Horizontal Transfer and Evolution of Prokaryote Transposable Elements in Eukaryotes

Clément Gilbert* and Richard Cordaux

Université de Poitiers, UMR CNRS 7267 Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, Poitiers, France

*Corresponding author: E-mail: clement.gilbert@univ-poitiers.fr.

Accepted: April 2, 2013

Abstract

Horizontal transfer (HT) of transposable elements (TEs) plays a key role in prokaryotic evolution, and mounting evidence suggests that it has also had an important impact on eukaryotic evolution. Although many prokaryote-to-prokaryote and eukaryote-to-eukaryote HTs of TEs have been characterized, only few cases have been reported between prokaryotes and eukaryotes. Here, we carried out a comprehensive search for all major groups of prokaryotic insertion sequences (ISs) in 430 eukaryote genomes. We uncovered a total of 80 sequences, all deriving from the IS607 family, integrated in the genomes of 14 eukaryote species belonging to four distinct phyla (Amoebozoa, Ascomycetes, Basidiomycetes, and Stramenopiles). Given that eukaryote IS607-like sequences are most closely related to cyanobacterial IS607 and that their phylogeny is incongruent with that of their hosts, we conclude that the presence of IS607-like sequences in eukaryotic genomes is the result of several HT events. Selection analyses further suggest that our ability to detect these prokaryote TEs today in eukaryotes is because HT of these sequences occurred recently and/or some IS607 elements were domesticated after HT, giving rise to new eukaryote genes. Supporting the recent age of some of these HTs, we uncovered intact full-length, potentially active IS607 copies in the amoeba *Acanthamoeba castellani*. Overall, our study shows that prokaryote-to-eukaryote HT of TEs occurred at relatively low frequency during recent eukaryote evolution and it sets IS607 as the most widespread TE (being present in prokaryotes, eukaryotes, and viruses).

Key words: horizontal transfer, transposable elements, eukaryotes, prokaryotes.

Introduction

Horizontal transfer (HT) of DNA is the movement of genetic information between nonmating organisms. The frequency, evolutionary consequences, and the mechanisms underlying this phenomenon are well understood in prokaryotes (Frost et al. 2005; Bichsel et al. 2010). The rates of HT are such in this group that evolutionary relationships may often be better represented as a network rather than as a simple bifurcating tree (Bapteste et al. 2009; Puigbò et al. 2010). Bacteria indeed possess a large arsenal of molecular vehicles, including plasmids (Smillie et al. 2010), integrative and conjugative elements (Wozniak and Waldor 2010), and gene transfer agents (Lang et al. 2012) that allow them to efficiently exchange various gene sets often crucial to their adaptation to new environments (Ochman et al. 2000).

In eukaryotes, no mechanism dedicated to HT has been described so far, and the presence of a nuclear envelope and the soma/germ division (in metazoans) are thought to constitute strong barriers to HT. In addition, the patterns and mechanisms underlying eukaryotic HT have been less thoroughly studied than in prokaryotes, owing to methodological limitations imposed by larger genome sizes. However, a relatively large number of HTs have been characterized in eukaryotes, many of which resulted in evolutionary innovations pivotal to adaptation and colonization of new environments (Andersson 2005; Keeling and Palmer 2008; Keeling 2009).

In both prokaryotes and eukaryotes, gene HT has long received far more attention than HT of nongenic DNA, likely because genes are generally better characterized than other parts of the genome, and the biological significance of gene HT can be more directly assessed than that of nongenic DNA. Interestingly, however, transposable elements (TEs), the main component of nongenic DNA in most genomes, exhibit several properties that make them better candidates to successful HT than genes. Contrary to bona fide, static genes, TEs are able to move from a chromosomal locus to another, often duplicating themselves in the process (Craig et al. 2002). They often are the single most abundant component of eukaryotic genomes, reaching, for example, approximately 45%

and approximately 85% of the human and maize genomes, respectively (Lander et al. 2001; Schnable et al. 2009). In addition, a TE landing into a new genome through HT can constitute a powerful source of variation (Oliver and Greene 2009). However, unlike genes, TEs generally do not encode any function beneficial to the genome, and their movement and proliferation can also have various negative effects (Cordaux and Batzer 2009), decreasing the likelihood of successful HT. Though no study has yet compared the rates and evolutionary consequences of gene HT versus TE HT, it is now established that, much like gene HT, TE HT has occurred recurrently, across multiple phylogenetic scales both within prokaryotes and eukaryotes (Touchon and Rocha 2007; Wagner and de la Chaux 2008; Schaack et al. 2010; Cerveau et al. 2011; Syvanen 2012; Wallau et al. 2012) and that such HTs have sometimes been pervasive (Sánchez-Gracia et al. 2005; Thomas et al. 2010; Gilbert et al. 2010).

HT between domains of life is an intriguing exception to this pattern. Although hundreds of cases of prokaryote-to-eukaryote gene HTs have been characterized (e.g., Marcet-Houben and Gabaldón 2010; Moran et al. 2012; reviewed in Keeling and Palmer 2008), very few TEs are known to have jumped from prokaryotes to eukaryotes. One likely such event has been reported in the bdelloid rotifer *Adineta vaga*, in which a single-copy TE showing similarity to the IS5 family of prokaryotic TEs was found to be transcriptionally silent (Gladyshev and Arkhipova 2009). Novo et al. (2009) noted that the genome of the yeast *Saccharomyces cerevisiae* (strain EC1118) contains an apparent pseudogene similar to a gene found in another yeast, *Ashbya gossypii,* that itself encodes a bacterial transposase. In another study, Rolland et al. (2009) characterized six genes in the genome of the yeast *Lachancea kluyveri* that likely originated through segmental duplications following HT of a bacterial TE related to the IS607 family. Finally, several bacterial TEs were found embedded in larger genomic fragments transferred horizontally from bacterial endosymbionts to their eukaryotic hosts (Dunning Hotopp 2011). However, in these cases, the presence of these elements in eukaryotic genomes is not the result of TE HT per se.

Whether the extreme paucity of known prokaryote-to-eukaryote TE HTs compared with gene HTs has true biological underpinnings is still unclear because no systematic prokaryote-to-eukaryote TE HT investigation has been conducted so far. In this study, we performed a large-scale search of prokaryotic insertion sequences (ISs) in eukaryotic genomes. We focused on IS elements because 1) they are the most abundant and widespread prokaryotic TEs (Siguier et al. 2006); 2) they are frequently involved in HT among prokaryotes (e.g., Touchon and Rocha 2007; Wagner and de la Chaux 2008; Cerveau et al. 2011); and 3) their structure and diversity are well characterized (Siguier et al. 2006).

