Distribution of Conventional and Nonconventional Introns in Tubulin (α and β) Genes of Euglenids

Rafał Milanowski,*[,1] Anna Karnkowska,[1,2] Takao Ishikawa,[3] and Bożena Zakryś[1]

[1]Department of Plant Systematics and Geography, Institute of Botany, Faculty of Biology, University of Warsaw, Warsaw, Poland

[2]Department of Parasitology, Faculty of Science, Charles University in Prague, Prague, Czech Republic

[3]Department of Molecular Biology, Institute of Biochemistry, Faculty of Biology, University of Warsaw, Warsaw, Poland

***Corresponding author:** E-mail: milan@biol.uw.edu.pl.

Associate editor: John Logsdon

Abstract

The nuclear genomes of euglenids contain three types of introns: conventional spliceosomal introns, nonconventional introns for which a splicing mechanism is unknown (variable noncanonical borders, RNA secondary structure bringing together intron ends), and so-called intermediate introns, which combine features of conventional and nonconventional introns. Analysis of two genes, *tubA* and *tubB*, from 20 species of euglenids reveals contrasting distribution patterns of conventional and nonconventional introns—positions of conventional introns are conserved, whereas those of the nonconventional ones are unique to individual species or small groups of closely related taxa. Moreover, in the group of phototrophic euglenids, 11 events of conventional intron loss versus 15 events of nonconventional intron gain were identified. A comparison of all nonconventional intron sequences highlighted the most conserved elements in their sequence and secondary structure. Our results led us to put forward two hypotheses. 1) The first one posits that mutational changes in intron sequence could lead to a change in their excision mechanism—intermediate introns would then be a transitional form between the conventional and nonconventional introns. 2) The second hypothesis concerns the origin of nonconventional introns—because of the presence of inverted repeats near their ends, insertion of MITE-like transposon elements is proposed as a possible source of new introns.

Key words: **euglenids, nonconventional introns, conventional spliceosomal introns, tubulin gene.**

Article

Introduction

Euglenids (Euglenida) together with heterotrophic flagellates diplonemids (Diplonemea), symbiontids (Symbiontida), and kinetoplastids (Kinetoplastea) form an ancient group Euglenozoa within the supergroup Excavata (Hampl et al. 2009; Adl et al. 2012). Some reports, however, indicate the Euglenozoa as the first group to branch off from the main evolutionary lineage of eukaryotes (Cavalier-Smith 2010). No process of sexual reproduction or any other process of exchanging genetic material has been observed in euglenids so far.

The phylogeny of euglenids is under intensive study in recent years, and reliable and consistent phylogenetic trees describing the evolution within this group have been obtained (Linton et al. 1999; Marin et al. 2003; Milanowski et al. 2006; Triemer et al. 2006; Linton et al. 2010; Yamaguchi et al. 2012). They led to the conclusion that all phototrophic euglenids form a monophyletic lineage. Their common ancestor (a heterotrophic euglenid) acquired the chloroplast by secondary endosymbiosis with green algae, probably from *Pyramimonas* genus (Prasinophyceae; Turmel et al. 2009; Hrdá et al. 2012). The best established are phylogenetic relationships within phototrophic euglenids, which are grouped in 12 genera.

Our understanding of the organization of the genetic material in euglenids is limited. Their chloroplast genomes are the best described: the complete plastid genomes of *Euglena gracilis* (Hallick et al. 1993), *E. longa* (Gockel and Hachtel 2000), *Eutreptiella gymnastica* (Hrdá et al. 2012), *Eutreptia viridis* (Wiegert et al. 2012), *Monomorphina aenigmatica* (Pombert et al. 2012), *Colacium vesiculosum*, *Strombomonas costata* (Wiegert et al. 2013), and *E. viridis* (Bennett et al. 2012) are known. Very little is known about mitochondrial genomes of euglenids, but their structure is probably unique. Earlier studies indicated that the mitochondrial genome of *E. gracilis* comprises numerous, short and long, circular and linear DNA molecules (Roy et al. 2007). The organization of the nuclear genetic material in euglenids is also unclear. Surprisingly, even the number of chromosomes in the best-examined species *E. gracilis* is uncertain (Dooijes et al. 2000). Currently, the *E. gracilis* genome sequencing project is in progress: its size is unexpectedly large and is estimated at about 250,000,000 bp (Goldstamp: Gi07537). Several unusual features of its organization are also observed: rRNA genes are located on the extrachromosomal, circular molecules present in the cell in hundreds to thousands of copies (Cook and Roxby 1985; Ravel-Chapuis 1988); there is no single full-length 28S rRNA—instead, 13 short RNA molecules form its equivalent (Schnare and Gray 1990). Moreover, in euglenids, a splice leader is transferred to the 5′ end of most pre-mRNAs in the process of spliceosome-dependent trans-splicing (Ebel et al. 1999).

An unprecedented feature of euglenid genomes is the presence of three types of introns: 1) conventional spliceosomal introns with canonical GT/C-AG borders, 2) nonconventional introns for which a splicing mechanism is unknown (noncanonical and variable borders, RNA secondary structure bringing together ends of the intron), and 3) so-called intermediate introns. It should be noted here that an analysis of complete genomes of trypanosomatids (relatives of euglenids) has revealed very few introns. In these parasitic flagellates, only two genes—those encoding tRNA (Tyr) (Schneider et al. 1994) and poly(A) polymerase (Mair et al. 2000)—are disrupted by introns, whereas in *E. gracilis* most examined genes contain introns. Conventional spliceosomal introns were found in genes encoding fibrillarin (Breckenridge et al. 1999; Russell et al. 2005), α-, β-, and γ-tubulin (Canaday et al. 2001), and in genes encoding proteins targeted to chloroplasts: *eno29*, *gapA*, *pbgD*, *petA*, *psaF*, *psbM*, and *psbW* (Vesteg et al. 2010). The presence of conventional introns has also been confirmed in the heterotrophic euglenids *Enthosiphon sulcatum* (*tubB*; Ebel et al. 1999) and *Peranema trichophorum* (*hsp90*; Breglia et al. 2007). *Euglena gracilis* genes also contain so-called nonconventional introns, which do not have GT/C-AG borders (nor AT-AC borders characteristic of some minor spliceosomal introns). However, very little is known about this type of introns—neither the mechanism of their removal nor any factors involved have been recognized. They are apparently removed by a spliceosome-free mechanism, because their 5′ ends are not complementary to the U1 snRNA (Breckenridge et al. 1999). It has only been noted that all introns of this type form a stable RNA secondary structure bringing the two splice sites together, which is probably needed for their proper removal. This structure, however, is not conserved and shows no common features with the self-splicing group I, II, or III introns (Muchhal and Schwartzbach 1994; Henze et al. 1995; Tessier et al. 1995; Canaday et al. 2001). It also seems that the nonconventional introns are excised before the addition of the spliced leader at the 5′ end of nuclear pre-mRNAs (trans-splicing) and before the excision of conventional spliceosomal introns (Tessier et al. 1995). Initially, introns of this type were only observed in nuclear genes of chloroplast origin (*lhcp2*, *rbcS*; Muchhal and Schwartzbach 1994; Tessier et al. 1995) and in the *gapC* gene of eubacterial origin (Henze et al. 1995). On this basis it was hypothesized that nonconventional introns were derived from the genome of a secondary endosymbiont (Ebel et al. 1999). Later, however, nonconventional introns were also found in *E. gracilis* genes encoding α- and β-tubulin (Canaday et al. 2001), which derived from the genome of the host cell. A comparison of intron positions in the tubulin genes of diverse species suggested that most of the conventional introns were evolutionarily old, as they occurred in the same positions in other organisms, whereas the nonconventional introns were present in unique positions, suggesting that they were evolutionarily younger than the spliceosomal ones (Canaday et al. 2001). In 2007, the presence of nonconventional introns, besides conventional ones, was revealed in the *hsp90* gene from the primarily heterotrophic euglenid species *P. trichophorum* (Breglia et al. 2007), which put into question the earlier hypothesis on an endosymbiotic origin of the nonconventional introns. Recently, nonconventional introns were also found in two genes encoding plastid-targeted proteins *petJ* and *psbW* from *E. gracilis* (Vesteg et al. 2010). Some authors also distinguish a third type of nuclear introns in *E. gracilis*, so-called "intermediate" introns, as observed in the genes encoding α- and β-tubulin (Canaday et al. 2001) and fibrillarin (Russell et al. 2005). They form a stable secondary structure bringing together the ends of the introns, one or even both intron borders are consistent with the GT/C-AG rule, and the 5′ end of the intron is to some extent complementary to the U1 snRNA (however, the complementarity is much weaker than that of conventional introns).

