcn1			
Tcn1 <i>C. neoformans</i> / Ccchromovir-1 <i>C. cinereus</i>	684 aa	53.6	0.441
Tcn1 <i>C. neoformans</i> / PcMetavir6 <i>Ph. chrysosporium</i>	684 aa	52.2	ND
Tcn1 <i>C. neoformans</i> / BatDenTy3-1 <i>B. dendrobatidis</i>	677 aa	49.1	0.381
PcMetavir6 <i>Ph. chrysosporium</i> / Ccchromovir-1 <i>C. cinereus</i>	688 aa	72.4	ND
PcMetavir6 <i>Ph. chrysosporium</i> / BatDenTy3-1 <i>Batrachochytrium</i> <i>dendrobatidis</i>	677 aa	47.0	0.412
BatDenTy3-1 <i>B. dendrobatidis</i> / Ccchromovir-1 <i>C. cinereus</i>	677 aa	46.5	0.411
Tcn1 <i>C. neoformans</i> / PpatensLTR1 <i>P. patens</i>	675 aa	52.0	0.216 (HT)
Tcn1 <i>C. neoformans</i> /SM-Tcn1 <i>Selaginella moellendorffii</i>	682 aa	50.3	0.235 (HT)
PcMetavir6 <i>Ph. chrysosporium</i> / PpatensLTR1 <i>P. patens</i>	675 aa	51.8	0.217 (HT)
PcMetavir6 <i>Ph. chrysosporium</i> / SM-Tcn1 <i>S. moellendorffii</i>	682 aa	49.5	0.230 (HT)
PpatensLTR1 <i>P. patens</i> /SM- Tcn1 <i>S. moellendorffii</i>	675 aa	58.6	0.429
Pyggy			
PyrTriTy3-2 <i>Pyrenophora tritici-</i> <i>repentis</i> / <i>NecHaemTy3-</i> <i>4Nectria haematococca</i>	706 aa	77.0	0.259 (HT)
PyrTriTy3-2 <i>P. tritici-repentis</i> / ChaGloTy3-8 <i>Chaetomium</i> <i>globosum</i>	706 aa	48.5	0.706

Table 3: Amino acid divergences of proteins and RT-Int fragments of CHD-containing Gypsy LTR retrotransposons from Tcn1, Pyggy and Pyret clades (Continued)

PyrTriTy3-2 <i>P. tritici-repentis</i> / grh <i>Magnaporthe grisea</i>	701 aa	49.2	0.703
PyrTriTy3-2 <i>P. tritici-repentis</i> / Dane4 <i>Aspergillus nidulans</i>	695 aa	55.5	0.531
AltBraTy3-2 <i>Alternaria brassicicola</i> / NecHaemTy3-4 <i>N. haematococca</i>	709 aa	47.7	0.880
AltBraTy3-2A. <i>brassicicola</i> / Dane4 <i>A. nidulans</i>	695 aa	49.2	0.649
grh <i>M. grisea</i> /Dane4 <i>A. nidulans</i>	695 aa	48.5	0.670
Pyret			
skippy <i>Fusarium oxysporum</i> / PyrTriTy3-1 <i>P. tritici-repentis</i>	648 aa	40.4	0.772
skippy <i>F. oxysporum</i> /AFLAV <i>Aspergillus flavus</i>	673 aa	41.4	0.766
AFLAV <i>A. flavus</i> /PyrTriTy3-1 <i>P. tritici-repentis</i>	648 aa	44.3	0.656

^a The corresponding accession numbers in GenBank are provided in the brackets;

^b ND - not determined: information concerning time divergence between species groups is unavailable; HT - putative horizontal transmission

should not be neglected. In the absence of information about genomic sequences from monilophytes, PCR screening seems to be very useful for isolation and characterization of new LTR retrotransposons from this group of non-seed vascular plants. Two new well-supported clades, Galahad and Mordred, as well as several other previously unknown lineages of CHD-containing Gypsy LTR retrotransposons were described based on the results of PCR-mediated survey of RT fragments from ferns, horsetails and lycophytes.

One of the clades originally described for fungal genomes, Tcn1, appeared to be present in genomes of mosses (Bryophytes) and lycophytes (genus *Selaginella*) (Figure 5). Such an interkingdom distribution is not typical for CHD-containing LTR retrotransposons clades which are usually very specific for a particular taxonomic group [see [8]]. Data suggested that horizontal transmission took place between fungi and non-seed plants (probably mosses and lycophytes). Horizontal transmissions or horizontal transfers (HTs) of mobile elements are usually

recognized by the presence of very closely related mobile elements in distant host taxa [29-33]. HT is well known for gypsy LTR retrotransposons in *Drosophila* [30] and has been suggested to have occurred in plants [2,24]. Recently, the evidence was provided for HT of RIRE1 LTR retrotransposon between representatives of genus *Oryza* [32] and Route66 LTR retrotransposon between representatives of Panicoideae (Poaceae) and several species of the genus *Oryza* [33].