Materials and Methods

Uncovering IS-Like Sequences in Eukaryote Genomes

A search for eukaryotic sequences exhibiting significant homology to prokaryotic IS TEs was conducted as follows. First, we extracted the amino acid (aa) sequence of 48 ISs representing all 24 known IS families and most IS subgroups from the IS reference database ISfinder (Siguier et al. 2006). Using this library as a query (provided in supplementary data set S1, Supplementary Material online), we performed tBLASTn (Basic Local Alignment Search Tool [BLAST]) searches against all eukaryote whole-genome sequences available in GenBank ($n = 431$ as of July 2012; fig. 1). This search yielded 1,515 hits longer than 50 aa with an *e* value $\leq 10^{-6}$ that were subjected to further analysis.

Ruling Out Contamination

An important issue to consider when studying HT is that sequences that apparently look like they were horizontally transferred could instead be the result of contamination. This is especially true when dealing with prokaryote-to-eukaryote HT because many eukaryotic taxa live in close association with bacterial organisms (e.g., endosymbionts), genomic DNA of which can be difficult to completely separate from host DNA before extraction. Furthermore, it has been reported that IS elements residing in the genomes of *Escherichia coli* strains used for cloning can transpose into cloned inserts in the laboratory; as a result, a substantial number of eukaryotic whole-genome sequences contain IS insertions that are experimental artifacts (Astua-Monge et al. 2002; Senejani and Sweasy 2010).

To rule out contamination, we first assessed whether the genomic sequences flanking the 1,515 candidate IS-like sequences were of eukaryotic or prokaryotic origin. We extracted 1.5 kb of upstream and downstream flanking sequences for each candidate and used them as queries in BLASTx searches against all protein sequences available in GenBank (nr database downloaded in April 2012). We discarded all IS-like sequences flanked by prokaryotic proteins. Some IS-like sequences had no similarity to any known protein in their flanks. To be conservative, these sequences were also excluded. In the end, we only retained IS-like sequences that showed some similarity to a known eukaryotic protein in their flanking region. Importantly, none of the retained IS-like sequences was identical or nearly identical to a known IS, as would be expected if they resulted from insertion into a cloned insert during the cloning process (Astua-Monge et al. 2002; Senejani and Sweasy 2010). IS-like sequences flanked by eukaryotic proteins were used as queries to perform additional BLASTn searches to extract all copies similar to the queries from the queried genome. The genomic coordinates of all IS-like sequences considered for further analysis (either flanked by eukaryotic proteins or similar to IS-like sequences

flanked by eukaryotic proteins in the same genome) are given in supplementary table S1, Supplementary Material online.

We also verified the presence of selected IS-like sequences by polymerase chain reaction (PCR) and sequencing in two taxa (*Acanthamoeba castellani* and *Phytophthora ramorum*) representing two distinct eukaryotic phyla (Amoebozoa and Stramenopiles). We designed primers in the genomic regions flanking 12 IS-like sequences (supplementary data set S2, Supplementary Material online) and carried out PCRs on genomic DNA of the strains that were originally used to produce the whole-genome sequence data available in GenBank, that is, ATCC MYA-2949 (Pr-102) for *P. ramorum* and ATCC 30010D (Neff) for *Aca. castellani*. PCRs were conducted using the following temperature cycling: initial denaturation at 94 °C for 5 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 54–58 °C (depending on the primer set) for 30 s, and elongation at 72 °C for 1 min, ending with a 10 min elongation step at 72 °C. Purified PCR products were directly sequenced using ABI BigDye sequencing mix (1.4 μl template PCR product, 0.4 μl BigDye, 2 μl manufacturer supplied buffer, 0.3 μl primer, and 6 μl H_2O). Sequencing reactions were ethanol precipitated and run on an ABI 3730 sequencer. Presence and sequences of all selected IS-like sequences were confirmed as predicted in silico. Altogether, we conclude that our final set of 80 IS-like sequences is highly unlikely to result from contamination artifacts.

Sequence Analyses

IS-like sequences retained for analysis were aligned at the aa level by hand using BioEdit 7.0.5.3 (Hall 2004). Phylogenetic analyses were carried out after removing ambiguous regions from the alignment using PhyML 3.0 (Guindon and Gascuel 2003). The alignment of all eukaryotic IS-like sequences uncovered in this study together with their flanking regions is provided in supplementary data set S3, Supplementary Material online. The alignments used to perform the phylogenetic analyses are provided in supplementary data sets S4 and S5, Supplementary Material online. Models of aa evolution best fitting the two alignments subjected to phylogenetic analyses were chosen using the Akaike information criterion in ProtTest 3.0 (Darriba et al. 2011). Analyses of selection were carried out using the GA-Branch method (Kosakovsky Pond and Frost 2005) in the HyPhy package (Kosakovsky Pond et al. 2005) implemented on the Datamonkey server (Delport et al. 2010).

Results

IS607-Like Sequences in Multiple Eukaryote Taxa

Our search yielded 80 sequences resembling prokaryotic IS in 14 different species belonging to four distinct eukaryotic phyla: Amoebozoa, Ascomycetes, Basidiomycetes, and Stramenopiles (fig. 1). The number of IS-like sequences per genome varied from 1 (in the brown algae *Ectocarpus siliculosus* and the yeasts *Ash. gossypii* and *S. cerevisiae*) to 22 (in the oomycete *P. ramorum*). Although our IS library used to query eukaryotic genomes included representatives of all known IS families, the 80 sequences that came out of our search were all most similar to a single IS family, namely IS607. We did not recover the IS5-like sequence described by Gladyshev and Arkhipova (2009) in the bdelloid rotifer *A. vaga* because whole-genome sequence is not available for this species. However, our search independently identified the six IS607-like sequences reported by Rolland et al. (2009) in *L. kluyveri* and the IS-like sequence uncovered in *Ash. gossypii* and *S. cerevisiae* (EC1118) by Novo et al. (2009). Furthermore, we found the IS-like sequence identified by Novo et al. (2009) in two additional *S. cerevisiae* strains (EC9 and LALVIN), and we report here that these yeast IS-like sequences all derive from the IS607 family.

Structurally, IS607 is made of two overlapping open reading frames in prokaryotes (ORFA and ORFB; Kersulyte et al. 2000). Among prokaryotic IS607 elements, those found in Cyanobacteria are the most similar to eukaryotic IS-like sequences retrieved in our study. A conserved domain search using ISArma1 (a canonical IS607 described in the cyanobacterium *Arthrospira maxima*) as a query in Pfam (Punta et al. 2012) indicates that 1) ORFA contains a N-terminal helix-turn-helix (HTH) DNA-binding domain related to the MerR family of regulatory proteins, followed by a resolvase domain and 2) ORFB is made of a N-terminal HTH DNA-binding domain, followed by a transposase domain and a C-terminal Zn ribbon DNA-binding domain (fig. 2). Several of the eukaryotic IS-like sequences match the entire IS607 sequence with various levels of aa similarity (fig. 2): from 42% to 51% over 590 aa in *Ect. siliculosus* and *Aca. castellani*. Most eukaryotic IS-like sequences, however, correspond to ORFA only, ORFB only, or fragments of ORFA or ORFB (fig. 2). Overall, the average level of aa similarity between all 80 eukaryotic IS-like sequences and their most similar prokaryotic IS607 element is 51% over an average length of 189 aa and aa similarity ranges from 42% over 408 aa to 68% over 64 aa (see supplementary table 2, Supplementary Material online, for more details).