To date, there are no data about the evolution of introns in euglenids, and little is known about the distribution of introns in nuclear genes in the context of the group's phylogeny. It is thus impossible to answer basic questions such as whether the conventional, nonconventional, and intermediate introns occur in conserved positions in different evolutionary lineages or whether introns of one type can be replaced by introns of another type. Answering these questions seems to be crucial for addressing more complex issues, such as the mechanism of excision of nonconventional introns. These answers should also help to understand the role of nonconventional introns in the functioning and evolution of euglenid genomes; a recent study of genes encoding chloroplast-targeted proteins suggests that both conventional and nonconventional introns may be hot spots of DNA recombination. It was proposed that this mechanism leads to the replacement or acquisition of plastid-targeting leader sequences (Vesteg et al. 2010).

As reviewed above, understanding the distribution of conventional, nonconventional, and intermediate introns may be key to resolving basic issues concerning the evolution of euglenids and their genomes. Here, we report an analysis of intron distribution in two nuclear genes, *tubA* and *tubB*, from 20 species (18 phototrophic or secondary heterotrophic and 2 primary heterotrophic) representing 13 genera of euglenids.

Results and Discussion

Types of Introns in Tubulin Genes

Using nested polymerase chain reaction (PCR) amplification with degenerate primers on genomic DNA, 30 *tubA* and 39 *tubB* sequences were obtained, encompassing 78% and 86% of coding regions, respectively. More than one form of the *tubA* gene was obtained for 8 and of the *tubB* gene for 11 species; the differences between the forms from the same species were minor and mostly concerned sequence differences within introns and the third position of codons. To determine the intron positions, the genomic sequences were compared with those of cDNAs obtained by reverse transcription of mRNA. Introns were absent in both genes from two representatives of Eutreptiales, *Eutreptia viridis* and *Eut. pomquetensis*, whereas for another representative of Eutreptiales, *Eut. braarudii*, and for the primary heterotroph *Ent. sulcatum*, no introns were found in the *tubA* gene. In the remaining genes between one and five introns were present,

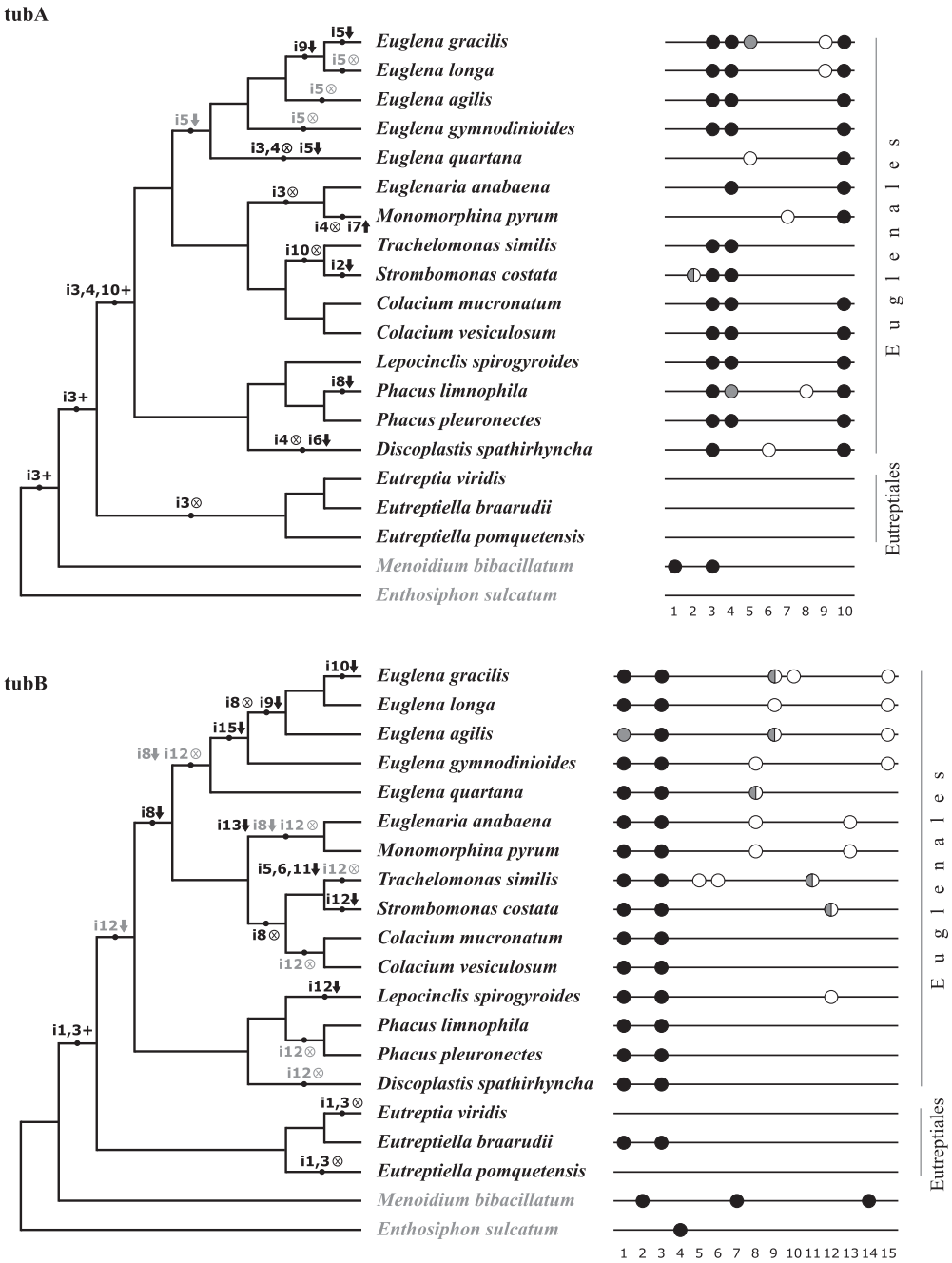

Fig. 1. Distribution and postulated evolution of introns in *tubA* and *tubB* genes of euglenids. Left: schematic phylogenetic trees reflecting phylogeny of euglenids (Busse et al. 2003; Marin et al. 2003; Linton et al. 2010). Species names of primary heterotrophs in gray, phototrophs and secondary heterotrophs in black. Predicted intron presence in common ancestors (+) and postulated events of intron gain (↓) and loss (x) are indicated on branches (alternative scenarios in gray). Right: distribution of introns in *tubA* and *tubB* genes. Circles on horizontal bars next to species names indicate introns in the specific positions; black circles: conventional introns; gray: conventional/intermediate; gray and white: intermediate/nonconventional; white: nonconventional.