Several criteria can be used for HT event recognition. The first criterion is inconsistencies between the phylogenies of transposable elements (TEs) and host species [29,34]. There are potential problems with application of this criterion for HT detection. Multiple transposable element lineages can be present within genomes. Moreover, transposable elements are multicopy components of genomes. Comparisons of paralogous copies instead of orthologs along with varying rates of their sequence evolution are the main sources for incongruence in phylogenetic analysis, this could be misidentified as HT. The

second criterion, which seems to offer the strongest evidence, is a higher degree of observed sequence similarity for transposable elements than for functional genes, so called 'slowdown effect on evolutionary rates'. Once inserted, a new copy of transposable element is presumed to evolve without functional constraints. Thus, all types of mutations should have an equal chance to be fixed [35]. The lower than expected sequence divergence of TEs in comparison with non-mobile nuclear genes of the host species can be explained either by strong selective constraints in TE sequence coupled with a strict vertical transmission, or by horizontal transfer [31,36,37]. The third criterion of inferring HT is the discontinuous distribution of TEs among closely related taxa, i.e., presence of a TE in one lineage and its absence in a sister lineage. Such discontinuous distribution could be due to random loss of TEs, ancestral polymorphism, or independent sorting of copies into descendant species. By itself, this kind of evidence provides only weak support for HT since TE can be lost through population dynamics or ecological forces that are difficult to reconstruct [38,39].

All three criteria are satisfied in case of Tcn1-like LTR retrotransposons. They demonstrated patchy distribution among fungi and plants (Figure 5). They were found in all investigated mosses, but only in a few lycophytes and they absent in basal lineages of green plants such as green

and red algae as well as in all seed plants investigated so far. The hypothesis of 'retained' genes in moss *Physcomitrella* represented an attractive alternative to horizontal transmission as an explanation of the phylogenetic inconsistencies as well as the existence of a number of functional genes in *Physcomitrella* genome, which seem to have non-plant origin and can be found in bacteria, fungi and protozoa but not in higher plants [28]. Nevertheless, the most important feature of Tcn1-like LTR retrotransposons in the context of HT is their lower evolutionary rates in comparison with other groups of CHD-containing LTR retrotransposons. The close examination and comparison of evolutionary rates for LTR retrotransposons including representatives of Tcn1, Pyggy and Pyret clades, and evolutionary rates estimated for putatively 'retained' genes suggests that a horizontal transmission of Tcn1-like LTR retrotransposons took place among fungi and the last common ancestor (LCA) of mosses and lycophytes (Table 3 and Figure 5) [27]. Alternatively, it is possible, but highly unlikely, that two independent acts of HT occurred. First HT event could happen among fungi and LCA of mosses since all investigated mosses contain Tcn1-like LTR retrotransposons [8]. The second HT could occur among fungi and LCA of *Selaginella* since only representatives of this genus carry this group of retrotransposons among all investigated lycophytes (Figure 4 and Figure 5). It is necessary to note that despite HT seeming to be a preferable explanation for the observed distribution. The evidence is not strong enough to discard other explanations; such as selective pressure coupled with vertical transmission of retrotransposons in genomes of non-seed plants and loss of these elements by other plants.

Another putative case of HT based on the results of present survey of LTR retrotransposons from fungal species was found for PyrTriTy3-2 LTR retrotransposon from *Pyrenophora tritici-repentis* Pt-1C-BFP (Dothideomycetes). PyrTriTy3-2 belongs to Pyggy clade and appeared to be more closely related to LTR retrotransposons from Sordariomycetes (NecHaemTy3-4 from *Nectria haematococca* MPVI and ChaGloTy3-8 from *Chaetomium globosum* CBS 148.51) than to the elements from other Dothideomycetes such as AltBraTy3-2 from *Alternaria brassicicola* ATCC 96866, REAL from *Alternaria alternata* (AB025309) [42], and PYGGY from *Pyrenophora graminea* (AF533704) [43] (Figure 2). The pairwise comparisons of RT-Int fragments and investigation of evolutionary rates for retrotransposons from Pyggy and Pyret clades revealed the unexpectedly high similarity between PyrTriTy3-2 and NecHaemTy3-4 (77% identical amino acids), much higher than between any other retrotransposons from Pyggy or Pyret clades, and at least two times lower evolutionary rate in the couple