HT of IS607-Like Sequences

The occurrence of bacterial IS-like sequences in eukaryote genomes begs the question of whether these sequences are the product of HT from prokaryotes to eukaryotes or whether they were vertically inherited from a protoeukaryote ancestor. It is noteworthy that five superfamilies of eukaryote transposase-carrying TEs (i.e., class II TEs), namely Tc1/mariner, Mutator, Merlin, PIF-Harbinger, and ISL2EU/IS4EU, are known to be evolutionary linked to prokaryotic IS families IS630, IS256, IS1595, IS5, and IS4, respectively (Doak et al. 1994; Eisen et al. 1994; Kapitonov and Jurka 1999, 2007;

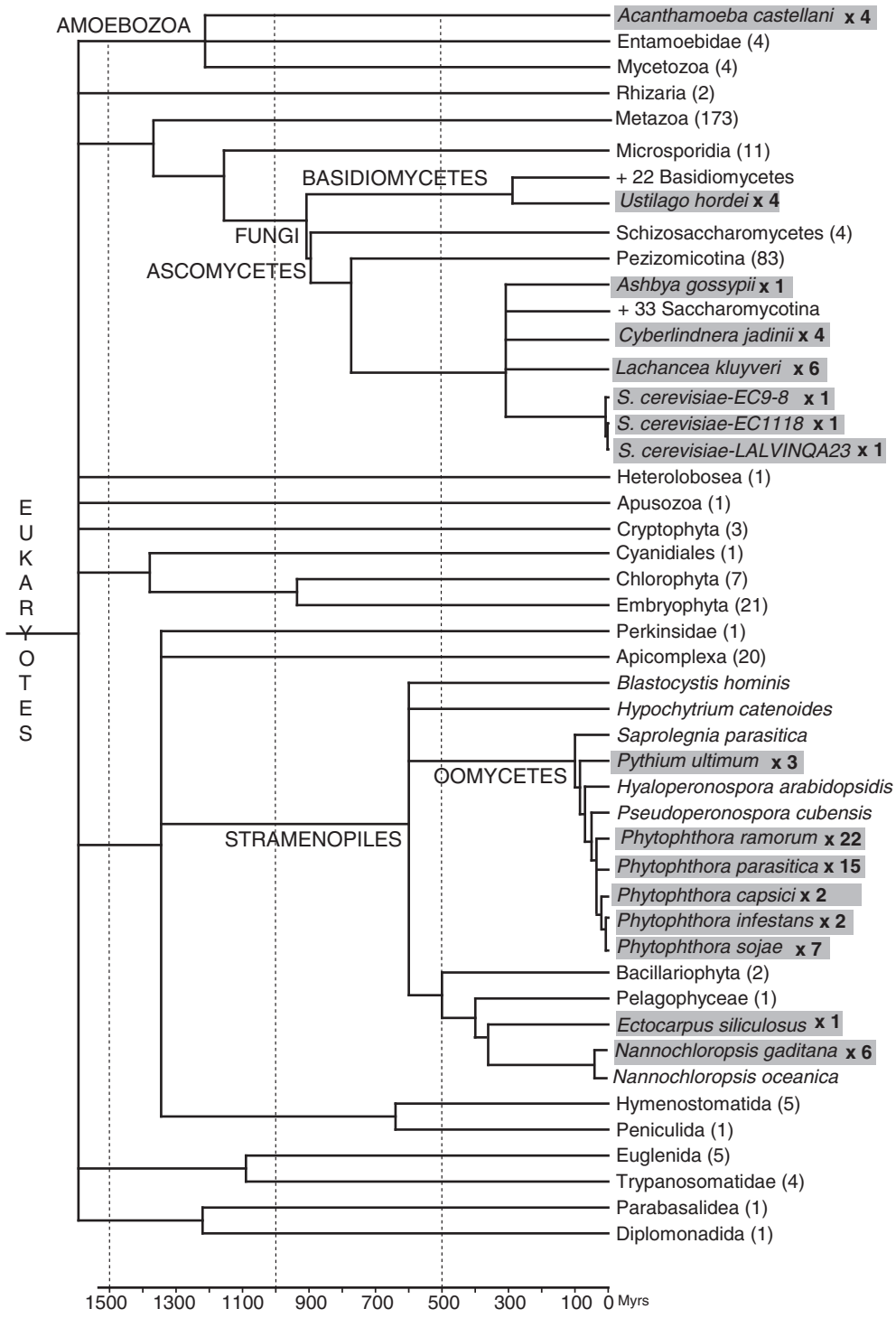

Fig. 1.—Timetree of all eukaryote species that were searched for the presence of bacterial ISs. The tree includes all 431 eukaryote species for which whole-genome sequence data were available in GenBank as of July 2012. Phylogenetic relationships and divergence times were taken from Blair et al. (2008), Brown and Sorhannus (2010), Lahr et al. (2011), Kurtzman (2003), and Hedges et al. (2006). Divergence times within oomycetes and between the two *Nannochloropsis* species are unknown and are represented arbitrarily for illustrative purposes. The name of the species in which IS607-like sequences were found and the number of sequences per species are shown in gray boxes. For taxa ranking above the species level, the number of available whole-genome sequences is given between brackets.

Robertson 2002; Feschotte 2004). Members of these superfamilies are generally widespread in multiple eukaryotic phyla, in which their long-lasting proliferation has produced diverse families of elements. Furthermore, aa similarities between representatives of these superfamilies and their cognate IS elements are usually low and limited to patches of residues surrounding the transposase catalytic domain (Eisen et al. 1994; Kapitonov and Jurka 1999; Robertson 2002; Feschotte 2004). Though it remains difficult to trace the origin of these superfamilies back to a precise date, together these lines of evidence suggest that their presence in eukaryotic genomes is likely to be very ancient. As proposed for eukaryotic non-LTR retrotransposon that are thought to be related to prokaryotic group II introns, their origin may even go back to the eukaryote ancestor (Eickbush and Malik 2002; Feschotte and Pritham 2007). In contrast, eukaryotic IS607-like sequences exhibit a very patchy phylogenetic distribution, not only at the scale of the eukaryotic tree but also within taxonomic groups of more recent origin. For example, IS607-like sequences were detected in only one (*Ustilago hordei*), six (three strains of *S. cerevisiae*, *Ash. gossypii*, *L. kluyveri,* and *Cyberlindnera jadinii*), and one (*Aca. castellani*) of the 23 basidiomycete, 177 ascomycete, and 9 Amoebozoa genomes available at the time of our search (fig. 1). In addition, the levels of aa similarity between IS607-like sequences and their most closely related prokaryotic counterparts are higher and/or extend over larger stretches of sequence than those typically observed between members of the Tc1/mariner, Mutator, Merlin, PIF-Harbinger, and ISL2EU/IS4EU superfamilies and their most closely related prokaryotic homologs (supplementary table 3, Supplementary Material online, and fig. 2). Together, these lines of evidence strongly suggest that the origin of IS607-like sequences in eukaryotic genomes is more recent than that of the five superfamilies of Class II eukaryotic TEs that are evolutionary linked to IS elements. Therefore, we conclude that the acquisition of IS607-like sequences in eukaryotic genomes most likely occurred after the origin of eukaryotes, which implies that these sequences result from one or more prokaryote-to-eukaryote HTs.

