Origin and Spread of Spliceosomal Introns: Insights from the Fungal Clade *Zymoseptoria*

Baojun Wu[1], Allison I. Macielog, and Weilong Hao*

Department of Biological Sciences, Wayne State University

[1]Present address: Department of Biology, Clark University, Worcester, MA, USA

*Corresponding author: E-mail: haow@wayne.edu.

Accepted: September 29, 2017

Abstract

Spliceosomal introns are a key feature of eukaryote genome architecture and have been proposed to originate from selfish group II introns from an endosymbiotic bacterium, that is, the ancestor of mitochondria. However, the mechanisms underlying the wide spread of spliceosomal introns across eukaryotic genomes have been obscure. In this study, we characterize the dynamic evolution of spliceosomal introns in the fungal genus *Zymoseptoria* at different evolutionary scales, that is, within a genome, among conspecific strains within species, and between different species. Within the genome, spliceosomal introns can proliferate in unrelated genes and intergenic regions. Among conspecific strains, spliceosomal introns undergo rapid turnover (gains and losses) and frequent sequence exchange between geographically distinct strains. Furthermore, spliceosomal introns could undergo introgression between distinct species, which can further promote intron invasion and proliferation. The dynamic invasion and proliferation processes of spliceosomal introns resemble the life cycles of mobile selfish (group I/II) introns, and these intron movements, at least in part, account for the dramatic processes of intron gain and intron loss during eukaryotic evolution.

Key words: spliceosomal introns, transposition, introgression, gain and loss.

Introduction

Since the discovery of spliceosomal introns in eukaryotes (Berget et al. 1977; Gilbert 1978), much progress has been made toward understanding the origin, evolution and function of spliceosomal introns (Cavalier-Smith 1991; Palmer and Logsdon 1991; Koonin 2006; Heyn et al. 2015). Spliceosomal introns are thought to originate from self-splicing group II introns, which are present in eubacteria, archaea, and eukaryotic organelles (Koonin 2006). Such hypothesis is supported by the fact that the spliceosome complex and group II introns share similar structures (Zimmerly and Semper 2015). After the birth of spliceosomal introns, they have undergone extensive gain and loss in a variety of lineages. As a result, intron densities in modern eukaryotic species vary greatly from a few introns in the entire genome of *Giardia* genus (Morrison et al. 2007) to over six introns per gene in most vertebrates and land plants (Csuros et al. 2011), many of which are either substantially higher or significantly lower than the intron density of the Last Eukaryotic Common Ancestor (estimated at 4.3 introns per gene) (Csuros et al. 2011). Several mechanisms have been proposed for the dynamic intron gain and loss processes (Farlow et al. 2011; Yenerall and Zhou 2012). For instance, variable intron densities among eukaryotic species could result from the interplay between nonhomologous end joining (NHEJ) and homologous recombination (HR), supposedly favoring DNA insertion and intron gain, and intron loss via reverse splicing, respectively.

Much higher (50- to 3,000-fold) rates of intron loss than intron gain have been observed in eukaryote genomes, including many intron-rich lineages (Roy and Gilbert 2005; Csuros et al. 2011). The presence of abundant intron-rich organisms requires mechanisms that can count against intron loss and/or lead to significant intron gain. The fast-growing genomic data have provided direct insights into some of the mechanisms. Particular spliceosomal introns share high similarity with other introns from unrelated genes, and are referred to as introner elements (IEs) (Worden et al. 2009). IEs are commonly present in a wide range of eukaryotes, including fungi, algae, tunicate and nematodes (Roy 2004; Worden et al. 2009; Denoeud et al. 2010; van der Burgt et al. 2012;

Verhelst et al. 2013). Many IEs are flanked by 3–8 nt short direct repeats or target site duplications (TSDs), which are hallmarks of NHEJ mediated sequence transposition (Li et al. 2009; Huff et al. 2016). In a recent study, transposition of IEs has been suggested to account for hundreds to thousands of spliceosomal introns in the green alga *Micromonas pusilla* and the heterokont alga *Aureococcus anophagefferens* (Huff et al. 2016). Together, these IEs of transposon-like features are potentially important for the transposition and proliferation of modern splicesomal introns.

To better understand the origin and spread of spliceosomal introns, we turned to the fungal genus *Zymoseptoria*, in which transposition of spliceosomal introns between unrelated-genes and IPAPs (intron presence–absence polymorphisms) at homologous loci among conspecific strains have been characterized (Torriani et al. 2011; van der Burgt et al. 2012; Collemare et al. 2013; Brunner et al. 2014). Taking advantage of the population genomics data of this genus, we evaluated intron dynamics at different evolutionary scales, namely, within the genome, and among conspecific and interspecific genomes. Our results suggest that spliceosomal introns can proliferate via intergenic homolog elements within the genome, invade other genomes within the same species and also between distinct species.

Materials and Methods

Introns Verification and Improved Annotation of Intergenic Regions in the *Z. tritici* IPO323 Genome

The annotated genome (containing 21 chromosomes) of *Zymoseptoria tritici* IPO323 (accession: ACPE00000000) (Goodwin et al. 2011) was downloaded from the GenBank database (supplementary table S1, Supplementary Material online). Annotated intron sequences were extracted based on their GenBank annotation. The raw transcriptome data (SRA accessions: ERR789217 and ERR789232) (Rudd et al. 2015) were downloaded from the NCBI SRA database and mapped to the IPO323 genome using TopHat (Kim et al. 2013). Splice junctions of each annotated intron were verified based on the mapping results. The expression level of each protein gene was measured by Cufflinks v2.2.1 (Trapnell et al. 2012). Putative RNAs within annotated introns were predicted by tRNAscan (Lowe and Eddy 1997) and Rfam 12.0 (Nawrocki et al. 2015). We considered an annotated intron as a verified intron only if 1) the whole protein coding region is transcribed with FPKM >1; 2) there are at least three reads supporting the intron being spliced out; 3) and there are no detected tRNA/small-ncRNA sequences within the intron region. To obtain a conservative view of intergenic regions, *ab initio* gene prediction was performed on the annotated intergenic regions using the Augustus program (Stanke and Morgenstern 2005) and regions free of any predicted genes were considered further as verified intergenic regions.