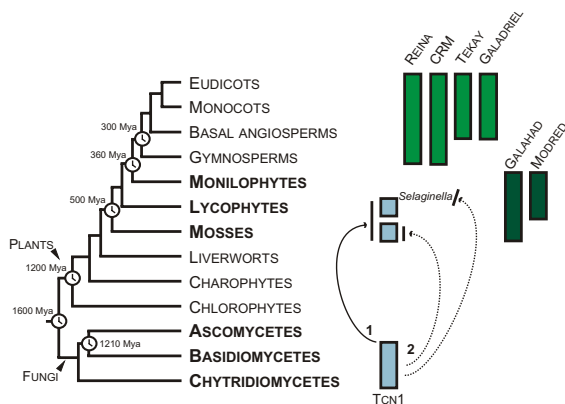


Figure 5 Distribution of different clades of CHD-containing Gypsy LTR retrotransposons in plants. Evolutionary tree is represented according to Bowman et al., 2007 [40] and Berbee and Taylor, 2001 [41] with minor modifications. Divergence times (Mya - million years ago) are indicated according to Hedges, 2002 [27]. Data suggest that Tcn1-like LTR retrotransposons were horizontally transmitted between fungi and non-seed plants (indicated by arrows). Presumably HT took place among fungi and the last common ancestor (LCA) of mosses and lycophytes (indicated as 1). Alternatively, it is possible that two independent acts of HT occurred (indicated as 2). First HT event could happen among fungi and LCA of mosses since all investigated mosses contain Tcn1-like LTR retrotransposons. The second HT could occur among fungi and LCA of *Selaginella* since only representatives of this genus carry this group of retrotransposons among all the investigated lycophytes.

PyrTriTy3-2/NecHaemTy3-4 than in comparisons of other LTR retrotransposons (Table 3).

The high similarity, phylogenetic inconsistencies, as well as lower evolutionary rates could be explained by very strict evolutionary constraints or a HT event. However, taking into consideration that the high selective pressure could be implemented only in the case of functional importance of the PyrTriTy3-2 or NecHaemTy3-4, HT looks more preferable for the explanation of the described case. It is known that transposable elements can alter gene expression since they carry their own regulatory sequences and insertions can be selectively advantageous. However, only those transposable elements, which were involved in regulation, evolve under strict selective pressure [44,45].

While extremely rare, horizontal transfer seems to be quite common and recurrent in eukaryotes. An incomplete list of putative HT events includes: HT as a key event in the evolution of several fungal genes [46-48]; HT from fungi to rice weevil *Sitophilus oryzae* proposed for pectinase gene [49]; numerous HT events described for eukaryotic transposable elements [30-38]; as well as HTs of mitochondrial genes, for example, multiple angiosperm-angiosperm HTs of homing group I intron in the mitochondrial *cox1* gene (for a review, see [50]); and a HT of the intron II and two adjacent exons of the mitochondrial *nad1* gene from the flowering plants (angiosperms) to *Gnetum* (gymnosperms) [51].

The actual mechanisms of horizontal transfer for eukaryotic genes and transposable elements are still unknown since it is not possible to show experimentally how HT can occur. Parasites, symbionts, bacteria, or viruses all could be suggested as potential vectors for horizontal transfer. Moreover, based on an example of massive HT from a land plant donor to the basal angiosperm *Amborella trichopoda*, it has been demonstrated that direct plant-to-plant transfer can take a place [52]. The associations between biotrophic fungi and their plant hosts are ubiquitous in nature and range from mutually beneficial to potentially fatal pathogenic interactions. Mycorrhiza refers to an association or symbiosis between plants and fungi that colonize the cortical tissue of plant roots. Ectomycorrhizal fungi are mostly basidiomycetes that grow between root cortical cells of many tree species whereas arbuscular mycorrhizal (AM) fungi belong to the order Glomales (Glomeromycota) and form highly branched structures called arbuscules, within root cortical cells of wide range of land plant species [53-55]. Both types of mycorrhiza represent intimate association and could provide suitable conditions for HT of transposable elements. AM-like mycorrhiza is widely distributed among mosses, ferns and lycophytes (for review, [53]).

Conclusions

Tcn1-like LTR retrotransposons were found in basidiomycota fungi and non-seed plants, including all investigated mosses and lycophytes from genus *Selaginella*. Such interkingdom distribution is not typical for chromodomain-containing LTR retrotransposons clades which are usually very specific for a particular taxonomic group and can be explained by strong selective constraints and the 'retained' genes theory or by horizontal transmission. The close examination and comparison of evolutionary rates for LTR retrotransposons including representatives of Tcn1 and two other clades of LTR retrotransposons, and evolutionary rates estimated for putatively 'retained' genes from mosses and fungi suggests that a horizontal transmission of Tcn1-like LTR retrotransposons took place among fungi and mosses/lycophytes. However the evidence is not strong enough to discard other explanations; such as selective pressure coupled with vertical transmission of retrotransposons in genomes of non-seed plants and loss of these elements by other plants.

Methods

Genomic sequences screening, sequence and phylogenetic analysis

Fungal genomic sequences are available at: Fungal Genome Initiative [56]; The DOE Joint Genome Institute [57]; and The Sanger Institute [58]. The source of individual genomes can be found in table represented in Additional file 5.