To further investigate the evolutionary processes underlying the distribution of IS607-like sequences in eukaryotes, we carried out phylogenetic analyses using the most conserved domains of ORFA (MerR HTH and resolvase domains, 163 aa) and ORFB (Zn ribbon domain, 110 aa). In both resulting trees (fig. 3, supplementary fig. 3, Supplementary Material online), the relationships between prokaryotic and eukaryotic sequences as well as most deep nodes are unresolved (bootstrap values <50%), which prevents drawing precise conclusions on the number of IS607 prokaryote-to-eukaryote HTs. Among eukaryotes, all sequences sharing similarity in their flanking region form well-supported groups (e.g., *U. hordei* sequences 1–4, *P. parasitica* 8–11, and *P. ramorum* 11–18), which is consistent with these sequences having experienced segmental duplications (fig. 3, supplementary table 1, Supplementary Material online). Interestingly, several strongly supported IS607-like groupings are inconsistent with the phylogeny of their host taxa (fig. 3). For example, within oomycetes, *Pythium ultimum* is distantly related to all species of the genus *Phytophthora* included in this study (Blair et al. 2008). However, its IS607-like sequences are sister to two of the 12 *P. parasitica* sequences included in the ORFB phylogeny, within a group that also includes a subset of *P. sojae* sequences (figs. 1 and 3). In addition, although the tree topology of yeast IS607 ORFB-like sequences is congruent with that of their host taxa (*Ash. gossypii* and the four strains of *S. cerevisiae*; figs. 1 and 3), the IS607-like sequence of *S. cerevisiae* EC1118 lies in a genomic region unique to this strain and closely related *S. cerevisiae* strains, and this region was horizontally transferred between *S. cerevisiae* and a member of a clade containing *Ash. gossypii* (Novo et al. 2009). Together, these data indicate that several events of eukaryote-to-eukaryote HT may have shaped the distribution of IS607-like sequences in eukaryotes.

Evolution of IS607-Like Sequences in Eukaryote Genomes

As 11 eukaryotic species possess two or more IS-like sequences, we investigated the potential mechanisms underlying their presence in multiple copies, for example, transposition or segmental duplication. IS-like sequences corresponding to fragments of ORFA or ORFB do not contain all the protein domains necessary for transposition. Instead, many of these sequences show homology in their flanking regions to one or several other IS-like sequences within the same genome (supplementary table 1, Supplementary Material online), indicating that they were generated by segmental duplication. Unlike many bacterial IS, IS607 elements do not possess terminal inverted repeats (TIRs) and do not necessarily generate target site duplications (TSDs) upon transposition (Kersulyte et al. 2000). Similarly, we could not find evidence of TIRs and TSDs in any of the eukaryotic IS607-like sequences. Thus, we cannot formally conclude on whether transposition of IS607-like sequences has occurred (and is still occurring) in any of the eukaryote genomes considered in this study. However, we note that several of the sequences corresponding to at least one complete ORF are free of nonsense mutations (stop or frameshift; fig. 2) and could potentially encode a functional transposase in various taxa (*P. parasitica*, *Pyt. ultimum*, *Ash. gossypii*, *S. cerevisiae*, *Aca. castellani,* and *L. kluyveri*). For example, one IS607-like sequence found in the amoeba *Aca. castellani* (Ac3) contains all protein domains (free of nonsense mutations) found in both ORFA and ORFB of IS607Arma1. It structurally differs from ISArma1 by a 100-aa insertion at the N-terminus of ORFB (fig. 2, supplementary fig. 1, Supplementary Material online). Another *Aca. castellani* IS607-like sequence (Ac4) is also devoid of nonsense mutations and contains all domains found in ORFA and ORFB of ISArma1, with the exception of a truncation in the resolvase C-terminus (supplementary fig. 1,