Analysis of Population Genomic Data from the *Zymoseptoria* Genus

The draft genome sequences of two *Z. tritici* (synonym *Mycosphaerella graminicola*) strains isolated from Iran were download from GenBank [accessions: AEYQ00000000 and AFIP00000000 (Stukenbrock et al. 2011)]. Raw sequencing data of 22 *Z. tritici* strains were downloaded from the NCBI SRA database. Nine were isolated from Switzerland, 13 were isolated from Australia, and their BioProject numbers are PRJNA178194 (Croll et al. 2013) and PRJNA299857 (McDonald et al. 2016), respectively. The draft genome sequences of two other *Zymoseptoria* species, *Z. pseudotritici* and *Z. ardabiliae*, were also analyzed. There are five *Z. pseudotritici* strains and four *Z. ardabiliae* strains first described in Stukenbrock et al. (2011). *De novo* assembly was performed using SPAdes with four different *K*-mers (21, 33, 55, and 77) (Bankevich et al. 2012). The assembled genomes are available upon request. The statistics of genome assembly is shown in supplementary table S2, Supplementary Material online.

Identification of IIHEs

The verified introns were used as queries to search for potential homologs in the redefined intergenic regions. Only homologous sequences with $\geq$ 80% sequence identity and $\geq$ 80% match length were collected. To further minimize the possibility of identified intergenic regions being protein genes, the ±200 nt flanking sequences of each identified IIHE were extracted for BLASTX searches (*E*-value $< 1 \times 10^{-5}$) against all predicted proteomes of currently available (as of July 2017) species ($n = 43$) from the *Dothideomycetes* class in GenBank (supplementary table S3, Supplementary Material online), which *Zymoseptoria* belongs to.

Sequence Analysis of IPAP among 25 *Z. tritici* Strains

Fifty-two *Z. tritici* introns annotated in the *Z. tritici* genome have been previously identified as IPAP (Torriani et al. 2011). Forty-two introns were supported by available RNA-seq data and EST data (as described earlier). We conducted BLASTN searches ($\geq$80% sequence identity and $>$ 80% match length) using these 42 introns and their flanking exon sequences as query sequences against the assembled *Zymoseptoria* genomes. An intron was deemed as presence if the intron and flanking exon sequences all have significant matches, and the matched intron sequences is flanked by the matched exon sequences; an intron was deemed as absence if its flanking exon sequences both have significant matches, and the matched exon sequences are adjacent in the examined genome. Orthologous introns were determined by best reciprocal BLASTN searches at intron regions and also flanking exonic sequences. Finally, 20 IPAP introns have high-quality sequence alignment for both the introns and their flanking regions, and are subject to further analyses. The nucleotide diversity (π) of exonic sequences flanking these IPAP introns was calculated by DnaSP (Librado and Rozas 2009) in sliding windows of 50 nt and step sizes of 10 nt.

Quantification of Intron Turnover and Intron Exchange

Intron turnover and intron exchange at homologous sites were modeled as continuous-time Markov processes. Intron presence and absence were modeled as a two-state process with state 0 for absence and state 1 for presence. For intron exchange at loci of fixed introns, different intron types were treated as different character states. The rates of intron turnover and intron exchange were measured by mapping character states on a phylogeny using the tree branch length as a relative time scale using the R package DiscML (Kim and Hao 2014). As detailed in our previous studies, the rate is relative to nucleotide substitution rate, and the unit is expressed as the number of gains/losses per site per nucleotide substitution (Hao and Golding 2006, 2010; Wu and Hao 2014; Wu et al. 2015).

Identification of Introgression between Different *Zymoseptoria* Species

Pairwise genetic distance was calculated among the *Z. tritici* strains, between the *Z. tritici* and *Z. pseudotritici* strains, and between the *Z. tritici* and *Z. ardabiliae* strains using the Jukes-Cantor model (Jukes and Cantor 1969). If an interspecific (between species) genetic distance is smaller than any conspecific (within-species) distance, this suggests the occurrence of introgression between the two species. To determine whether the introgression phenomenon is more predominant in introns, we compared the results between intron sequences and their protein coding genes.

Phylogenetic Analysis of Species and Strains Relationship

The 1,829 single-copy genes that are universally present in all examined 25 *Z. tritici* strains, five *Z. pseudotritici* strains and four *Z. ardabiliae* strains were used to construct phylogenetic relationships. Intron sequences were removed from the phylogenetic analyses. Each gene was aligned individually using MUSCLE (Edgar 2004). The concatenated sequences of all gene alignments were used to reconstruct the phylogenetic relationship of these strains. Phylogenetic trees were constructed using the RAxML program (Stamatakis 2006) under a GTR $+ \Gamma +$ I substitution model, and 100 bootstrap iterations were performed.

Results

Intergenic Regions as Reservoirs of Modern Spliceosomal Introns

Modern spliceosomal introns have been suggested to have transposon-like features since the discovery of IEs, which are introns sharing high sequence similarity with other introns within the genome (Worden et al. 2009). A recent study has indicated that IEs could transpose between different genes by mechanisms similar to transposons (Huff et al. 2016). If spliceosomal introns are similar to mobile elements, one would expect introns and related elements to transpose to regions beyond gene regions. To test this, we searched for intron homologs in intergenic regions within the only annotated *Zymoseptoria tritici* genome, IPO323 (Goodwin et al. 2011). In the *Z. tritici* IPO323 genome, there are 17, 357 annotated introns. Among these introns, 10, 192 introns met our verification criteria (see Materials and Methods) and were considered as verified introns. We then performed *ab initio* gene predictions on all annotated intergenic sequences, identified 738 putative genes and removed these regions from the intergenic sequences. Through a stringent BLAST search ($\geq$ 80% sequence identity and $\geq$ 80% match length) using the 10,192 verified introns against the redefined intergenic sequences, we identified 83 introns from 74 genes having homologs in intergenic regions. To further minimize unrecognized genes in intergenic sequences, the flanking sequences ($\pm$200 nt) of intron homologs in intergenic regions were used to search against the proteomic sequences of 43 species in the Dothideomycetes class. After the exclusion of the intron homologs whose flanking sequences in intergenic regions match proteomic sequences, 50 introns from 44 genes were considered to have homologs in the re-defined intergenic regions. Hereafter, we call these intron homologs in intergenic regions as intergenic intron homolog elements (IIHEs).