We used UniPro uGENE software [59] for LTR retrotransposons identification. The designed pipeline for Gypsy LTR retrotransposons identification and classification included: loading genomic sequence, translation of genomic sequence over six possible reading frames to amino acids, and subsequent search for homologous regions performed using "HMMER search" options of UniPro uGENE. The algorithm of HMMER search is based on profile hidden Markov models, which can perform amino acid sequence searches by use of an appropriate profile [60]. For the analyses, we used a multiple alignment consensus sequence, which contains Gypsy LTR retrotransposon reverse transcriptase (RT) and partial integrase (Int) domains. The profile HMM, based on this consensus sequence, was built using UniPro uGENE software. An additional test for the presence of RT and partial Int domains was performed using BLAST (blastp) which also was incorporated in the designed pipeline. All BLAST analysis was essentially performed using sequence databases accessible from the National Center for Biotechnology Information [61]. The classification of the newly identified elements was performed by a comparative analysis of their sequences. Newly identified ele-

ments and their accession numbers in public databases are listed in Additional file 6.

The whole nucleotide sequences of the transposable elements, if possible, were also extracted with the assistance of UniPro uGENE software. After localization of amino acid sequences obtained during HMMER search in the initial genomes in its nucleotide representation, the sequences were expanded up to 15 Kb and used for long terminal repeats (LTRs) search. The algorithm for repeats search, 'Repeat Find', is included to the UniPro uGENE as well as the visualization feature and 'ORF Find' option which were used to identify the putatively intact copies of LTR retrotransposons. Structural features of newly identified LTR retrotransposons can be found in Additional file 1.

Tcn1-like LTR retrotransposon search was carried out using BLAST (blastp and blastx). BLAST analysis was performed using sequence databases accessible from the National Center for Biotechnology Information (NCBI) server [59], The U.S. Department of Energy Joint Genome Institute [57], and Broad Institute of MIT and Harvard [56] as well as Phytosome, a tool for green plant comparative genomics [62]. The described copy of SM-Tcn1 from spikemoss *Selaginella moellendorffii* (Lycopodiophyta) is located in scaffold_0 (1426925-1421008) of genomic sequence version 1.0 which is available at The U.S. Department of Energy Joint Genome Institute website [57]. The whole sequence of SM-Tcn1 with annotations can be found in Additional file 3. Other websites used in the present study were: Repbase [63], NCBI conserved domain database and search service [64], ESTs from *Porphyra yezoensis* at Kazusa DNA Research Institute [65], *Cyanidioschyzon merolae* Genome Project [66], The Plant Genomics Consortium [67], The Institute for Genomic Research [68], Cassava and Leafy Spurge EST Project [69].

All multiple DNA alignments were performed by ClustalW [70] and edited manually in UniPro uGENE. Phylogenetic analyses were performed using the Neighbor-Joining (NJ) method in MEGA 4.0 program [71]. Statistical support for the NJ tree was evaluated by bootstrapping (number of replications, 1000) [72]. Evolutionary rates were estimated by standard methods [73]. Poisson correction distances (d) were estimated from the equation $d = -\ln(1 - p)$, where p represents the proportion of different amino acids. The rate of amino acid substitution (r) was estimated by the standard equation $r = d/2T$, where T is the divergence time of the last common ancestor of the compared species. The estimated divergence times used were: Plants/Fungi, 1500 Myr and Basidiomycetes/Ascomycetes, 1200 Myr according to Hedges (2002) [27]; Homobasidiomycetes/Chytridiomycetes, 900 Myr, Sordaryomycetes/Eurotiomycetes, 540 Myr, and

Sordaryomycetes/Dothideomycetes, 490 Myr according to Padovan et al. (2005) [74]; and Heterobasidiomycetes (or Tremellomycetes)/Homobasidiomycetes (Agaricomycetes and Dacrymycetes), 700 Myr according to Hibbett et al. (2007) [75] and Taylor et al. (2004) [76].

Species collection and total DNA isolation

Table 2 lists plant species, and Table 3 lists fungal species used in present study. The taxonomy of vascular non-seed plants (monilophytes and lycophytes) is given after Pryer et al. (2004) [22], Smith et al. (2006) [21], and Korall et al. (2007) [77]. Plant species (monilophytes) were collected in nature. The detailed label data are available from the authors. The genomic DNA of lycophytes was provided by the Royal Botanic Gardens, Kew, London, UK [78]. Genomic DNA was isolated from the leaves. Extraction was performed using the QIAGEN DNeasy Plant Mini Kit (QIAGEN). Isolated DNA was used directly in PCR amplifications.