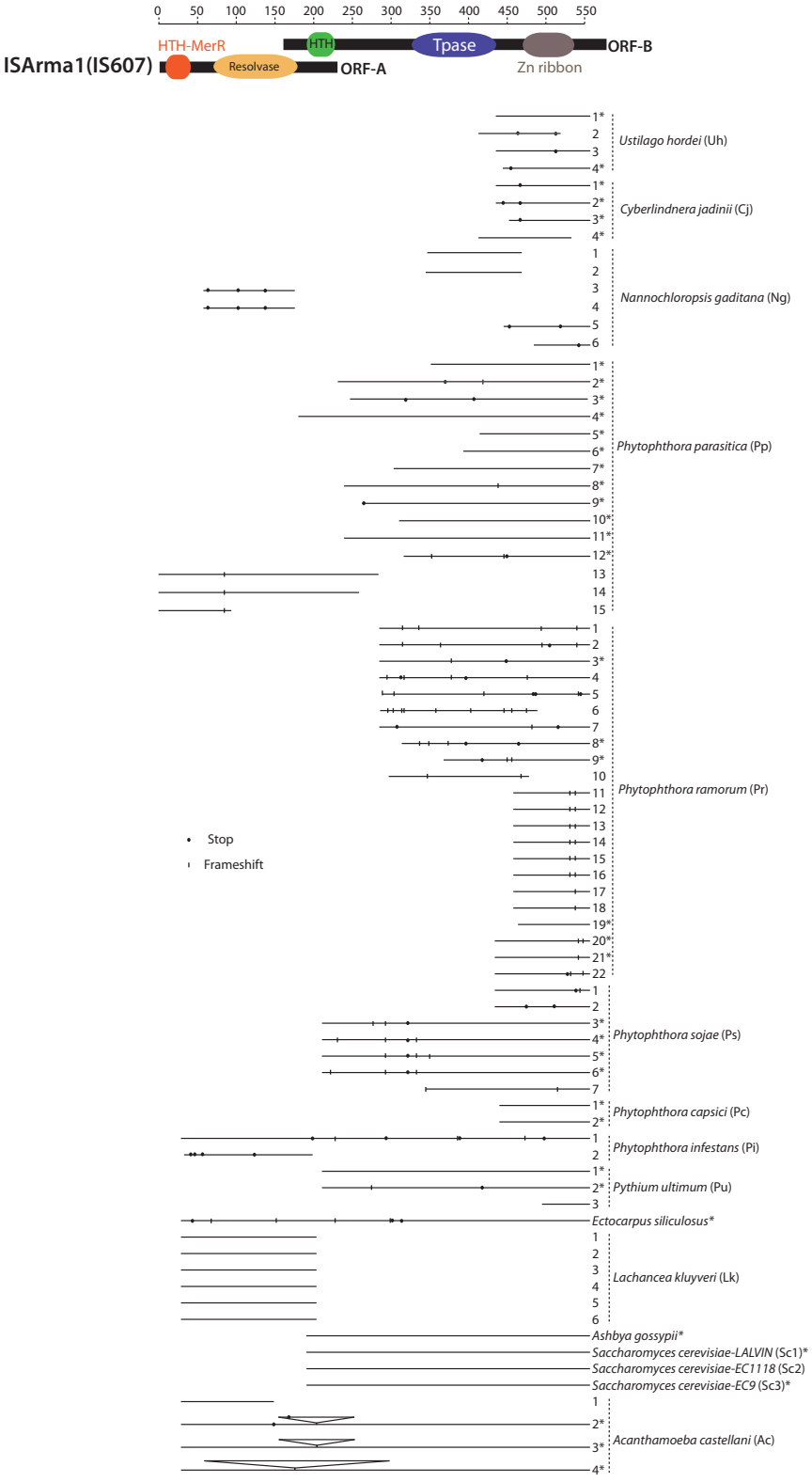

Fig. 2.—Eukaryotic IS607-like sequences mapped onto ISArma1. The domain structure of ISArma1 was determined in Pfam (Punta et al. 2012). Each horizontal line corresponds to one IS607-like sequence. Filled circles represent stop codons and vertical lines represent frameshifts. Sequences included in the selection analyses are marked with an asterisk. Inverted triangles indicate large insertions in the *Acanthamoeba castellani* sequences.

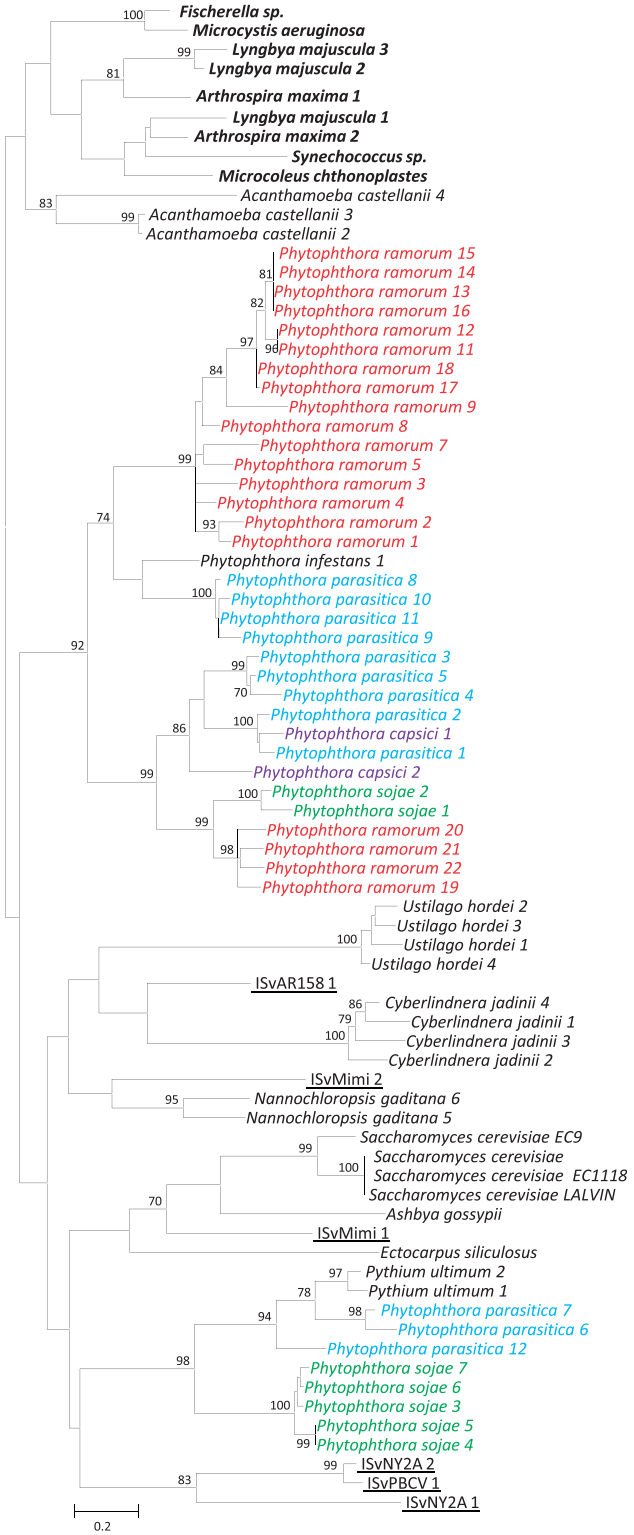

Fig. 3.—Phylogenetic tree of IS607-like ORFB sequences. Maximum-likelihood phylogenetic analyses were carried out using the WAG + G + F model of aa substitution. An interesting outcome of this phylogenetic analyses is that sequences found in *Phytophthora ramorum* (red), *P. capsici* (purple), *P. sojae* (green), and *P. parasitica* (blue) are polyphyletic. The relationships between IS607-like sequences are therefore incongruent with the host phylogeny. Bootstrap values above 70% are indicated. IS607-like sequences extracted from viral genomes are underlined and bacterial IS607 are in bold.

Supplementary Material online), which may inactivate ORFA. Although ORFB is not necessary for IS607 transposition in *E. coli*, it has been proposed that ORFA or ORFB could be needed for transposition in a different set of host species (Kersulyte et al. 2000). Interestingly, we found six *Aca. castellani* transcripts that map to the 3′-region of Ac4 (100% nucleotide identity, supplementary fig. 1, Supplementary Material online), suggesting that Ac4 is transcribed and that it may be transposing in natural *Aca. castellani* populations. It will be relevant to functionally characterize the catalytic activity of Ac3 and Ac4 in the future because if their ability to transpose can be confirmed, Ac3 and Ac4 would be, to our knowledge, the first prokaryotic TEs known to be capable of transposing in natural settings in a eukaryotic genome.