The levels of sequence similarity between introns and their intergenic homologs are often very high, but decrease dramatically in the flanking regions (fig. 1*A*). We found that insertion sites of introns and their intergenic homologs could be flanked by duplicated sequences (TSD in supplementary fig S1, Supplementary Material online). TSD is a characteristic mark of transposon insertion or remnant of NHEJ, and both processes have been shown to be involved in intron mobility (Li et al. 2009; Huff et al. 2016). Introns having IIHE(s) are of different sequence propensities compared with introns not having IIHE(s). Introns having IIHE(s) are longer (*P*-value $<$ 0.0001, Wilcoxon Test) and skewed towards the middle and 3′ end of the gene, while the ones not having IIHE(s) are shorter and skewed towards the 5′ end of the gene (fig. 1*B* and *C*). This is consistent with the notion that intron loss preferentially takes place toward the 3′-end of the gene (Mourier and Jeffares 2003), resulting in empty intron sites known as protosplice sites where new introns tend to reinvade (Sverdlov et al. 2004).

Among the introns having IIHEs, 40% of them share high sequence similarity with at least one other intron ($\geq$80% sequence identity and $\geq$80% match length). These introns have homologs both in intron and intergenic regions, suggesting a dynamic transposition of intron homologs within a genome.

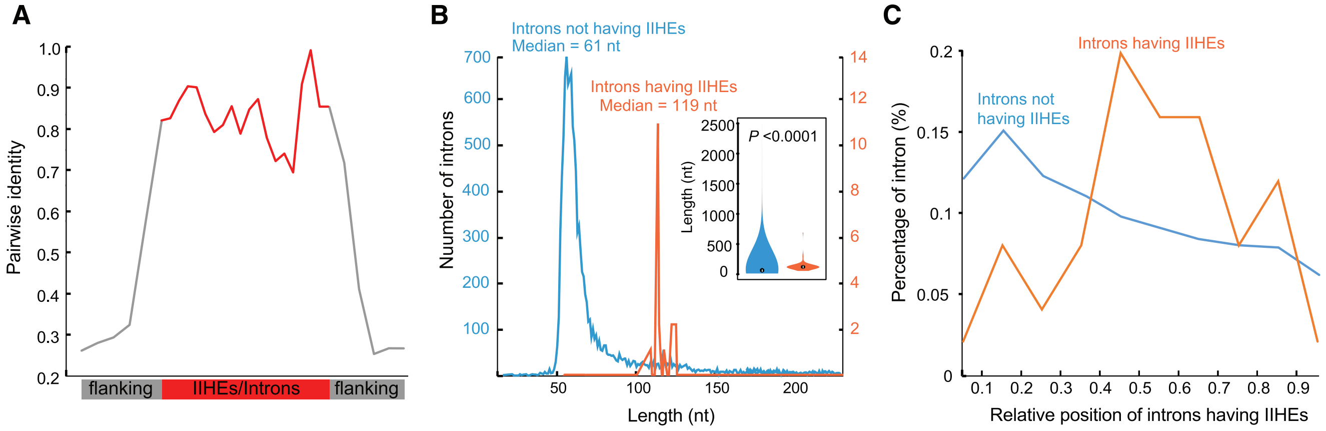

Fig. 1.—Characteristics of IIHEs in the *Z. tritici* IPO323 genome. (*A*) High sequence similarities are evident between IIHEs and their corresponding introns (in red), but not between flanking sequences (in gray). (*B*) Introns having IIHEs tend to have longer sequence-lengths than introns not having IIHEs (P-value $<$ 0.0001, Wilcoxon Test). (*C*) Comparison of intron position along protein genes between introns having IIHEs and introns not having IIHEs. The relative position of each intron along the protein gene is calculated, and introns are grouped into 10 bins (ranging from 0 to 1) by their relative positions in the coding regions.

We further investigated the association between IIHEs and intron dynamics at the population level using IPAP as an indicator for recent intron movement. A total of 52 IPAP loci have been previously reported in Torriani et al. (2011), 42 intron loci are verified by available transcriptome data (Rudd et al. 2015) and EST data (Torriani et al. 2011). Among the 42 verified introns, nine of them (21%) have IIHEs. In comparison, among the 10, 150 fixed introns, only 41 introns (0.4%) have IIHEs. This significant contrast (P-value $<$ 0.0001, Fisher's exact test) suggests that IIHEs are more strongly associated with IPAP than with fixed intron, and likely contribute to recent intron turnover in *Z. tritici*.

Rapid Intron Turnover and Exchange among Conspecific Strains

We first took advantage of the introns of IPAP to estimate intron turnover rates among 25 conspecific strains. Introns from 20 IPAP loci (supplementary fig. S2, Supplementary Material online) were mapped on the *Z. tritici* phylogeny (supplementary fig. S3, Supplementary Material online) and the rates of intron gain and intron loss were measured. Our results show a higher overall rate of intron gain (152.7 ± 16.5) than intron loss (83.9 ± 9.1) (fig. 2*A*) and the likelihood ratio test favors different rates between intron gain and intron loss ($2\Delta \ln L = 4.2$, P-value < 0.05, df $= 1$). As the estimated turnover rates are relative to the nucleotide substitution rate (Hao and Golding 2006; Wu and Hao 2014; Wu et al. 2015), both the rates of intron gain and intron loss are considered to be fast. A difference between rates of intron gain and intron loss is also evident in population genomics and genetic crosses data (Torriani et al. 2011; Brunner et al. 2014). For example, among 43 F1 progenies after cross between ID39491i2-containing and ID39491i2-lacking strains, 30 progenies have intron-present alleles (IPAs) and 13 have intron-absent alleles (IAAs), showing a significant bias towards intron invasion (P-value $= 0.009$, χ^2 test).

To gain insights into the molecular mechanisms underlying the spread of introns, we carefully analyzed the ±500 nt (nucleotides) exonic sequences immediately flanking intron insertion sites from 20 IPAP loci (supplementary fig. S2, Supplementary Material online). Overall, nucleotide diversity (π) of exonic sequences is low near intron insertion sites (flanking ±100 nt) among IPAs, and also among IAAs (fig. 2*B*). Strikingly, however, the same regions are among regions with the highest nucleotide diversity between IPAs and IAAs (fig. 2*B*), suggesting a tight association between the absence/presence of introns and the evolutionary history of exonic sequences immediately flanking intron insertion sites. We then sought to map the patterns of intron presence/absence and ±100 nt exonic sequences (only nucleotide sites that have variation among strains are shown) immediately flanking intron-insertion sites onto the organismal tree (supplementary fig. S3, Supplementary Material online). Exonic sequences near intron insertion sites show a strong association with the presence or absence of the corresponding intron, but little association with the organismal phylogenetic relationship or geographic distribution (fig. 2*C* and supplementary fig. S2, Supplementary Material online). The narrow flanking regions (±100 nt) associated with intron presence/absence are unlikely due to chromosomal recombination events during meiosis, whose interval distances in *Z. tritici* have been measured to be 3 kb or longer (Croll et al. 2015). Such narrow regions flanking intron presence/absence resemble the co-conversion tracts flanking selfish group I/II introns generated by gene conversion (Lazowska et al. 1994; Sanchez-Puerta et al. 2008; Wu and Hao 2014).