Gypsy LTR retrotransposons PCR amplification and sequencing

Previously designed degenerate PCR primers for chromodomain-containing Gypsy LTR retrotransposons were used in present study: GyRT1 = 5'-MRNATGTGYGT-NGAYTAYMG-3' [24] and ty3-A = 5'-AATTTCGCTGC-CGCTAAGATNARNADRTRC-3' [8], where M = A + C, Y = C + T, R = A + G, D = A + G + T and N = A + G + C + T. These primers were designed to amplify the most conserved part of the reverse transcriptase (RT) domain of LTR retrotransposons and were proved to be efficient [8,24]. The expected length of PCR products was about 320 bp. PCR amplification with degenerate primers was performed using 0.1 μ g of genomic DNA in 10- μ l volume of 10 mM Tris-HCl (pH 8.9), 1 mM $(\text{NH}_4)_2\text{SO}_4$, 4 mM MgCl_2 , 200 μ M each of four dNTPs, 0.5 μ M primers, and 2.5 units of Taq polymerase. After an initial denaturation step for 3 min at 94°C, the PCR reactions were subjected to 30 cycles of amplification consisting of 30 sec denaturation at 94°C, 42 sec annealing at 50°C, and 1 min extension at 72°C. PCR products were separated by agarose gel electrophoresis. The resulting PCR products were directly ligated into a pGEM vector using a pGEM-T-Easy cloning kit (Promega) for sequence determination.

Clones were amplified by PCR with M13 primers, and 40 ng of the product was used in a 10 μ l cycle sequencing reaction with the ABI BigDye Terminator Kit on an ABI 310 Genetic Analyser (Applied Biosystems); or sequencing reactions were performed with Dye Terminator Cycle Sequencing Kit (Beckman Coulter) and analyzed on CEQ 8000 Genetic Analysis System. Sequences were deposited to GenBank under Acc. Numbers [GQ443314-GQ443445](#) and [AY959294-AY959313](#).

List of abbreviations used

LTR: long terminal repeat; SINE: short interspersed nuclear element; CHD: chromodomain; RT: reverse transcriptase; Int: integrase; ORF: open reading frame; PR: proteinase; dUTPase: deoxyuridine triphosphatase domain; PPT: polypurine tract; PBS: primer-binding site; HT: horizontal transfer or horizontal transmission; TE: transposable element; LCA: last common ancestor; AM: arbuscular mycorrhiza; HMM: hidden Markov models; Myr: million years; Mya: million years ago.

Additional material

Additional file 1 Structure of novel LTR retrotransposons from Fungi.

Table contained list of novel LTR retrotransposons from Fungi detected in present study, their copy numbers and putative structure including predicted enzymatic domains.

Additional file 2 Phylogenetic analysis of dUTPase. Neighbor-Joining phylogenetic tree reconstructed based on dUTPase amino acid sequences from eukaryotes, viruses, and dUTPase domains from CHD-containing Gypsy LTR retrotransposons.

Additional file 3 SM-Tcn1 CHD-containing Gypsy LTR retrotransposon. Sequence of SM-Tcn1 CHD-containing Gypsy LTR retrotransposon from spikemoss *Selaginella moellendorffii* (Lycopodiophyta) with annotations in GenBank format.

Additional file 4 Phylogenetic analysis of Athila-like LTR retrotransposons. Neighbor-joining (NJ) phylogenetic tree based on RT nucleotide sequences of Athila-like LTR retrotransposons including newly described elements.

Additional file 5 List of fungal species, genomes of which were analyzed. Table contained the list of fungal species, genomes of which were analyzed *in silico* in the present study and the sources of genomic sequences.

Additional file 6 Novel Gypsy LTR retrotransposons from Fungi. Table contained the list of novel Gypsy LTR retrotransposons from Fungi detected in present study and their accession numbers.

Authors' contributions

ON participated in the design of the study, carried out the analysis, participated in the sequence analysis and drafted the manuscript. GS contributed to data acquisition, participated in analysis and interpretation. AB participated in the design of the study, performed coordination and has given final approval of the submitted version. All authors read and approved the final manuscript.

Acknowledgements

Author thanks Dr. Mark L. Farman (Department of Plant Pathology, University of Kentucky, USA) for the helpful comments and Dr. David Thornbury (Department of Plant Pathology, University of Kentucky, USA) for his stylistic suggestions. This work was supported by the Russian Foundation for Basic Research (grant number RFBR 09-04-00360-a) and by state contract 10002-251/П-25/155-270/200404-082 and Siberian Branch of the Russian Academy of Sciences (project No. 10.4).

The sequence data for *Laccaria bicolor*, *Trichoderma reesei*, *Trichoderma virens*, *Aspergillus niger*, *Stagonospora nodorum*, *Postia placenta* and *Sporobolomyces roseus* were produced by the US Department of Energy Joint Genome Institute <http://www.jgi.doe.gov/>. Preliminary sequence data for *Alternaria brassicicola* were obtained from Genome Sequencing Center at Washington University Medical School <http://genome.wustl.edu/>.

The genomic DNA of lycophytes was provided by the Royal Botanic Gardens, Kew, London, UK <http://www.kew.org>. Monilophytes material was kindly provided by Dr. Alexander Shmakov (Altai State University, Barnaul, Russia) and Dr. Elena Korolyuk (Central Siberian Botanical Garden, Novosibirsk, Russia).