occupying 10 and 15 unique positions in *tubA* and *tubB*, respectively. Their locations are depicted schematically in figure 1 together with the proposed type and a phylogenetic tree of the respective species; supplementary tables S1 and S2, Supplementary Material online, show sequences of intron junctions in the tubulin genes.

Introns with the consensus border sequence GT/C-AG are present in *tubA* genes in positions indicated in figure 1 as 1 (in the primary heterotroph *Menoidium bibacillatum*), 3 (*M. bibacillatum* and representatives of Euglenales), 4 and 10 (Euglenales), and 5 (*E. gracilis*); in *tubB* genes in positions 1 and 3 (Euglenales), 2, 7, and 14 (*M. bibacillatum*), and 4 (*Ent. sulcatum*). They were classified as conventional spliceosomal introns (C), except for introns in positions 4 and 5 in *tubA* genes from *Pha. limnophila* and *E. gracilis*, as well as the intron in position 1 in the *tubB* gene from *E. agilis*, for which

stable secondary structure bringing together intron ends was predicted—they were classified as conventional/intermediate introns (C/I).

Most introns without the consensus sequence GT/C-AG at both ends are present in positions unique to single species (*tubA*: 2, 6, 7, and 8; *tubB*: 5, 6, 10, and 11) or common for several closely related species composing well-defined clades (*tubA*: 9; *tubB*: 8, 9, 13, and 15). The exceptions are *tubB* introns in position 12, present in two relatively distant species *S. costata* and *Lepocinclis spirogyroides*; another exception is the *tubA* intron in *E. quartana* in position 5, where a C/I intron is also present in *E. gracilis*. The *E. gracilis* intron contains direct CAG repeats at the intron–exon junctions, and four alternative positions can be defined for it; the position preserving the GC-AG ends of intron is one nucleotide downstream from the position of the *E. quartana* intron.

The introns lacking the consensus GT/C-AG sequences at ends were initially classified as nonconventional ones (N); all of them can form a stable stem-loop RNA structure bringing the splice sites together. As most of them have direct repeats at the intron–exon junctions, it is difficult to determine their exact positions. To predict the most likely locations, all known nonconventional introns lacking repeats, for which the locations could be determined unambiguously (14 introns), were compared and the nucleotide logo of their junctions was created (supplementary fig. S1, Supplementary Material online). The secondary structure of these introns was predicted as well. The ends of all the nonconventional introns in the tubulin genes were then fitted to the consensus obtained as above (supplementary tables S1 and S2, Supplementary Material online) and according to the secondary structures predicted for the introns used to obtain the consensus. Six introns with the 5′ end conforming to the GT/C rule were excluded from the group of nonconventional introns and were classified as intermediate/nonconventional ones (I/N).

Tubulin Genes with Nonconventional Introns—Active or Inactive Forms?

As it was mentioned above, in some cases more than one form of gene was obtained suggesting the presence of intraspecies polymorphisms or many forms of the gene in a single genome. In both cases, the question is whether all cloned forms of tubulin genes are active. This problem especially concerns genes in which nonconventional introns are observed—if these genes are inactive, nonconventional introns should be considered as undefined insertion elements, which are not spliced out, rather than as introns. To avoid including inactive forms of genes in the analysis, the genomic and cDNA sequences were compared (for details, see Materials and Methods); however, we cannot rule out the possibility that two copies of the gene—active form without introns and another, inactive one, with introns and/or insertion elements—coexist in a single genome. On the other hand, each PCR amplification of genomic DNA generated individual products corresponding to the forms containing introns, whereas products corresponding to the intronless forms were not detected. This result suggests that nonconventional introns are indeed present in active forms of genes.

Phylogenetic Analysis

The trees obtained using the *tubA* and *tubB* sequences do not faithfully reflect the phylogeny of euglenids, because of the rather small amount of data related to a highly conserved protein (supplementary fig. S2, Supplementary Material online). Nevertheless, all well-supported relationships from those trees are consistent with a more reliable phylogenetic tree developed previously with the use of more sophisticated phylogenetic analyses of larger data sets (fig. 1).

Toward a Model of Nonconventional Intron Structure

To find common features of the nonconventional introns, the sequences of all known introns (74) were compared and a sequence logo for euglenid nonconventional intron junctions was created (fig. 2). Comparing the obtained sequence logo with the RNA secondary structure of nonconventional introns, two nucleotides at positions +4, +5 (5′ end of intron) and complementary nucleotides at positions −7, −6 (3′ end of intron) stand out as their most conserved feature; in most cases, CA/TG are present at these positions (a representative structure is shown in fig. 2). Other nucleotides involved in the maintenance of the stem-loop structure are less conserved or not conserved at all. Other conserved features observed for most nonconventional introns is the presence of a pyrimidine at the 3′ of the upstream exon and at the 3′ end of the intron, a purine at the 5′ end of the intron and at the 5′ end of the downstream exon, and a C at the third position of the downstream exon. Although nucleotides at positions +2 and +3 and −2, −3, −4, and −5 of the intron are not conserved, their ability to pair is preserved in some introns (e.g., intron in fig. 2). The conserved features mentioned above are in agreement with the observation of Muchhal and Schwartzbach (1994) for nonconventional introns in LHCPII-coding genes of *E. gracilis*.

Conventional Introns

Conventional spliceosomal introns are unique to eukaryotes. Although the intron density varies among different evolutionary lineages of eukaryotes, no representative of this group lacking spliceosomal introns are known. The lack of introns in the nucleomorph genome of *Hemiselmis andersenii* (Lane et al. 2007) is a special case—the nucleomorph located in the periplastid space of the chloroplast is just a remnant nucleus of a secondary endosymbiont. Due to their similar mechanism of excision, it is widely accepted that spliceosomal introns have evolved from self-splicing group II introns. It appears that the common ancestor of eukaryotes acquired them with the genome of an endosymbiotic α-proteobacterium, the progenitor of mitochondria. However, there are also opinions indicating that group II introns and eukaryotic spliceosomes only share a common ancestor, namely the proto-spliceosome, which evolved in the RNA world as a mechanism to excise functional RNAs from the ancient RNA genomes (Vesteg et al. 2012).