We extended our investigation on whether intron exchange takes place at loci of fixed introns (introns present in

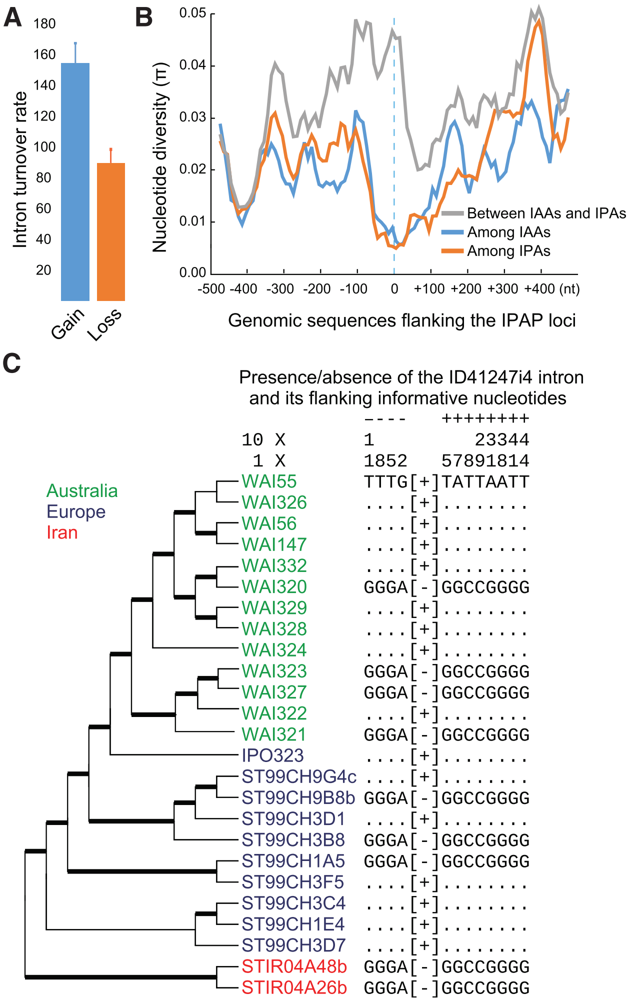

Fig. 2.—Characteristics of intron loci of IPAP in *Z. tritici* strains. (*A*) Overall turnover rates (± SE) of introns estimated at IPAP loci. (*B*) Comparison of the nucleotide diversity (π) values of the ±500 nt exon sequences flanking the insertion sites of PAP introns, among IAAs, among IPAs, and between IAA and IPA alleles. (*C*) Exonic sequence alignments flanking the insertion site of the ID41247i4 intron (i.e., the fourth intron in the ID41247 gene) are shown for illustration purpose. Only informative sites (nucleotide sites that show difference among strains) flanking ±100 nt of the intron site are shown with the nucleotide co-ordinates at the top. Dots indicate identities relative to the top sequence, whereas letters represent nucleotide differences. Intron presence is indicated by [+], while intron absence is shown as [−]. The colors of strain names represent their geographical origins, and internal branches supported by > 80 bootstrap values are shown in thick lines.

all Z. tritici strains), which would not result in IPAP. We employed BLASTClust (100% sequence coverage and 100% nucleotide identity) to group intron sequences from each homologous locus. Introns with no sequence variation and the ones forming more than four BLASTClust groups were excluded. Finally, we found introns from 463 loci (5% of total verified introns) showing evidence of sequences exchange between strains according to their phylogenetic positions on the organismal tree (supplementary fig. S3, Supplementary Material online). Among the *Z. tritici* strains isolated from Australia, nine of them show low levels of overall intron sequence diversity (the top nine strains in fig. 3*A*) due to a recent and perhaps severe population bottleneck (Banke et al. 2004). Nevertheless, among the 463 intron loci, 186 loci (40%) clearly support that at least one of the nine postbottleneck strains is different from other postbottleneck strains, but identical with some "organismally" more distantly related strains. Thus, there is a significant amount of intron exchange between some of the postbottleneck strains and prebottleneck strains. For example, the ID64923i1 intron in the WAI147 from Australia is identical with the one in the STIR04A48b strain from Iran but very different from the other eight postbottleneck strains (fig. 3*C* and supplementary fig. S4, Supplementary Material online). Furthermore, we modeled different intron types as different states and estimated the rates of intron exchange at each of the 463 loci of fixed introns. Under an equal rate model, the rates of intron exchange (± SE) were estimated from 14.8 ± 10.7 to 555.6 ± 249.1 with a median of 98 (fig. 3*B*). This suggests that fixed introns can undergo intron exchange at rates comparable to intron turnover at IPAP loci.

Intron Introgression among Different *Zymoseptoria* Species

We expanded comparative analysis of intron homologs among the *Z. tritici* strains to their closely related species *Zymoseptoria pseudotritici* (five strains) and *Z. ardabiliae* (four strains). Twenty-two intron loci were found to be of IPAP among the three *Zymoseptoria* species (Torriani et al. 2011). As illustrated in three intron loci, the exonic sequences immediately flanking intron insertion site are tightly associated with the presence or absence of the intron in different strains, but independent of their organismal relationship (supplementary fig. S5, Supplementary Material online). For example, the exonic sequences immediately flanking the ID77869i3 intron in *Z. tritici* break into two utterly different groups, with the 13 intron-containing *Z. tritici* strains clustered together with seven intron-containing strains in *Z. pseudotritici* and *Z. ardabiliae*. In contrast, the entire coding sequences of the ID77869 gene support a clear separation of the three well-defined species (fig. 4). The different phylogenetic relationships between the intron and coding sequences suggest that the introgression of the ID77869i3 intron is restricted to a very short region, not likely mediated by exchange in large sequences fragments. The strikingly similar or identical sequences immediately flanking the intron insertion site but not the entire protein gene cannot be simply explained by random point mutation, but favor the notion that introgression takes place at regions restricted to introns and their short flanking regions. This type of fine-scale sequence exchange between different species, as sources for diversity, can also explain the high sequence diversity flanking intron insertion