Author Details

¹Laboratory of Molecular Genetic Systems, Institute of Cytology and Genetics, Novosibirsk, Russia and ²Novosibirsk State University, Novosibirsk, Russia

Received: 15 October 2009 Accepted: 8 April 2010

Published: 8 April 2010

References

1. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhou B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH: **A unified classification system for eukaryotic transposable elements.** *Nat Rev Genet* 2007, **8**:973-982.
2. Marin I, Llorens C: **Ty3/Gypsy retrotransposons: description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data.** *Mol Biol Evol* 2000, **17**:1040-1049.
3. Gorinsek B, Gubensek F, Kordis D: **Evolutionary genomics of chromoviruses in eukaryotes.** *Mol Biol Evol* 2004, **21**:781-798.
4. Paro R, Hogness DS: **The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*.** *Proc Natl Acad Sci USA* 1991, **88**:263-267.
5. Koonin EV, Zhou S, Lucchesi JC: **The chromo superfamily: new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin.** *Nucleic Acids Res* 1995, **23**:4229-4233.
6. Brehm A, Tufeland KR, Aasland R, Becker PB: **The many colours of chromodomains.** *Bioessays* 2004, **26**:133-140.
7. Gao X, Hou Y, Ebina H, Levin HL, Voytas DF: **Chromodomains direct integration of retrotransposons to heterochromatin.** *Genome Res* 2008, **18**:359-369.
8. Novikova O, Mayorov V, Smyshlyaev G, Fursov M, Adkison L, Pisarenko O, Blinov A: **Novel clades of chromodomain-containing Gypsy LTR retrotransposons from mosses (Bryophyta).** *Plant J* 2008, **56**:562-574.
9. Novikova O: **Chromodomains and LTR retrotransposons in plants.** *Comm & Integr Biol* 2009, **2**:158-162.
10. Neuvéglise C, Feldmann H, Bon E, Gaillardin C, Casaregola S: **Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts.** *Genome Res* 2002, **12**:930-943.
11. Schmid-Berger N, Schmid B, Barth G: **Ylt1, a highly repetitive retrotransposon in the genome of the dimorphic fungus *Yarrowia lipolytica*.** *J Bacteriol* 1994, **176**:2477-2482.
12. Goodwin TJ, Poulter RT: **Multiple LTR-retrotransposon families in the asexual yeast *Candida albicans*.** *Genome Res* 2000, **10**:174-191.
13. Goodwin TJ, Poulter RT: **The diversity of retrotransposons in the yeast *Cryptococcus neoformans*.** *Yeast* 2001, **18**:865-880.
14. Novikova OS, Blinov AG: **dUTPase-containing Metaviridae LTR retrotransposons from the genome of Phanerochaete chrysosporium (Fungi: Basidiomycota).** *Dokl Biochem Biophys* 2008, **420**:146-149.
15. Riccioni C, Rubini A, Belfiori B, Passeri V, Paolucci F, Arcioni S: **Tmt1: the first LTR-retrotransposon from a Tuber spp.** *Curr Genet* 2008, **53**:23-34.
16. Payne SL, Elder JH: **The role of retroviral dUTPases in replication and virulence.** *Curr Protein Pept Sci* 2001, **2**:381-388.
17. Varmus H, Brown P: **Retroviruses.** In *Mobile DNA* Edited by: Berg DE, Howe MM. Washington: American Society for Microbiology; 1989:53-108.
18. Wilhelm M, Wilhelm FX: **Reverse transcription of retroviruses and LTR retrotransposons.** *Cell Mol Life Sci* 2001, **58**:1246-1262.
19. Levin HL: **A novel mechanism of self-primed reverse transcription defines a new family of retroelements.** *Mol Cell Biol* 1995, **15**:3310-3317.
20. Wang W, Tanurdzic M, Luo M, Sisneros N, Kim HR, Weng JK, Kudrna D, Mueller C, Arumuganathan K, Carlson J, Chapple C, de Pampillis C, Mandoli D, Tomkins J, Wing RA, Banks JA, et al.: **Construction of a bacterial artificial chromosome library from the spikemoss *Selaginella moellendorffii*: a new resource for plant comparative genomics.** *BMC Plant Biol* 2005, **5**:10.
21. Smith AR, Pryer KM, Schuettpeiz E, Korall P, Schneider H, Wolf PG: **A classification for extant ferns.** *Taxon* 2006, **55**:705-731.
22. Pryer KM, Schuettpeiz E, Wolf PG, Schneider H, Smith AR, Cranfill R: **Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences.** *Am J Bot* 2004, **91**:1582-1598.
23. Wikstrom N, Kenrick P: **Evolution of Lycopodiaceae (Lycopsidea): estimating divergence times from rbcL gene sequences by use of nonparametric rate smoothing.** *Mol Phylog Evol* 2001, **19**:177-186.