In addition to these potentially active IS607 elements, several other sequences contain at least one ORFA or ORFB domain free of nonsense mutations. The most frequent of these domains is the ORFB C-terminal Zn ribbon DNA-binding domain, which is intact in more than half of all IS607-like sequences uncovered in this study (43/80; fig. 2). To assess which forces governed its evolution, we carried out selection analyses of all 43 sequences containing an intact Zn ribbon DNA-binding domain, using the GAbranch method (Kosakovsky Pond and Frost 2005) and the tree presented in figure 3 as a reference. These analyses revealed that 65% (55/84) of the internal branches of this tree are characterized by d*N*/d*S* ratios lower than 0.4, the majority of which (40/55) having d*N*/d*S* ratios lower than 0.07. Furthermore, about half (20/41) of the external branches of the tree are characterized by d*N*/d*S* ratios lower than 0.4, many of which (13/20) having d*N*/d*S* ratios lower than 0.07 (supplementary fig. 3, Supplementary Material online). These results indicate that some IS607-like sequences have been evolving and are possibly still evolving under purifying selection in their respective eukaryotic genomes.

In some instances, this pattern of purifying selection could conceivably stem from functional constraints acting at the level of the element itself during HT, as demonstrated for the *Mariner* transposon in insects (Lampe et al. 2003). This would imply that at least some IS607 HTs occurred relatively recently during eukaryote evolution, so that there has not been enough time to erase the signal of purifying selection acting on the elements. Another line of evidence supporting a scenario of recent HTs for at least some IS607 elements is that all but one eukaryote IS607-like sequences are species specific, that is, we did not find them at orthologous loci in other species, even in the genus *Phytophthora* for which the

genomes of five species are available. Alternatively, the purifying selection signal may be explained by some IS607 sequences having been domesticated upon their arrival into eukaryote genomes, that is, they may have been exapted or recruited as novel host genes to fulfill new cellular functions (Gould and Vrba 1983; Miller et al. 1999). For example, one IS607-like locus is orthologous in three *S. cerevisiae* strains (corresponding to sequences Sc1–3, fig. 2), and some of the orthologous sequence pairs exhibit high nucleotide divergence (e.g., 30% between Sc3 and Sc1/Sc2). Together, these observations do not support a scenario of recent HT for this IS607-like sequence in *S. cerevisiae* strains. However, our selection analysis indicates that at least one sequence (Sc3) has been evolving under strong purifying selection ($dN/dS = 0.06$), thereby supporting the hypothesis that one or more domains of this sequence might have been domesticated. Another IS607-like sequence of interest in this regard is Pc1 in *P. capsici*. This sequence is evolving under strong purifying selection ($dN/dS = 0.06$), and its annotation as a predicted gene in the genome of *P. capsici* is supported by a cDNA sequence (accession number in the JGI genome browser: BT032152). Furthermore, five IS607-like sequences evolving under strong purifying selection ($dN/dS \leq 0.1$) are annotated as predicted genes in the genome of four species and could represent additional events of IS607 domestication (supplementary table 1, Supplementary Material online).

Discussion

We uncovered 80 sequences related to bacterial IS607 in the genomes of 13 eukaryotic species belonging to four different phyla (Amoebozoa, Ascomycetes, Basidiomycetes, and Stramenopiles). Though this is the highest number of IS-like sequences reported so far in eukaryotes, this number is relatively modest given that our search consisted in a screen of more than 400 eukaryote genomes using at least one representative of all known IS families as queries. Because we used conservative criteria in our search, it is possible that we have missed a number of eukaryotic IS-like sequences. In particular, to avoid retaining IS-like sequences resulting from contamination, we filtered out all sequences not flanked by at least one eukaryotic exon. Nevertheless, our study firmly establishes that TE HT has occurred between prokaryotes and eukaryotes, albeit quite rarely during recent eukaryote evolution.

Regarding the route(s) followed by IS607 to transfer between domains of life, it is noteworthy that several of the taxa in which we found IS607-like sequences have already been involved in various gene HT events. For example, a large number of genes with cyanobacterial ancestry have been uncovered in *P. sojae* and *P. ramorum* that are believed to result either from the secondary endosymbiosis through which the ancestor of Stramenopiles acquired its plastid (Tyler et al. 2006; Maruyama et al. 2009; Morris et al. 2009; Keeling 2010) or from later gene HTs specific to oomycetes (Archibald 2009). Several hundred gene HT events have also been reported between bacteria and fungi (Uo et al. 2001; Gojković et al. 2004; Hall et al. 2005; Wenzl et al. 2005; Hall and Dietrich 2007; Fitzpatrick et al. 2008; Marcet-Houben and Gabaldon 2010). In addition, it is known that some of the genes conferring to oomycetes the ability to parasitize plants were acquired via gene HT between fungi and oomycetes (Richards et al. 2011). Though the lack of resolution of the IS607-like sequence tree does not allow us to favor a precise scenario to explain the current taxonomic distribution of these sequences in eukaryotes, it is conceivable that at least some of them followed the same route as the horizontally transferred genes described earlier. Another notable aspect regarding the possible routes that IS607 may have followed to end up in several eukaryotic genomes is that free-living amoebae are known to act as "gene exchange platforms," facilitating gene HT between the numerous bacterial symbionts they host, the bacteria they feed on through phagocytosis, and their parasitic mimiviruses (Moliner et al. 2010; Bertelli and Greub 2012). Gene HT occurs both between members of the community of organisms that live within amoeba cells and between these organisms and their amoeba hosts (Filée and Chandler 2008; Moliner et al. 2009). In this context, it is remarkable that IS607 elements have been found in the genomes of several giant nucleocytoplasmic large DNA viruses, including that of the mimivirus, which is known to replicate in species of the *Acanthamoeba* genus (Filée et al. 2007). These IS elements apparently became integrated in viral genomes as parts of larger bacterial genome fragments, and much like the Ac3 and Ac4 IS607-like sequences, we uncovered in *Aca. castellani*, some copies are apparently intact, suggesting that their integration is relatively recent and that they may still be able to transpose (Filée et al. 2007). The other viruses known to harbor IS607 elements (*Chlorella* phycodnaviruses NY2A, AR158, and PBCV1) belong to the Phycodnaviridae (Filée et al. 2007), a family of viruses that infect marine and freshwater algae. It is noteworthy that a member of this family (the *Ectocarpus* phaeovirus Esv-1) infects the brown algae *Ect. siliculosus* and that a copy of the Esv-1 genome is integrated into the *Ect. siliculosus* genome (Delaroque and Boland 2008; Cock et al. 2010). The IS607-like sequence we found in *Ect. siliculosus* lies in contig0028 that flanks the contig containing the integrated Esv-1 genome (contig 0052), but we did not find any IS607 element in any of the integrated and nonintegrated Esv1 genomes. Together, these data suggest that viruses may have acted as vectors facilitating eukaryote-to-eukaryote IS607 HTs, as previously proposed for other TEs (Piskurek and Okada 2007; Dupuy et al. 2011; Routh et al. 2012).