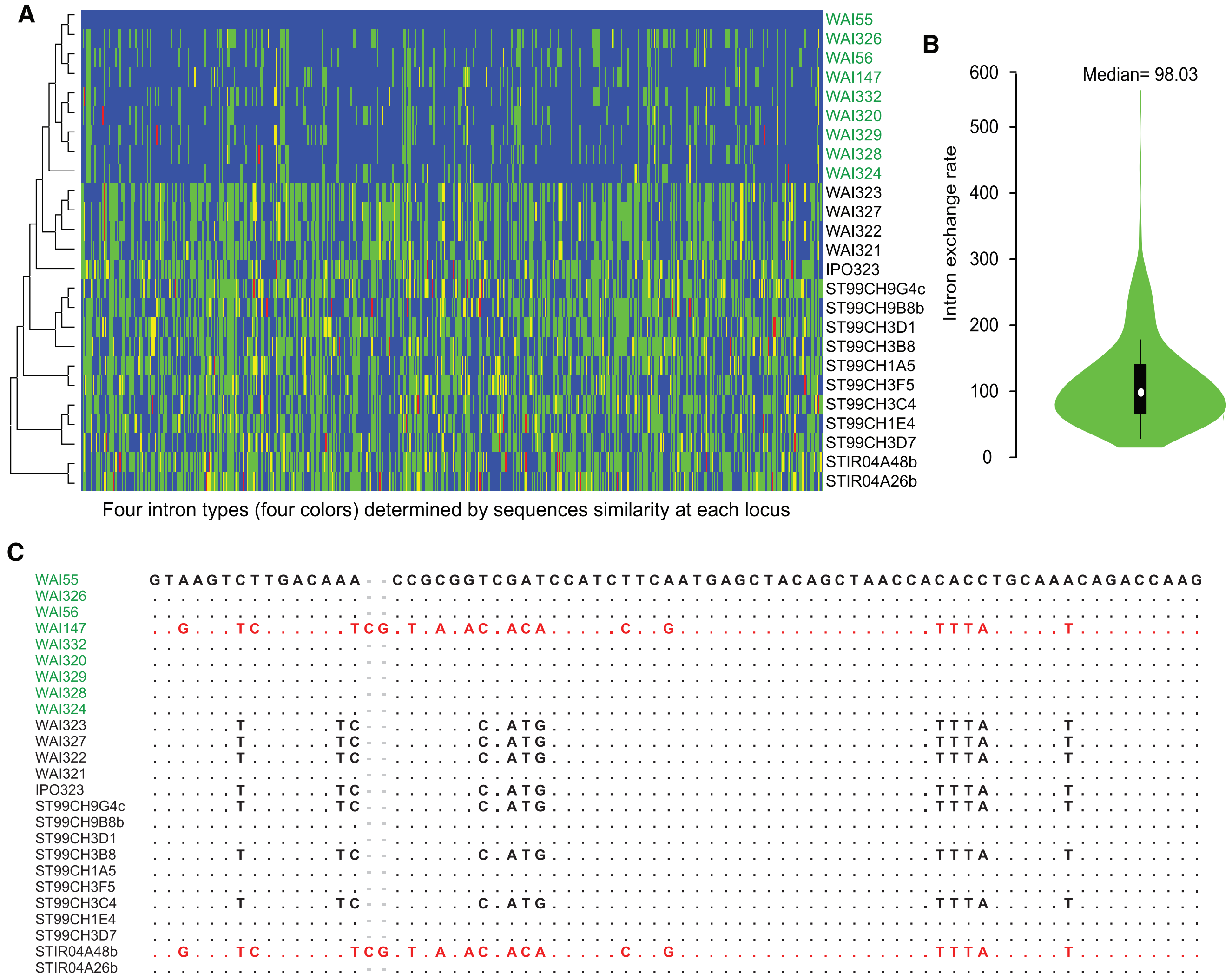

Fig. 3.—Intronic sequence exchange at 463 loci of fixed introns among 25 *Z. tritici* strains. (*A*) Heatmap of intron types categorized based on sequence similarity. For each intron locus (each column), identical intron sequences are shown in the same color. By looking vertically in panel A, identical introns (in the same color) can be shared between distantly related strains, while closely related strains may have different introns (in different colors). (*B*) Rates of intron exchange calculated for each of the 463 intron loci by assuming an equal-rate model. (*C*) The ID64923i1 intron sequences are shown as an example to illustrate intron exchange. The WAI147 and STIR04A48b strains are distantly related, but have identical intron sequences. Extended alignment of the exonic sequences flanking ID64923i1 can be found in supplementary figure S4, Supplementary Material online.

sites between IAAs and IPAs among conspecific strains (fig. 2C and supplementary fig. S2, Supplementary Material online).

Intron introgression also takes place extensively at loci of fixed introns. Among 5,506 orthologous introns commonly shared by all 25 examined *Z. tritici* and 4 *Z. pseudotritici* strains, 824 introns support a higher interspecific sequence similarity than between conspecific ones. The 824 introns involved in introgression are from 712 protein genes, but 544 (76%) of these protein genes support a monophyletic relationship of all *Z. tritici* strains based on phylogenetic analysis of entire protein gene. For instance, the five orthologs of the ID97336i1 intron in *Z. pseudotritici* share 100% identity with two *Z. tritici* strains (fig. 5A), but the sequences of the entire ID97336 gene show a clear separation between *Z. pseudotritici* and *Z. tritici* (fig. 5B). Similarly, 484 introns were involved in introgression between *Z. tritici* and *Z. ardabiliae*. These introgressed introns are from 438 protein genes but 371 (85%) of these genes support a monophyletic relationship of all *Z. tritici* strains (one case shown in supplementary fig. S6, Supplementary Material online). These results suggest that intron introgression is highly specific to the intron regions. Such targeted and extensive intron introgression is important in the wide spread of introns among a broad spectrum of organisms.