24. Friesen N, Brandes A, Heslop-Harrison JS: **Diversity, origin, and distribution of retrotransposons (gypsy and copia) in conifers.** *Mol Biol Evol* 2001, **18**:1176-1188.
25. Wright DA, Voytas DF: **Potential retroviruses in plants: Tat1 is related to a group of Arabidopsis thaliana Ty3/gypsy retrotransposons that encode envelope-like proteins.** *Genetics* 1998, **149**:703-715.
26. Vitte C, Panaud O: **LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model.** *Cytogenet Genome Res* 2005, **110**:91-107.
27. Hedges SB: **The origin and evolution of model organisms.** *Nat Rev Genet* 2002, **3**:838-849.
28. Rensing SA, Fritzwosky D, Lang D, Reski R: **Protein encoding genes in an ancient plant: analysis of codon usage, retained genes and splice sites in a moss, *Physcomitrella patens*.** *BMC Genomics* 2005, **6**:43.
29. Hartl DL, Lohe AR, Lozovskaya ER: **Modern thoughts on an ancient mariner: function, evolution, regulation.** *Annu Rev Genet* 1997, **31**:337-358.
30. Jordan IK, Matyunina LV, McDonald JF: **Evidence for the recent horizontal transfer of long terminal repeat retrotransposon.** *Proc Natl Acad Sci USA* 1999, **96**:12621-12625.
31. Novikova O, Sliwińska E, Fet V, Settele J, Blinov A, Woyciechowski M: **CR1 clade of non-LTR retrotransposons from Maculinea butterflies (Lepidoptera: Lycaenidae): evidence for recent horizontal transmission.** *BMC Evol Biol* 2007, **7**:93.
32. Roulin A, Piegu B, Wing RA, Panaud O: **Evidence of multiple horizontal transfers of the long terminal repeat retrotransposon RIRE1 within the genus *Oryza*.** *Plant J* 2008, **53**:950-959.
33. Roulin A, Piegu B, Fortune PM, Sabot F, D'Hont A, Manicacci D, Panaud O: **Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfers of the LTR-retrotransposon Route66 in Poaceae.** *BMC Evol Biol* 2009, **9**:58.
34. Kidwell MG: **Horizontal transfer.** *Curr Opin Genet Dev* 1992, **2**:868-873.
35. Volff JN, Körtling C, Schartl M: **Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes.** *Mol Biol Evol* 2000, **17**:1673-1684.
36. Kordis D, Gubensek F: **Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes.** *Proc Natl Acad Sci USA* 1998, **95**:10704-10709.
37. Novikova O, Fet V, Blinov A: **Non-LTR retrotransposons in fungi.** *Funct Integr Genomics* 2009, **9**:27-42.
38. Kaplan N, Darden T, Langley CH: **Evolution and extinction of transposable elements in Mendelian populations.** *Genetics* 1985, **109**:459-480.
39. Lohe AR, Moriyama EN, Lidholm DA, Hartl DL: **Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements.** *Mol Biol Evol* 1995, **12**:62-72.
40. Bowman JL, Floyd SK, Sakakibara K: **Green genes-comparative genomics of the green branch of life.** *Cell* 2007, **129**:229-234.
41. Berbee ML, Taylor JW: **Fungal Molecular Evolution: Gene Trees and Geologic Time.** In *The Mycota: a comprehensive treatise on fungi as experimental systems for basic and applied research. Systematics and Evolution, Part B Volume VII*. Edited by: McLaughlin DJ, McLaughlin EG, Lemke PA. New York: Springer-Verlag; 2001:229-246.
42. Kaneko I, Tanaka A, Tsuge T: **REAL, an LTR retrotransposon from the plant pathogenic fungus *Alternaria alternata*.** *Mol Gen Genet* 2000, **263**:625-634.
43. Taylor EJ, Konstantinova P, Leigh F, Bates JA, Lee D: **Gypsy-like retrotransposons in Pyrenophora: an abundant and informative class of molecular markers.** *Genome* 2004, **47**:519-525.
44. Medstrand P, Landry JR, Mager DL: **Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans.** *J Biol Chem* 2001, **276**:1896-1903.
45. Ono R, Kobayashi S, Wagatsuma H, Aisaka K, Kohda T, Kaneko-Ishino T, Ishino F: **A retrotransposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21.** *Genomics* 2001, **73**:232-237.
46. Wenzl P, Wong L, Kwang-won K, Jefferson RA: **A functional screen identifies lateral transfer of beta-glucuronidase (gus) from bacteria to fungi.** *Mol Biol Evol* 2005, **22**:308-316.
47. Slot JC, Hibbett DS: **Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: a phylogenetic study.** *PLoS ONE* 2007, **2**:e1097.
48. Khaldi N, Collemare J, Lebrun MH, Wolfe KH: **Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi.** *Genome Biol* 2008, **9**:R18.
49. Shen Z, Denton M, Mutti N, Pappan K, Kanost MR, Reese JC, Reeck GR: **Polygalacturonase from *Sitophilus oryzae*: possible horizontal transfer of a pectinase gene from fungi to weevils.** *J Insect Sci* 2003, **3**:24.