The fact that many prokaryote-to-eukaryote gene HT events have been reported suggests that the paucity of IS-like sequences in eukaryotic genomes most likely does not result from a lack of opportunity for transfer. One may argue that prokaryotic genes may be more likely to be

successful at transferring into eukaryotes than TEs because they encode domains that can be readily used to fulfill a cellular function in the receiving eukaryote host. However, TEs are not only intrinsically equipped for excision and integration, two steps that are necessary for HT, but they can duplicate and reach high copy number, which increases the likelihood of fixation in populations. Therefore, we believe that the scarcity of prokaryote-to-eukaryote TE HT more likely stems from incompatibilities at the transcriptional level (Gladyshev and Arkhipova 2009) and/or at the level of transposase targeting between the different compartments of the eukaryotic cells, which prevent prokaryote TEs from being able to transpose and duplicate in eukaryote genomes. In this respect, the fact that all transferred IS sequences detected in this study are from the same family may reflect a pronounced flexibility of IS607 in terms of host factor requirements for transposition. The two transposases that it encodes may allow it to transpose in a large spectrum of hosts as proposed by Kersulyte et al. (2000), including some eukaryote species. Though we were not able to find any evidence of IS607 transposition in the various eukaryote species in which we found this element, the presence of seemingly full-length and intact copies in *Aca. castellani* provides an interesting eukaryote system in which to test for IS607 transpositional activity. Finally, our selection analyses suggest that in some cases, our ability to see these prokaryotic TEs in eukaryotic genomes today is because some of them may have been domesticated as new eukaryote genes, a possibility that will be interesting to assess at the functional level in future studies.

Supplementary Material

Supplementary figure S1–S4, tables S1–S3, and data sets S1–S5 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Myriam Badawi, Isabelle Giraud, and Jérôme Lesobre for technical assistance and Pierre Capy, Cédric Feschotte, Jonathan Filée, and Aurélie Hua-Van for discussions. We further acknowledge Bouziane Moumen for his assistance with the bioinformatic analyses. This work was supported by a European Research Council Starting Grant (FP7/2007-2013, grant 260729 EndoSexDet) to R.C.

Literature Cited

Andersson JO. 2005. Lateral gene transfer in eukaryotes. Cell Mol Life Sci. 62:1182–1197.

Archibald JM. 2009. The puzzle of plastid evolution. Curr Biol. 19:R81–8.

Astua-Monge G, Lyznik A, Jones V, Mackenzie SA, Vallejos CE. 2002. Evidence for a prokaryotic insertion-sequence contamination in eukaryotic sequences registered in different databases. Theor Appl Genet. 104:48–53.

Bapteste E, et al. 2009. Prokaryotic evolution and the tree of life are two different things. Biol Direct. 4:34.

Bertelli C, Greub G. 2012. Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms. Front Cell Infect Microbiol. 2:110.

Bichsel M, Barbour AD, Wagner A. 2010. The early phase of a bacterial insertion sequence infection. Theor Popul Biol. 78:278–88.

Blair JE, Coffey MD, Park SY, Geiser DM, Kang S. 2008. A multi-locus phylogeny for *Phytophthora* utilizing markers derived from complete genome sequences. Fungal Genet Biol. 45:266–77.

Brown JW, Sorhannus U. 2010. A molecular genetic timescale for the diversification of autotrophic stramenopiles (Ochrophyta): substantive underestimation of putative fossil ages. PLoS One 5:e12759.

Cerveau N, Leclercq S, Leroy E, Bouchon D, Cordaux R. 2011. Short- and long-term evolutionary dynamics of bacterial insertion sequences: insights from *Wolbachia* endosymbionts. Genome Biol Evol. 3:1175–86.

Cock JM, et al. 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature 465:617–21.

Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. Nat Rev Genet. 10:691–703.

Craig NL, Craigie R, Gellert M, Lambovitz AM. 2002. Mobile DNA II. Washington (DC): ASM Press.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27: 1164–1165.

Delaroque N, Boland W. 2008. The genome of the brown alga *Ectocarpus siliculosus* contains a series of viral DNA pieces, suggesting an ancient association with large dsDNA viruses. BMC Evol Biol. 8:110.

Delport W, Poon AF, Frost SD, Kosakovsky Pond SL. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26:2455–2457.

Doak TG, Doerder FP, Jahn CL, Herrick G. 1994. A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc Natl Acad Sci U S A. 91: 942–946.

Dunning Hotopp JC. 2011. Horizontal gene transfer between bacteria and animals. Trends Genet. 27:157–163.

Dupuy C, et al. 2011. Transfer of a chromosomal Maverick to endogenous bracovirus in a parasitoid wasp. Genetica 139:489–496.

Eickbush TH, Malik HS. 2002. Origins and evolution of retrotransposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington (DC): ASM Press. p. 1111–1144.

Eisen JA, Benito MI, Walbot V. 1994. Sequence similarity of putative transposases links the maize Mutator autonomous element and a group of bacterial insertion sequences. Nucleic Acids Res. 22: 2634–2636.

Feschotte C. 2004. Merlin, a new superfamily of DNA transposons identified in diverse animal genomes and related to bacterial IS1016 insertion sequences. Mol Biol Evol. 21:1769–1760.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet. 41:331–368.

Filée J, Chandler M. 2008. Convergent mechanisms of genome evolution of large and giant DNA viruses. Res Microbiol. 159:325–331.

Filée J, Siguier P, Chandler M. 2007. I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses. Trends Genet. 23: 10–15.

Fitzpatrick DA, Logue ME, Butler G. 2008. Evidence of recent interkingdom horizontal gene transfer between bacteria and *Candida parapsilosis*. BMC Evol Biol. 8:181.

Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol. 3: 722–732.

Gilbert C, Schaack S, Pace JK 2nd, Brindley PJ, Feschotte C. 2010. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. Nature 464:1347–1350.

Gladyshev EA, Arkhipova IR. 2009. A single-copy IS5-like transposon in the genome of a bdelloid rotifer. Mol Biol Evol. 26:1921–1929.