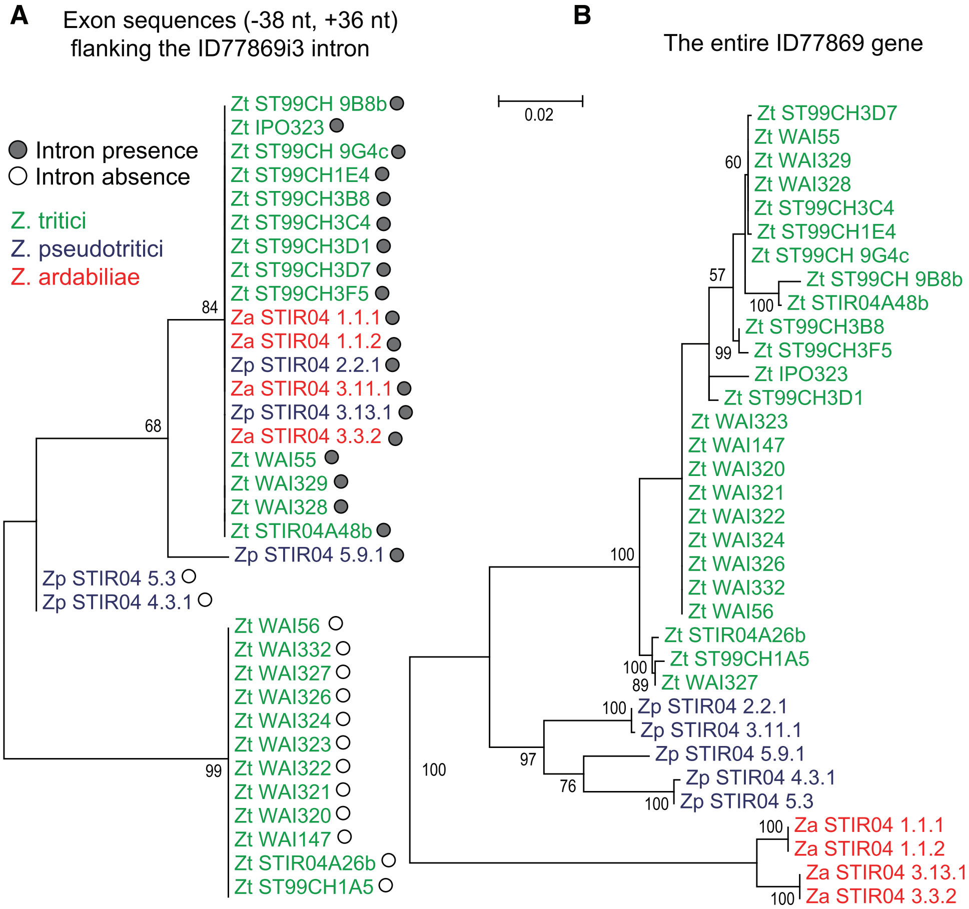

Fig. 4.—Intron introgression at a presence-absence polymorphic locus. (*A*) A phylogenetic tree based on exonic sequences flanking the ID77869i3 intron (i.e., the third intron in the ID77869 gene). (*B*) A phylogenetic tree based on the entire coding sequences of the ID77869 gene. Bootstrap values when > 50% are shown. The boundaries of the flanking coconversion tracts were identified by visual examination.

Discussions

Our findings show that IIHEs are associated with intron movement within the genome and rapid intron turnover among different genomes in *Z. tritici*. It is noteworthy that the detected introns having IIHEs account for only a small fraction (~0.5%) of the total number of verified introns. This is due to our stringent criteria used to define IIHEs, and also the fact that many introns are lost during the processes of rapid intron turnover and exchange. The detected introns having IIHEs may be viewed as the visible tip of the largely hidden iceberg of all (both extant and extinct) introns once having IIHEs. IIHE-mediated intron movement is perhaps a general phenomenon during intron evolution. In fact, intergenic intron homologs have also been observed in algae species *Micromonas* (van Baren et al. 2016), which has undergone recent massive intron invasion (Simmons et al. 2015).

It has been acknowledged that direct introduction of group II introns into nuclear-encoded protein genes is deleterious and unlikely to reach fixation (Doolittle 2014; Qu et al. 2014). IIHEs can fill the gap between the proposed origin of likely deleterious group II intron and observed wide-distribution of modern spliceosomal introns. With intergenic homologs as intermediate stage, insertion of group II introns into intergenic regions in eukaryotes would be less deleterious, provide a buffered environment for subsequent mutations to accumulate, lead to minimal deleterious effects and ultimately facilitate intron proliferation. Unlike typical group II introns, which often encode reverse transcriptase genes, sequence analysis of IIHEs in this study revealed no transposase or reverse transcriptase homologs. The IIHEs are rather similar to the miniature group II introns or miniature transposons that do not encode genes for their transposition. Unfortunately, the genes responsive for the transposition of spliceosomal introns and related IIHEs are still unknown. Further studies are needed to determine the key genes and underlying molecular mechanisms.

We measured very fast rates of intron gain and loss at IPAP loci in *Z. tritici*, which have been suggested in *Neurospora tetrasperma* and *Daphnia* (Li et al. 2014; Sun et al. 2015). Similar to the dynamic IPAP loci, the loci of fixed introns also show fast rates of intron exchange. Furthermore, a

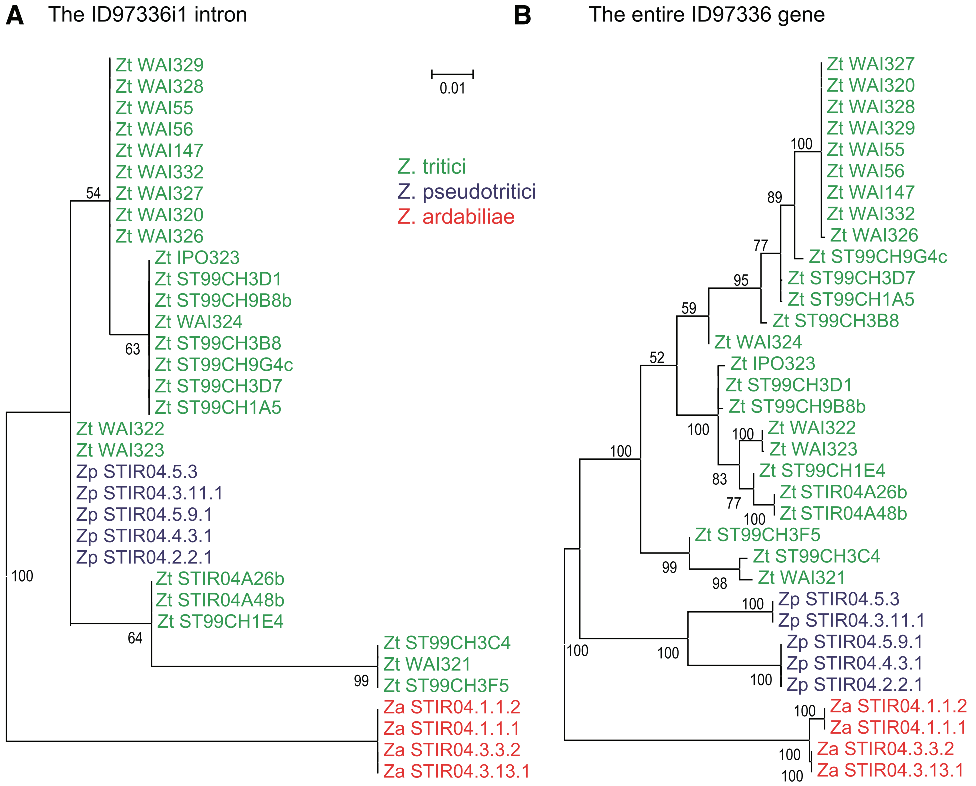

Fig. 5.—Intron introgression between *Z. tritici* and *Z. pseudotritici* at a locus with fixed introns. The phylogenetic tree on the left is based on sequences of the ID97336i1 intron (i.e., the first intron in the ID97336 gene), while the phylogenetic tree on the right is based on the entire coding sequences of the ID97336 gene. Bootstrap values when > 50% are shown.