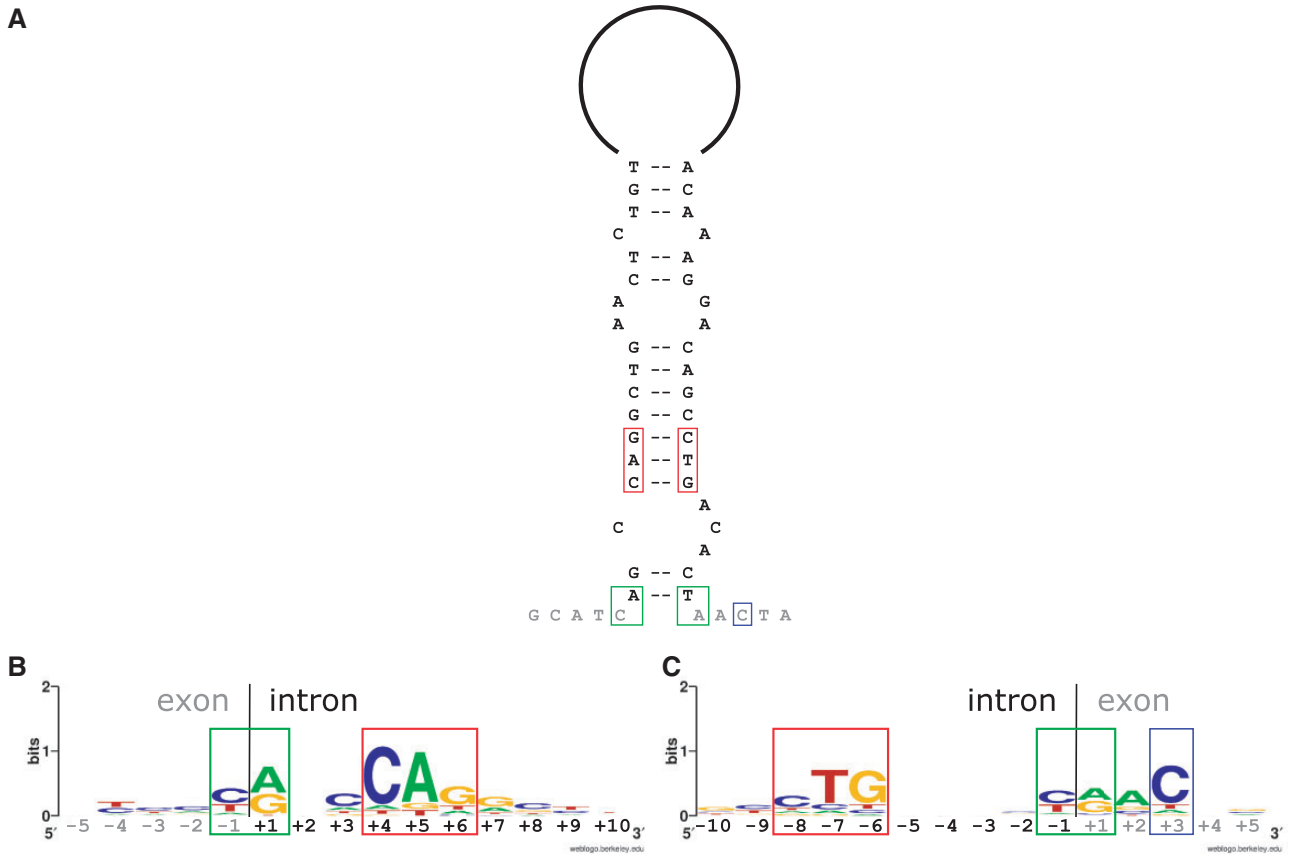

FIG. 2. Model of nonconventional intron junctions. (*A*) Intron 9 in *tubA*2 gene from *Euglena gracilis*—an example of nonconventional intron secondary structure; exons in gray, intron in black; the most conserved nucleotides are boxed. Sequence logo of 5′ (*B*) and 3′ (*C*) exon/intron junctions (splice sites indicated by vertical lines) created from 74 nonconventional intron sequences (36 introns used in this study, introns in *tubA*, *tubB*, *rbcS*, *lhcp2*, *gapC*, *nop1p*, *psbJ*, *psbO*, and *psbW* genes from *E. gracilis*, and two introns in *hsp90* gene from *P. trichophorum*).

When considering the evolution of spliceosomal introns, one has to mention a long-lasting debate between supporters of the "intron-early" and "intron-late" hypotheses. The first hypothesis assumed that introns were present in the prokaryotic ancestor of eukaryotes, and the present differences in their occurrence in different evolutionary lineages are due to their independent loss. According to the latter theory, spliceosomal introns, being a typically eukaryotic "invention," were inserted into originally intron-free genes. At present, a synthetic theory of the evolution of spliceosomal introns, combining both hypotheses, is widely accepted. It can be summarized as "many introns early in eukaryotic evolution" and has a strong support in a wide range of data (for review, see Rogozin et al. 2012).

Spliceosomal introns are common in tubulin-coding genes throughout eukaryotes; some of them are unique to individual groups, whereas others occur in the same positions in distantly related species (Perumal et al. 2005), which is consistent with the above synthetic concept. A similar situation is also observed for the conventional spliceosomal introns in the euglenid genes coding for α- and β-tubulins: some occur in positions shared by distantly related species (introns 3 and 4 in *tubA*), while the positions of others are unique. Unfortunately, in other representatives of Excavata examined to date, the *tubA* and *tubB* genes are intronless and it is difficult to reconstruct the pattern of intron distribution in the putative common ancestor of all the taxa analyzed in this study. It is much easier to analyze the distribution of conventional introns exclusively in photosynthetic euglenids—the most parsimonious hypothesis is that the common ancestor of all phototrophic taxa had introns in positions 1 and 3 in the *tubB* gene and at least one intron (position 3) in *tubA*. The common ancestor of Euglenales had two more introns in the *tubA* gene—in positions 4 and 10. Based on an analysis of intron distribution in the photosynthetic euglenids, 11 events of intron loss are predicted: in the *tubA* gene, introns 3 and 4 were lost in *E. quartana*, intron 3 in the common ancestor of *Euglenaria anabaena* and *M. pyrum*, intron 4 in *M. pyrum*, intron 10 in the ancestor of *Trachelomonas similis* and *S. costata*, intron 4 in *Discoplastis spathirhyncha*, and intron 3 in the common ancestor of Eutreptiales; in the *tubB* gene, introns 1 and 3 were lost independently in *Eutreptia viridis* and in *Eut. pomquetensis* (see fig. 1). On the other hand, there is no clear-cut evidence indicating a conventional intron gain in the *tubA* or *tubB* genes of photosynthetic euglenids. However, such an event could have taken place in their common ancestor—the distribution of introns differed substantially between hetero- and phototrophs.