50. Richardson AO, Palmer JD: **Horizontal gene transfer in plants.** *J Exp Bot* 2007, **58**:1-9.
51. Won H, Renner SS: **Horizontal gene transfer from flowering plants to *Gnetum*.** *Proc Natl Acad Sci USA* 2003, **100**:10824-10829.
52. Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD: **Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*.** *Proc Natl Acad Sci USA* 2004, **101**:17747-17752.
53. Albrecht C, Geurts R, Bisseling T: **Legume nodulation and mycorrhizae formation; two extremes in host specificity meet.** *EMBO J* 1999, **18**:281-288.
54. Wang B, Qiu YL: **Phylogenetic distribution and evolution of mycorrhizas in land plants.** *Mycorrhiza* 2006, **16**:299-363.
55. Hibbett DS, Matheny PB: **The relative ages of ectomycorrhizal mushrooms and their plant hosts estimated using Bayesian relaxed molecular clock analyses.** *BMC Biol* 2009, **7**:13.
56. Broad Institute [<http://www.broad.mit.edu>]
57. The DOE Joint Genome Institute [<http://www.jcvi.org/>]
58. The Sanger Institute [<http://www.sanger.ac.uk/>]
59. UniPro uGENE software [<http://genome.unipro.ru>]
60. Eddy SR: **A probabilistic model of local sequence alignment that simplifies statistical significance estimation.** *PLoS Comput Biol* 2008, **4**:e1000069.
61. National Center for Biotechnology Information [<http://www.ncbi.nlm.nih.gov/>]
62. Phytozome, a tool for green plant comparative genomics [<http://www.phytozome.net/>]
63. Repbase [<http://www.girinst.org/>]
64. National Center for Biotechnology Information conserved domain database and search service [<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>]
65. ESTs from *Porphyra yezoensis* at Kazusa DNA Research Institute [<http://est.kazusa.or.jp/>]
66. Cyanidioschyzon merolae Genome Project [<http://merolae.biol.s.u-tokyo.ac.jp/>]
67. The Plant Genomics Consortium [<http://nypg.bio.nyu.edu/>]
68. The Institute for Genomic Research [<http://www.tigr.org/>]
69. Cassava and Leafy Spurge EST Project [<http://titan.biotech.uiuc.edu/cassava/>]
70. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
71. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**:1596-1599.
72. Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap.** *Evolution* 1985, **39**:783-791.
73. Nei M, Kumar S: *Molecular evolution and phylogenetics* New York: Oxford University Press; 2000.
74. Padovan AC, Sanson GF, Brunstein A, Briones MR: **Fungi evolution revisited: application of the penalized likelihood method to a Bayesian fungal phylogeny provides a new perspective on phylogenetic relationships and divergence dates of Ascomycota groups.** *J Mol Evol* 2005, **60**:726-735.
75. Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Lücking R, Thorsten Lumbsch H, Lutzoni F, Matheny PB, McLaughlin DJ, Powell MJ, Redhead S, Schoch CL, Spatafora JW, Stalpers JA, Vilgalys R, Aime MC, Aptroot A, Bauer R, Begerow D, Benny GL, Castlebury LA, Crous PW, Dai YC, Gams W, Geiser DM, Griffith GW, Gueidan C, Hawksworth DL, Hestmark G, Hosaka K, Humber RA, Hyde KD, Ironside JE, Kõljalg U, Kurtzman CP, Larsson KH, Lichtwardt R, Longcore J, Miadlikowska J, Miller A, Moncalvo JM, Mozley-Standridge S, Oberwinkler F, Parmasto E, Reeb V, Rogers JD, Roux C, Ryvarden L, Sampaio JP, Schüssler A, Sugiyama J, Thorn RG, Tibell L, Untereiner WA, Walker C, Wang

- Z, Weir A, Weiss M, White MM, Winka K, Yao YJ, Zhang N: **A higher-level phylogenetic classification of the Fungi.** *Mycol Res* 2007, **111**:509-547.
76. Taylor JW, Spatafora J, O'Donnell K, Lutzoni F, James T, Hibbett DS, Geiser D, Bruns TD, Blackwell M: **The relationships of fungi.** In *Assembling the tree of life* Edited by: Cracraft J, Donoghue MJ. New York: Oxford University Press; 2004:171-196.
77. Korall P, Conant DS, Metzgar JS, Schneider H, Pryer KM: **A molecular phylogeny of scaly tree ferns (Cyatheaceae).** *Am J Bot* 2007, **94**:873-886.
78. **The Royal Botanic Gardens, Kew, London, UK** [<http://www.kew.org>]

doi: 10.1186/1471-2164-11-231

Cite this article as: Novikova *et al.*, Evolutionary genomics revealed interkingdom distribution of Tcn1-like chromodomain-containing Gypsy LTR retrotransposons among fungi and plants *BMC Genomics* 2010, **11**:231

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