substantial amount of spliceosomal introns are involved in introgression between different species. It is important to note that these introgression cases of spliceosomal introns are fundamentally different from the published horizontal transfer cases of nuclear introns in fungi and algae (Hibbett 1996; Nikoh and Fukatsu 2001; Coates et al. 2002; Simon et al. 2005), which are self-splicing group I introns. As spliceosomal introns are believed to originate from group II introns (Rogers 1990; Lambowitz and Zimmerly 2004), similar features between sequences and structures have been well documented (Zimmerly and Semper 2015). Here, we observed additional features shared between spliceosomal introns and self-splicing introns with respect to their mobility. 1) Both spliceosomal introns (e.g., in *Z. tritici*) and group I/II introns (Dai and Zimmerly 2002; Wu et al. 2015) may have IPAP among conspecific strains and show high mobility. 2) Splicesomal introns and self-splicing introns could have co-conversion tracts flanking their insertion sites, which are strongly associated with intron invasion and recombination (Lazowska et al. 1994; Moran et al. 1995; Sanchez-Puerta et al. 2008; Wu and Hao 2014). 3) Both spliceosomal introns and self-splicing introns could show invasive nonmendelian transmission with a significant bias towards intron gain in genetic crosses between intron-containing and intron-lacking strains. This has been seen in a spliceosomal intron (the ID39491i2 intron) in *Z. tritici* (Torriani et al. 2011) and in group I/II introns in *Saccharomyces cerevisiae* (Jacquier and Dujon 1985; Moran et al. 1995).

Conclusion

In this study, we demonstrated three crucial stages in the lifecycle of spliceosomal introns, where intron elements can proliferate within the genome, spread horizontally among conspecific strains, and introgress between related species. These processes are very similar to those involving mobile introns and transposable elements (Goddard and Burt 1999; Williams 1999; Schaack et al. 2010), and also consistent with the hypothesis that spliceosomal introns are of a transposon origin (Roy 2004). These processes can partially account for the paradox that rich modern introns exist in a variety of eukaryotic genomes under generally very high rates of intron loss (Csuros et al. 2011). Given the highly variable tempo of spliceosomal intron movement among eukaryotic organisms (Worden et al. 2009; Torriani et al. 2011; van der Burgt et al. 2012; van Baren et al. 2016), we expect the rapidly growing population genomics data, high-quality genome assembly and annotation from diverse eukaryotic organisms to shed new lights on the evolution of modern spliceosomal introns.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

The authors thank three anonymous reviewers for their helpful comments. They are grateful for the grid computing service from Computing and Information Technology of Wayne State University. This work was supported by funds from Wayne State University (to W.H.).

Author Contributions

B.W. and W.H. designed the study. B.W. and A.I.M. performed experiments and analyzed data. B.W., A.I.M., and W.H. wrote the paper.

Literature Cited

Banke S, Peschon A, McDonald BA. 2004. Phylogenetic analysis of globally distributed Mycosphaerella graminicola populations based on three DNA sequence loci. Fungal Genet Biol. 41(2):226–238.

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci USA. 74(8): 3171–3175.

Brunner PC, Torriani SF, Croll D, Stukenbrock EH, McDonald BA. 2014. Hitchhiking selection is driving intron gain in a pathogenic fungus. Mol Biol Evol. 31(7):1741–1749.

Cavalier-Smith T. 1991. Intron phylogeny: a new hypothesis. Trends Genet. 7(5):145–148.

Coates BS, Hellmich RL, Lewis LC. 2002. Nuclear small subunit rRNA group I intron variation among Beauveria spp provide tools for strain identification and evidence of horizontal transfer. Curr Genet. 41(6):414–424.

Collemare J, van der Burgt A, de Wit PJ. 2013. At the origin of spliceosomal introns: is multiplication of introner-like elements the main mechanism of intron gain in fungi? Commun Integr Biol. 6(2): e23147.

Croll D, Lendenmann MH, Stewart E, McDonald BA. 2015. The impact of recombination hotspots on genome evolution of a fungal plant pathogen. Genetics 201(3): 1213–1228.

Croll D, Zala M, McDonald BA, Heitman J. 2013. Breakage-fusion-bridge Cycles and Large Insertions Contribute to the Rapid Evolution of Accessory Chromosomes in a Fungal Pathogen. PLoS Genet. 9(6):e1003567.

Csuros M, Rogozin IB, Koonin EV. 2011. A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. PLoS Comput Biol. 7(9):e1002150.

Dai L, Zimmerly S. 2002. The dispersal of five group II introns among natural populations of *Escherichia coli*. RNA 8(10):1294–1307.

Denoeud F, et al. 2010. Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. Science 330(6009):1381–1385.

Doolittle WF. 2014. The trouble with (group II) introns. Proc Natl Acad Sci USA. 111(18):6536–6537.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.

Farlow A, Meduri E, Schlotterer C. 2011. DNA double-strand break repair and the evolution of intron density. Trends Genet. 27(1):1–6.

Gilbert W. 1978. Why genes in pieces? Nature 271(5645):501.

Goddard MR, Burt A. 1999. Recurrent invasion and extinction of a selfish gene. Proc Natl Acad Sci USA. 96(24):13880–13885.

Goodwin SB, et al. 2011. Finished genome of the fungal wheat pathogen Mycosphaerella graminicola reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLoS Genet. 7(6):e1002070.

Hao W, Golding GB. 2006. The fate of laterally transferred genes: life in the fast lane to adaptation or death. Genome Res. 16(5):636–643.

Hao W, Golding GB. 2010. Inferring bacterial genome flux while considering truncated genes. Genetics 186(1):411–426.

Heyn P, Kalinka AT, Tomancak P, Neugebauer KM. 2015. Introns and gene expression: cellular constraints, transcriptional regulation, and evolutionary consequences. Bioessays 37(2):148–154.