Nonconventional Introns

Only limited data about nonconventional introns of euglenids are available. It is not known how they are removed or

what their origin is; consequently, the relationship between conventional spliceosomal and nonconventional introns remains a mystery. Considering the lack of sequence conservation of intron junctions and the weakness or absence of base pairing between the 5′ junction and U1 snRNA, a spliceosome-free mechanism of their excision seems the most likely. One is tempted to postulate the existence of an endonuclease, possibly similar to the endonucleases taking part in the splicing of tRNA introns in archaea and eukaryotes (for review, see Calvin and Li 2008), where a conserved secondary structure of the pre-tRNA is recognized and intron removed, followed by exons joining by a tRNA ligase. A similar mechanism is responsible for excising introns from mRNA and rRNA in archaea (for review, see Calvin and Li 2008). A case of nonconventional intron removal from pre-mRNA by endonuclease not involved in tRNA splicing is also known—in yeast, mammals, and plants, the endonuclease Ire1 removes an intron from mRNA encoding a transcription factor involved in the unfolded protein response (Gonzalez et al. 1999; Yoshida et al. 2001; Nagashima et al. 2011); however, no ortholog of this endonuclease has been found in the *E. gracilis* EST database (Russell et al. 2005).

The nonconventional introns of the euglenid tubulin genes show a distribution pattern different to that of the conventional ones. The nonconventional introns are present in five positions in the *tubA* gene and in nine in *tubB*; none of those 14 positions is common to all representatives of Euglenales, and 7 are unique to single species (fig. 1). This distribution pattern suggests that they are relatively recent and the process of nonconventional intron gain is much more frequent than in the case of conventional ones. Our conclusion is consistent with the original suggestion of Canaday et al. (2001), who reported the presence of nonconventional introns in unique positions in *E. gracilis* tubulin genes for the first time. An apparent recent gain of numerous introns is not limited in Euglenales to nuclear genes—in their chloroplast genomes, a massive gain of self-splicing introns is also observed (Pombert et al. 2012; Wiegert et al. 2013). Such a widespread occurrence of intron gain suggests the action of undefined evolutionary pressure promoting the presence of intervening sequences in various genomes or a common mechanism of their spreading in different genomes. There are no common features between the nonconventional nuclear and the self-splicing organellar introns; however, the secondary structures of group II and III introns dominating in the euglenid chloroplast genomes sometimes differ from the conserved model (Hrdá et al. 2012). Possible relationships between these two intron types deserve further investigation.

Do Intermediate Introns Really Exist?

Intermediate introns were originally defined by Canaday et al. (2001), who noted that some introns in *E. gracilis tubA* and *tubB* genes (intron 5 in *tubA* and 9 in *tubB*, according to the numbering in fig. 1) combine features of both conventional (base paring with U1 snRNA, presence of both consensus intron borders or at least the 5′ consensus GT/C) and nonconventional introns (stable RNA secondary structure). Another intermediate intron was defined in *E. gracilis* fibrillarin gene (junctions AG | gatc . . . ggag | GA, secondary structure present; Russell et al. 2005). The idea of distinguishing that type of introns as a potential transitional form between conventional and nonconventional introns seems attractive (Russell et al. 2005), although the direction of such a transition is unclear. In this study, intermediate introns were subdivided into two groups: C/I and I/N. The former group includes cases where junctions typical for conventional introns can be found at any of possible intron positions (more than one possible position is the effect of direct repeats at the junctions) and a stable secondary structure is observed. Three such introns were found: intron 4 in *Pha. limnophila tubA*, intron 5 in *E. garcilis tubA*, and intron 1 in *E. agilis tubB*. Intron 4 in the *Pha. limnophila tubA* gene and intron 1 in the *E. agilis tubB* gene are in positions shared with conventional introns—the question then is whether their secondary structures arose accidentally in the ancestral conventional introns or whether they reflect kinship with nonconventional introns? Stable secondary structures can be formed by many cis- and trans-spliced conventional introns (Chen and Stephan 2003; van der Burgt et al. 2012, Roy et al. 2012) and could be responsible for bringing together the two intron ends before the excision by spliceosome. The nucleotide sequences and the predicted secondary structures of the *Pha. limnophila* and *E. agilis* C/I introns do not fit well the model for nonconventional intron junctions (fig. 2), which could mean that they are in fact conventional ones. However, these introns should also be considered as a potential transitional form between conventional and nonconventional introns, because a mutation disturbing canonical GT-AG ends could lead to a change in their excision mechanism.

The third intron defined as C/I occurs in position 5 in the *E. garcilis tubA* gene—four alternative positions are possible for it, the last one giving GC-AG intron ends (fig. 3). In contrast to the *Pha. limnophila* and *E. agilis* C/I introns, no conventional introns are present at this position in other species, but one nucleotide upstream a nonconventional intron 5 is present in *E. quartana*, with an exactly defined position. When the *E. gracilis* intron is assumed to share position with the *E. quartana* one, its ends no longer conform to the classical intron rule but instead fit the model for nonconventional intron junctions (fig. 3). Thus, the question of the mechanism of intron 5 removal in *E. gracilis* remains open. The shared position with the *E. quartana* intron, the similarity to the model for nonconventional intron junctions, and a lack of a typical polypyrimidine track upstream of the 3′ end of the intron suggest that this intron is most likely of the nonconventional type. The fact that no other newly acquired conventional introns can be found in Euglenales also supports this conclusion. By chance, the direct repeats present at this intron's borders would also allow its excision by the conventional spliceosomal mechanism at a position one nucleotide downstream from the original one with preservation of the mRNA coding potential.

Six introns were defined as I/N (only 5′ end GT/C conserved); four of them occur in positions shared with nonconventional introns (8, 9, and 12 in *tubB*), whereas two are in

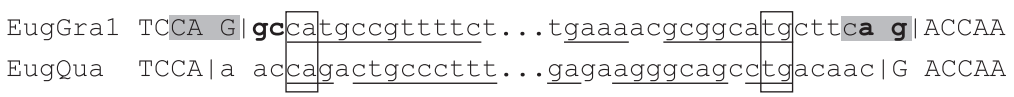

Fig. 3. Comparison of intron 5 sequences in *tubA* genes from *Euglena gracilis* and *E. quartana*. Four positions are possible for the *E. gracilis* intron because of the presence of direct repeats (shaded); in the fourth position the intron has GC-AG ends (in bold). If the *E. gracilis* intron is defined to coincide with the *E. quartana* intron position, the most conserved nucleotides in the secondary structure of nonconventional introns CA-TG (boxed) are present in conserved positions +4, +5 (5′ end of intron) and −7, −6 (3′ end of intron). Nucleotides involved in forming the intron secondary structure are underlined; exon sequences are shown in upper case and intron sequences in lower case.