Gojković Z, et al. 2004. Horizontal gene transfer promoted evolution of the ability to propagate under anaerobic conditions in yeasts. Mol Genet Genomics. 271:387–393.
Gould SJ, Vrba ES. 1983. Exaptation—a missing term in the science of form. Paleobiology 8:4–15.
Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52: 696–704.
Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22: 2971–2972.
Hall C, Brachat S, Dietrich FS. 2005. Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. Eukaryot Cell. 4: 1102–1115.
Hall C, Dietrich FS. 2007. The reacquisition of biotin prototrophy in *Saccharomyces cerevisiae* involved horizontal gene transfer, gene duplication and gene clustering. Genetics 177:2293–2307.
Hall T. 2004. BioEdit version 5.0.6. Available from: http://www.mbio.ncsu.edu/BioEdit/bioedit.html, last accessed July 2012.
Kapitonov VV, Jurka J. 1999. Molecular paleontology of transposable elements from *Arabidopsis thaliana*. Genetica 107:27–37.
Kapitonov VV, Jurka J. 2007. IS4EU, a novel superfamily of eukaryotic DNA transposons. Repbase Rep. 7:143–143.
Keeling PJ. 2009. Functional and ecological impacts of horizontal gene transfer in eukaryotes. Curr Opin Genet Dev. 19:613–619.
Keeling PJ. 2010. The endosymbiotic origin, diversification and fate of plastids. Philos Trans R Soc Lond B Biol Sci. 365:729–748.
Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 9:605–618.
Kersulyte D, Mukhopadhyay AK, Shirai M, Nakazawa T, Berg DE. 2000. Functional organization and insertion specificity of IS607, a chimeric element of *Helicobacter pylori*. J Bacteriol. 182:5300–5308.
Kosakovsky Pond SL, Frost SD. 2005. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. Mol Biol Evol. 22:478–485.
Kosakovsky Pond SL, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679.
Kurtzman CP. 2003. Phylogenetic circumscription of *Saccharomyces*, *Kluyveromyces* and other members of the Saccharomycetaceae, and the proposal of the new genera *Lachancea*, *Nakaseomyces*, *Naumovia*, *Vanderwaltozyma* and *Zygotorulaspora*. FEMS Yeast Res. 4:233–245.
Lahr DJ, Grant J, Nguyen T, Lin JH, Katz LA. 2011. Comprehensive phylogenetic reconstruction of amoebozoa based on concatenated analyses of SSU-rDNA and actin genes. PLoS One 6: e22780.
Lampe DJ, Witherspoon DJ, Soto-Adames FN, Robertson HM. 2003. Recent horizontal transfer of mellifera subfamily mariner transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer. Mol Biol Evol. 20: 554–562.
Lander ES, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.
Lang AS, Zhaxybayeva O, Beatty JT. 2012. Gene transfer agents: phage-like elements of genetic exchange. Nat Rev Microbiol. 10: 472–482.
Marcet-Houben M, Gabaldón T. 2010. Acquisition of prokaryotic genes by fungal genomes. Trends Genet. 26:5–8.
Maruyama S, Matsuzaki M, Misawa K, Nozaki H. 2009. Cyanobacterial contribution to the genomes of the plastid-lacking protists. BMC Evol Biol. 9:197.
Miller WJ, McDonald JF, Nouaud D, Anxolabéhère D. 1999. Molecular domestication—more than a sporadic episode in evolution. Genetica 107:197–207.
Moliner C, Fournier PE, Raoult D. 2010. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. FEMS Microbiol Rev. 34:281–94.
Moliner C, Raoult D, Fournier PE. 2009. Evidence of horizontal gene transfer between amoeba and bacteria. Clin Microbiol Infect. 15: 178–180.
Moran Y, Fredman D, Szczesny P, Grynberg M, Technau U. 2012. Recurrent horizontal transfer of bacterial toxin genes to eukaryotes. Mol Biol Evol. 29:2223–2230.
Morris PF, et al. 2009. Multiple horizontal gene transfer events and domain fusions have created novel regulatory and metabolic networks in the oomycete genome. PLoS One 4:e6133.
Novo M, et al. 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. Proc Natl Acad Sci U S A. 106: 16333–16338.
Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304.
Oliver KR, Greene WK. 2009. Transposable elements: powerful facilitators of evolution. Bioessays 31:703–714.
Piskurek O, Okada N. 2007. Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. Proc Natl Acad Sci U S A. 104:12046–12051.
Puigbò P, Wolf YI, Koonin EV. 2010. The tree and net components of prokaryote evolution. Genome Biol Evol. 2:745–756.
Punta M, et al. 2012. The Pfam protein families database. Nucleic Acids Res. 40:D290–D301.
Richards TA, et al. 2011. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. Proc Natl Acad Sci U S A. 108:15258–15263.
Robertson HM. 2002. Evolution of DNA transposons in eukaryotes. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington (DC): ASM Press. p. 1093–1110.
Rolland T, Neuvéglise C, Sacerdot C, Dujon B. 2009. Insertion of horizontally transferred genes within conserved syntenic regions of yeast genomes. PLoS One 4:e6515.
Routh A, Domitrovic T, Johnson JE. 2012. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. Proc Natl Acad Sci U S A. 109:1907–1912.
Sánchez-Gracia A, Maside X, Charlesworth B. 2005. High rate of horizontal transfer of transposable elements in *Drosophila*. Trends Genet. 21: 200–203.
Schaack S, Gilbert C, Feschotte C. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol. 25:537–546.
Schnable PS, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115.
Senejani AG, Sweasy JB. 2010. Eukaryotic gene invasion by a bacterial mobile insertion sequence element IS2 during cloning into a plasmid vector. Genome Integr. 1:2.
Siguier P, Filée J, Chandler M. 2006. Insertion sequences in prokaryotic genomes. Curr Opin Microbiol. 9:526–531.
Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EP, de la Cruz F. 2010. Mobility of plasmids. Microbiol Mol Biol Rev. 74:434–452.
Syvanen M. 2012. Evolutionary implications of horizontal gene transfer. Annu Rev Genet. 46:341–358.
Thomas J, Schaack S, Pritham EJ. 2010. Pervasive horizontal transfer of rolling-circle transposons among animals. Genome Biol Evol. 2: 656–664.
Touchon M, Rocha EP. 2007. Causes of insertion sequences abundance in prokaryotic genomes. Mol Biol Evol. 24:969–981.
Tyler BM, et al. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313: 1261–1266.

Uo T, Yoshimura T, Tanaka N, Takegawa K, Esaki N. 2001. Functional characterization of alanine racemase from *Schizosaccharomyces pombe*: a eukaryotic counterpart to bacterial alanine racemase. J Bacteriol. 183:2226–2233.

Wagner A, de la Chaux N. 2008. Distant horizontal gene transfer is rare for multiple families of prokaryotic insertion sequences. Mol Genet Genomics. 280:397–408.

Wallau GL, Ortiz MF, Loreto EL. 2012. Horizontal transposon transfer in eukarya: detection, bias, and perspectives. Genome Biol Evol. 4:689–699.

Wenzl P, Wong L, Kwang-won K, Jefferson RA. 2005. A functional screen identifies lateral transfer of beta-glucuronidase (gus) from bacteria to fungi. Mol Biol Evol. 22:308–316.

Wozniak RA, Waldor MK. 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol. 8:552–563.

Associate editor: Tal Dagan