Hibbett DS. 1996. Phylogenetic evidence for horizontal transmission of group I introns in the nuclear ribosomal DNA of mushroom-forming fungi. Mol Biol Evol. 13(7):903–917.

Huff JT, Zilberman D, Roy SW. 2016. Mechanism for DNA transposons to generate introns on genomic scales. Nature 538(7626):533–536.

Jacquier A, Dujon B. 1985. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. Cell 41(2):383–394.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro HN, editor. Mammalian protein metabolism. New York: Academic Press. p. 21–132.

Kim D, et al. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14(4):R36.

Kim T, Hao W. 2014. DiscML: an R package for estimating evolutionary rates of discrete characters using maximum likelihood. BMC Bioinformatics 15(1):320.

Koonin EV. 2006. The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? Biol Direct. 1:22.

Lambowitz AM, Zimmerly S. 2004. Mobile group II introns. Annu Rev Genet. 38:1–35.

Lazowska J, Meunier B, Macadre C. 1994. Homing of a group-Ii intron in yeast mitochondrial-DNA is accompanied by unidirectional co-conversion of upstream-located markers. Embo J. 13(20):4963–4972.

Li WL, Kuzoff R, Wong CK, Tucker A, Lynch M. 2014. Characterization of newly gained introns in daphnia populations. Genome Biol Evol. 6(9):2218–2234.

Li WL, Tucker AE, Sung W, Thomas WK, Lynch M. 2009. Extensive, recent intron gains in daphnia populations. Science 326(5957): 1260–1262.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25(11):1451–1452.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25(5):955–964.

McDonald MC, et al. 2016. Utilizing Gene tree variation to identify candidate effector genes in Zymoseptoria tritici. G3 (Bethesda) 6(4):779–791.

Moran JV, et al. 1995. Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. Mol Cell Biol. 15(5):2828–2838.

Morrison HG, et al. 2007. Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. Science 317(5846):1921–1926.

Mourier T, Jeffares DC. 2003. Eukaryotic intron loss. Science 300(5624):1393.

Nawrocki EP, et al. 2015. Rfam 12.0: updates to the RNA families database. Nucleic Acids Res. 43(Database issue):D130–D137.

Nikoh N, Fukatsu T. 2001. Evolutionary dynamics of multiple group I introns in nuclear ribosomal RNA genes of endoparasitic fungi of the genus Cordyceps. Mol Biol Evol. 18(9):1631–1642.

Palmer JD, Logsdon JM Jr. 1991. The recent origins of introns. Curr Opin Genet Dev. 1(4):470–477.

Qu GS, et al. 2014. RNA-RNA interactions and pre-mRNA mislocalization as drivers of group II intron loss from nuclear genomes. Proc Natl Acad Sci USA. 111(18):6612–6617.

Rogers JH. 1990. The role of introns in evolution. FEBS Lett. 268(2):339–343.

Roy SW. 2004. The origin of recent introns: transposons? Genome Biol. 5(12):251.

Roy SW, Gilbert W. 2005. Rates of intron loss and gain: implications for early eukaryotic evolution. Proc Natl Acad Sci USA. 102(16): 5773–5778.

Rudd JJ, et al. 2015. Transcriptome and metabolite profiling of the infection cycle of Zymoseptoria tritici on wheat reveals a biphasic interaction with plant immunity involving differential pathogen chromosomal contributions and a variation on the hemibiotrophic lifestyle definition. Plant Physiol. 167(3):1158–1185.

Sanchez-Puerta MV, Cho Y, Mower JP, Alverson AJ, Palmer JD. 2008. Frequent, phylogenetically local horizontal transfer of the *cox1* group I Intron in flowering plant mitochondria. Mol Biol Evol. 25(8):1762–1777.

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(9):537–546.

Simmons MP, et al. 2015. Intron invasions trace algal speciation and reveal nearly identical arctic and antarctic micromonas populations. Mol Biol Evol. 32(9):2219–2235.

Simon D, Moline J, Helms G, Friedl T, Bhattacharya D. 2005. Divergent histories of rDNA group I introns in the lichen family Physciaceae. J Mol Evol. 60(4):434–446.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22(21):2688–2690.

Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33(Web Server issue):W465–W467.

Stukenbrock EH, et al. 2011. The making of a new pathogen: insights from comparative population genomics of the domesticated wheat pathogen Mycosphaerella graminicola and its wild sister species. Genome Res. 21(12):2157–2166.

Sun Y, Whittle CA, Corcoran P, Johannesson H. 2015. Intron evolution in Neurospora: the role of mutational bias and selection. Genome Res. 25(1): 100–110.

Sverdlov AV, Rogozin IB, Babenko VN, Koonin EV. 2004. Reconstruction of ancestral protosplice sites. Curr Biol. 14(16):1505–1508.

Torriani SFF, Stukenbrock EH, Brunner PC, McDonald BA, Croll D. 2011. Evidence for extensive recent intron transposition in closely related fungi. Curr Biol. 21(23):2017–2022.

Trapnell C, et al. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 7(3):562–578.

van Baren MJ, et al. 2016. Evidence-based green algal genomics reveals marine diversity and ancestral characteristics of land plants. BMC Genomics 17:267.

van der Burgt A, Severing E, de Wit PJ, Collemare J. 2012. Birth of new spliceosomal introns in fungi by multiplication of introner-like elements. Curr Biol. 22(13):1260–1265.

Verhelst B, Van de Peer Y, Rouze P. 2013. The complex intron landscape and massive intron invasion in a picoeukaryote provides insights into intron evolution. Genome Biol Evol. 5(12):2393–2401.

Williams TM. 1999. The evolution of cost efficient swimming in marine mammals: limits to energetic optimization. Philos Trans R Soc Lond B Biol Sci. 354(1380):9.

Worden AZ, et al. 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas. Science 324(5924):268–272.

Wu B, Buljic A, Hao W. 2015. Extensive horizontal transfer and homologous recombination generate highly chimeric mitochondrial genomes in yeast. Mol Biol Evol. 32(10):2559–2570.

Wu B, Hao W. 2014. Horizontal transfer and gene conversion as an important driving force in shaping the landscape of mitochondrial introns. G3 (Bethesda) 4(4):605–612.

Yenerall P, Zhou LM. 2012. Identifying the mechanisms of intron gain: progress and trends. Biol Direct. 7:29.

Zimmerly S, Semper C. 2015. Evolution of group II introns. Mob DNA. 6:7.

Associate editor: Dennis Lavrov