positions unique to single species—intron 2 in *S. costata tubA* and intron 11 in the *T. similis tubB* gene. The I/N intron in the *S. costata tubB* gene is in the same position as nonconventional intron in *L. spirogyroides* (position 12); these species are not closely related and it is unclear whether these introns were gained independently or are derived from a common ancestor (both scenarios are shown in fig. 1). The first scenario needs two intron gains, whereas the second one needs a single intron gain and six independent intron losses. The position of intron 12 in *S. costata* is unequivocal, whereas in *L. spirogyroides* as many as nine positions are possible. However, the position shared with *S. costata* fits the best the model for nonconventional intron junctions. The fundamental question regarding the I/N introns is whether they really are intermediate introns or are simply nonconventional ones with the 5′ GC/T ends formed by random mutations. According to the sequence logo for the 5′ ends of nonconventional introns, most of them have a purine at the first position, while the second one is not conserved (fig. 2). Therefore, the probability of a GT/C end is quite high. A different situation is observed at the 3′ end of nonconventional introns, where a pyrimidine is preferred at the last position and probability of an AG end seems low; in fact, not a single intron in the α- and β-tubulin gene can be found with AG at the 3′ end and a nonconventional 5′ end. It therefore seems likely that the 5′ GT/C ends of I/N introns arose by mutations preserving the RN end (a puRine followed by aNy nucleotide) typical of nonconventional introns, and they are still excised by mechanism specific for nonconventional introns. It should be emphasized, however, that nonconventional introns with the GT/C 5′ ends are more susceptible to becoming transformed into conventional ones. Crucial is the transversion of the last pyrimidine at the 3′ end of nonconventional intron to G (pyrimidine is preferred at the 3′ end of nonconventional introns). Such a situation is observed at the 3′ end of an intermediate (I/N according to the nomenclature used in this study) intron in the *E. gracilis* fibrillarin gene (Russell et al. 2005), with GA at the 5′ and AG at the 3′ end; it can also be folded into a stable secondary structure fitting the model for nonconventional intron junctions very well. This intron seems to be a better example of the I/N type than are introns in the tubulin genes, which are possibly regular nonconventional introns.

The Origin of Nonconventional Introns

As discussed above, a change of the mechanism of intron excision through mutational changes of the intron sequence is plausible. However, when one takes into account the different patterns of distribution of conventional and nonconventional introns, it seems unlikely that such a change is the only mechanism of nonconventional intron gain. On the contrary, the 15 events of nonconventional intron gains versus none for conventional intron in Euglenales suggest an independent origin of nonconventional introns. It is widely accepted that conventional spliceosomal introns evolved from group II self-splicing introns in the common ancestor of all eukaryotes. What is more, at least seven mechanisms of spliceosomal intron gain in new positions have also been proposed (for review, see Yenerall and Zhou 2012). In contrast, for the nonconventional introns, both the time and the mechanism of their acquisition in the evolutionary past remain a mystery. Whether any of the mechanisms proposed to explain the gain of conventional introns also functions in the case of nonconventional intron gain or whether the latter is unique to euglenids is unknown; it cannot be excluded that mechanism of nonconventional intron gain is closely linked to the mechanism of intron removal. In this context, one of the mechanisms of conventional intron gain—transposon insertion—deserves particular attention (Giroux et al. 1994; Roy 2004). To date, no transposable elements have been identified in euglenid genomes, but it does not necessarily mean that they are absent—those genomes have only been explored rather cursorily. The acquisition of stable secondary structure by nonconventional intron RNA is made possible by the presence of inverted repeats at the intron ends. Inverted repeats are also characteristic for transposon ends—including miniature inverted-repeat transposable elements (MITEs) (Feschotte et al. 2002; Casacuberta and Santiago 2003; Jiang et al. 2004; Lu et al. 2012). MITEs are short (usually less then 1 kb), AT-rich structures which do not encode proteins; they are flanked by short (about 15 bp), often imperfect, inverted repeats that enable them to form stem-loop RNA structures. They have been found in numerous eukaryotic genomes, including plants, animals, and human; their location is often conserved between related taxa (Wessler 1998). MITEs are nonautonomous and require for mobilization enzymes encoded by other transposons from which they probably originated (Jiang et al. 2004). Despite being widespread in eukaryotic genomes, mechanism of their transposition remains unclear. Interestingly, they are preferentially located in euchromatin—within gene promoters, terminators, and introns or even coding sequences. The origin of nonconventional introns from MITE-like elements should be considered as a plausible hypothesis. Inserted in new exonic genome locations, MITE-like structures could become intervening sequences excised

by an unknown (transposition-related?) mechanism from pre-mRNA.

Conclusions and Perspectives

The analysis of *tubA* and *tubB* gene sequences from photosynthetic euglenids reveals that nonconventional intron gains were very common over their evolutionary history, in contrast to the behavior of conventional introns that seem only to have been lost in some lineages. It also suggests that the origin of nonconventional introns must be different than the origin of conventional ones. Insertion of MITE-like elements is proposed as a possible source of nonconventional introns. It seems likely that nonconventional introns can also arise as a consequence of mutational changes in conventional intron sequences—an intron with signals recognized by two types of splicing mechanism would be resistant to mutations affecting only one of the signals. However, an alternative scenario is also possible—conventional introns could arise as a consequence of mutations creating canonical GT-AG splice sites at the ends of nonconventional introns. The so-called intermediate introns would play an important role in both scenarios.

We hope that our highlighting of the common features of nonconventional introns will aid future studies of the mechanism of their removal and deeper analyses of euglenid genomes. Such analyses will be especially helpful in verifying the hypothesis presented here and establishing the moment in euglenids evolution when nonconventional introns appeared.

Materials and Methods

Strains and Culture Conditions

Euglenid strains whose sequences were used in this study are described in supplementary table S3, Supplementary Material online. All strains were cultivated under identical conditions in a liquid soil–water medium enriched with a small piece of garden pea (medium 3c, Schlösser 1994), in a growth chamber maintained at 17 °C and a 16:8 h light/dark cycle, ca. 27 μmol photons·m^{-2}·s^{-1} provided by cool white fluorescent tubes (Philips).

RNA Isolation, cDNA Synthesis, Amplification, Cloning, and Sequencing

Total genomic RNA was isolated with the RNeasy Kit (Qiagen) using the animal tissues protocol (incubation with proteinase K instead of mechanical disruption was performed). RNA was treated with DNase I (Qiagen) to eliminate DNA contamination according to the manufacturer's instructions. First-strand cDNA was synthesized from total RNA using SuperScript III Reverse Transcriptase (Invitrogen) and a 17 bp oligo-dT primer. Products of the synthesis were used as a template in a PCR reaction with the pair of degenerated primers F0/R0 (sequences of all primers used in this study are listed in supplementary table S4, Supplementary Material online). A 25 μl reaction mixture contained 0.5 U of Phusion High-Fidelity DNA Polymerase (Finnzymes), 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 10 pmol each primer, reaction buffer HR (Finnzymes), and 1 μl of the first strand cDNA (diluted 10×). The PCR protocol consisted of an initial 30 s at 98 °C, followed by 35 cycles comprising 10 s at 98 °C, 15 s at 54 °C (tubAF0/R0) or 53 °C (tubBF0/R0), and 30 s at 72 °C. The final extension step was performed for 5 min at 72 °C. PCR products were sized on agarose gels, purified using the QIAEXII Gel Extraction Kit (Qiagen), sequenced directly from both strands by cycle sequencing using BigDye Terminator Cycle Sequencing Ready Reaction Kit 3.1 (Life Technologies), and then analyzed on an ABI 3730 Genetic Analyser (Applied Biosystems). In some cases, PCR products were cloned into pGEM-T easy vector (Promega) after addition of adenosine at their 3′ ends using Taq Polymerase (Qiagen). For each euglenid strain, several clones were chosen and plasmids were isolated; then cycle sequencing of inserts was performed using BigDye Terminator Cycle Sequencing Ready Reaction Kit 3.1 (Life Technologies); universal vector primers M13F and M13R were used in the sequencing.

DNA Isolation, Amplification, Cloning, and Sequencing

Total genomic DNA was isolated from cultures with the DNeasy Tissue Kit (Qiagen) using the animal tissues protocol. DNA was treated with RNase A (Qiagen) to eliminate RNA contamination according to the manufacturer's instructions. Two-step nested amplification was performed to obtain a sufficient quantity of genomic PCR products. In the first reaction, the external degenerated primer pair F0/R0 was used, followed by amplification with the internal degenerated primer pair F1/R1. For *tubA* amplification from *M. pyrum* and *T. similis*, specific primers F0/R0 and F1/R1 were designed based on the cDNA sequences obtained earlier; for *tubB* amplification from *E. gymnodinioides*, specific primers F1/R1 were designed (see supplementary table S4, Supplementary Material online). In the first step, 25 μl reaction mixture contained 0.5 U Phusion High-Fidelity DNA Polymerase (Finnzymes), 0.2 mM dNTPs, 3.5 mM $MgCl_2$, 10 pmol of each primer, reaction buffer GC (Finnzymes), Q-solution (Qiagen), 1.2 μg of Taq Single-Stranded DNA Binding Protein (EURx), and 10–50 ng of DNA. We found that addition of the Taq SSB protein was crucial for reaction specificity and yield. The PCR protocol consisted of 2 min at 98 °C, followed by 7 initial cycles comprising 30 s at 98 °C, 30 s at 54 °C (tubAF0/R0) or 53 °C (tubBF0/R0), and 2.5 min at 72 °C, then by 35 cycles comprising 15 s at 98 °C, 15 s at 54/53 °C, and 2.5 min at 72 °C. The final extension step was performed for 5 min at 72 °C. In the second step, 25 μl reaction mixture contained 0.5 U of Phusion High-Fidelity DNA Polymerase (Finnzymes), 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 10 pmol of each primer, reaction buffer GC (Finnzymes), Q-solution (Qiagen), 0.6 μg of Taq Single-Stranded DNA Binding Protein (EURx), and 1 μl of undiluted mixture from the first step. The PCR protocol consisted of an initial 2 min at 98 °C, followed by 35 cycles comprising 15 s at 98 °C, 15 s at 54 °C (tubAF1/R1) or 63 °C (tubBF1/R1), and 2.5 min at 72 °C. PCR

products were sized on agarose gels, purified using the QIAEXII Gel Extraction Kit (Qiagen), and cloned into pGEM-T easy vector (Promega) after addition of adenosine at their 3′ ends using Taq Polymerase (Qiagen). For each euglenid strain, several clones were chosen and plasmids were isolated; then cycle sequencing of inserts was performed using BigDye Terminator Cycle Sequencing Ready Reaction Kit 3.1 (Life Technologies); universal vector primers M13F and M13R were used in the sequencing, as well as internal degenerated primers F2, R2, F3, and R3 corresponding to regions in the middle of *tubA* and *tubB* genes. In the case of *M. pyrum tubA* and *E. gymnodinioides* and *T. similis tubB*, additional internal sequencing primers were needed (see supplementary table S4, Supplementary Material online). Products of the sequencing reactions were analyzed on an ABI 3730 Genetic Analyser (Applied Biosystems).

Sequence Analysis

Sequences were assembled into contigs by the SeqMan program from the Lasergene package (DnaStar). To minimize the risk to include in the analysis inactive forms of genes, the genomic and cDNA sequences (direct sequencing of PCR products) were compared. When the genomic sequence matched perfectly to the cDNA sequence, it was taken into account in further analysis. In cases where more than one form of the gene was obtained, if at least one of them matched perfectly to the cDNA, they were also included in the analysis. In a few cases, where the mismatch was found between genomic and cDNA sequences, the cDNA-derived PCR products were cloned, and then several cloned products were sequenced. If at least one of genomic forms matched perfectly to the one of cloned cDNA sequences, such forms were included in the analysis. The forms of genes, which did not match perfectly to the cDNA, were retained in the analysis because they did not differ in the distribution of introns from forms whose expression was confirmed. To determine intron positions, the genomic and cDNA sequences were compared using the Mesquite program (Maddison WP and Maddison DR 2011). Prediction of the RNA secondary structure of introns was performed using the RNAfold WebServer (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi, last accessed December 5, 2013). Nucleotide logos of nonconventional intron junctions were created using Weblogo 2.8.2 (Crooks and Hon 2004; http://weblogo.berkeley.edu/logo.cgi, last accessed December 5, 2013).

Phylogenetic Analysis

After removing introns, *tubA* and *tubB* gene sequences were aligned separately using the Mesquite program (Maddison WP and Maddison DR 2011); to the set of sequences obtained in this study, previously published genes from *E. gracilis* and *Ent. sulcatum* were added. Maximum-likelihood analysis was performed with PhyML 3.0 (Guindon et al. 2010) using models for sequence evolution and their parameters estimated by JModeltest 2.1 (Posada 2008); for details, see legend to supplementary figure S2, Supplementary Material online.

Supplementary Material

Supplementary figures S1 and S2 and tables S1–S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Abiba Boulahdjel for her help with cloning and technical support. A.K. acknowledges support from the European Social Fund and the state budget of the Czech Republic (Project no. CZ.1.07/2.3.00/30.0061). This work was supported by grant 2011/01/B/NZ8/01658 from the National Science Centre, Poland.

References

Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, et al. 2012. The revised classification of eukaryotes. *J Eukaryot Microbiol.* 59:429–493.

Bennett MS, Wiegert KE, Triemer RE. 2012. Comparative chloroplast genomics between *Euglena viridis* and *Euglena gracilis* (Euglenophyta). *Phycologia* 51:711–718.

Breckenridge DG, Watanabe Y-I, Greenwood SJ, Gray MW, Schnare MN. 1999. U1 small nuclear RNA and spliceosomal introns in *Euglena gracilis*. *Proc Natl Acad Sci U S A.* 96:852–856.

Breglia SA, Slamovits CH, Leander BS. 2007. Phylogeny of phagotrophic euglenids (Euglenozoa) as inferred from hsp90 gene sequences. *J Eukaryot Microbiol.* 54:86–92.

Busse I, Patterson DJ, Preisfeld A. 2003. Phylogeny of phagotrophic euglenids (Euglenozoa): a molecular approach based on culture material and environmental samples. *J Phycol.* 39:828–836.

Calvin K, Li H. 2008. RNA-splicing endonuclease structure and function. *Cell Mol Life Sci.* 65:1176–1185.

Canaday J, Tessier LH, Imbault P, Paulus F. 2001. Analysis of *Euglena gracilis* alpha-, beta- and gamma-tubulin genes: introns and pre-mRNA maturation. *Mol Genet Genomics.* 265:153–160.

Casacuberta JM, Santiago N. 2003. Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. *Gene* 311:1–11.

Cavalier-Smith T. 2010. Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biol Lett.* 6:342–345.

Chen Y, Stephan W. 2003. Compensatory evolution of a precursor messenger RNA secondary structure in the *Drosophila melanogaster* Adh gene. *Proc Natl Acad Sci U S A.* 100:11499–11504.

Cook JR, Roxby R. 1985. Physical properties of a plasmid-like DNA from *Euglena gracilis*. *Biochim Biophys Acta.* 824:80–83.

Crooks G, Hon G. 2004. WebLogo: a sequence logo generator. *Genome Res.* 14:1188–1190.

Dooijes D, Chaves I, Kieft R, Dirks-Mulder A, Martin W, Borst P. 2000. Base J originally found in kinetoplastida is also a minor constituent of nuclear DNA of *Euglena gracilis*. *Nucleic Acids Res.* 28:3017–3021.

Ebel C, Frantz C, Paulus F, Imbault P. 1999. Trans-splicing and cis-splicing in the colourless Euglenoid, *Entosiphon sulcatum*. *Curr Genet.* 35: 542–550.

Feschotte C, Zhang X, Wessler SR. 2002. Miniature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington (DC): ASM Press. p. 1147–1158.

Giroux MJ, Clancy M, Baier J, Ingham L, McCarty D, Hannah LC. 1994. De novo synthesis of an intron by the maize transposable element dissociation. *Proc Natl Acad Sci U S A.* 91:12150–12154.

Gockel G, Hachtel W. 2000. Complete gene map of the plastid genome of the nonphotosynthetic Euglenoid flagellate *Astasia longa*. *Protist* 151:347–351.

Gonzalez TN, Sidrauski C, Dörfler S, Walter P. 1999. Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway. *EMBO J.* 18:3119–3132.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.

Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A, Stutz E. 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res.* 21:3537–3544.

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger AJ. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proc Natl Acad Sci U S A.* 106:3859–3864.

Henze K, Badr A, Wettern M, Cerff R, Martin W. 1995. A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. *Proc Natl Acad Sci U S A.* 92:9122–9126.

Hrdá Š, Fousek J, Szabová J, Hampl V, Vlček Č. 2012. The plastid genome of *Eutreptiella* provides a window into the process of secondary endosymbiosis of plastid in euglenids. *PLoS One* 7:e33746.

Jiang N, Feschotte C, Zhang X, Wessler SR. 2004. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Curr Opin Plant Biol.* 7:115–119.

Lane CE, Van den Heuvel K, Kozera C, Curtis BA, Parsons BJ, Bowman S, Archibald JM. 2007. Nucleomorph genome of *Hemiselmis andersenii* reveals complete intron loss and compaction as a driver of protein structure and function. *Proc Natl Acad Sci U S A.* 104:19908–19913.

Linton EW, Hittner D, Lewandowski C, Auld T, Triemer RE. 1999. A molecular study of euglenoid phylogeny using small subunit rDNA. *J Eukaryot Microbiol.* 46:217–223.

Linton EW, Karnkowska-Ishikawa A, Kim JI, Shin W, Bennett MS, Kwiatowski J, Zakryś B, Triemer RE. 2010. Reconstructing euglenoid evolutionary relationships using three genes: nuclear SSU and LSU, and chloroplast SSU rDNA sequences and the description of *Euglenaria* gen. nov. (Euglenophyta). *Protist* 161:603–619.

Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H. 2012. Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. *Mol Biol Evol.* 29: 1005–1017.

Maddison WP, Maddison DR. 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75 [Internet]. [cited 2013 Dec 5]. Available from: http://mesquiteproject.org.

Mair G, Shi H, Li H, Djikeng A, Aviles HO, Bishop JR, Falcone FH, Gavrilescu C, Montgomery JL, Santori MI, et al. 2000. A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *RNA* 6: 163–169.

Marin B, Palm A, Klingberg M, Melkonian M. 2003. Phylogeny and taxonomic revision of plastid-containing euglenophytes based on SSU rDNA sequence comparisons and synapomorphic signatures in the SSU rRNA secondary structure. *Protist* 154:99–145.

Milanowski R, Kosmala S, Zakryś B, Kwiatowski J. 2006. Phylogeny of photosynthetic euglenophytes based on combined chloroplast and cytoplasmic SSU rDNA sequence analysis. *J Phycol.* 42:721–730.

Muchhal US, Schwartzbach SD. 1994. Characterization of the unique intron - exon junctions of *Euglena* gene(s) encoding the polyprotein precursor to the light-harvesting chlorophyll a/b binding protein of photosystem II. *Nucleic Acids Res.* 22:5737–5744.

Nagashima Y, Mishiba K, Suzuki E, Shimada Y, Iwata Y, Koizumi N. 2011. Arabidopsis IRE1 catalyses unconventional splicing of bZIP60 mRNA to produce the active transcription factor. *Sci Rep.* 1:29.

Perumal BS, Sakharkar KR, Chow VT, Pandjassarame K, Sakharkar MK. 2005. Intron position conservation across eukaryotic lineages in tubulin genes. *Front Biosci.* 10:2412–2419.

Pombert J-F, James ER, Janouškovec J, Keeling PJ. 2012. Evidence for transitional stages in the evolution of Euglenid group II introns and twintrons in the *Monomorphina aenigmatica* plastid genome. *PLoS One* 7:e53433.

Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* 25:1253–1256.

Ravel-Chapuis P. 1988. Nuclear rDNA in *Euglena gracilis*: paucity of chromosomal units and replication of extrachromosomal units. *Nucleic Acids Res.* 16:4801–4810.

Rogozin IB, Carmel L, Csuros M, Koonin EV. 2012. Origin and evolution of spliceosomal introns. *Biol Direct.* 7:11.

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

Roy J, Faktorová D, Lukes J, Burger G. 2007. Unusual mitochondrial genome structures throughout the Euglenozoa. *Protist* 158: 385–396.

Roy SW, Hudson AJ, Joseph J, Yee J, Russell AG. 2012. Numerous fragmented spliceosomal introns, AT-AC splicing, and an unusual dynein gene expression pathway in *Giardia lamblia*. *Mol Biol Evol.* 29:43–49.

Russell AG, Watanabe Y, Charette JM, Gray MW. 2005. Unusual features of fibrillarin cDNA and gene structure in *Euglena gracilis*: evolutionary conservation of core proteins and structural predictions for methylation-guide box C/D snoRNPs throughout the domain Eucarya. *Nucleic Acids Res.* 33:2781–2791.

Schlösser UG. 1994. SAG - Sammlung von Algenkulturen at the University of Göttingen. Cataloge of strains 1994. *Botanica Acta* 107:113–187.

Schnare MN, Gray MW. 1990. Sixteen discrete RNA components in the cytoplasmic ribosome of *Euglena gracilis*. *J Mol Biol.* 215:73–83.

Schneider A, Martin J, Agabian N. 1994. A nuclear encoded tRNA of *Trypanosoma brucei* is imported into mitochondria. *Mol Cell Biol.* 14: 2317–2322.

Tessier LH, Paulus F, Keller M, Vial C, Imbault P. 1995. Structure and expression of *Euglena gracilis* nuclear rbcS genes encoding the small subunits of the ribulose 1,5-bisphoshate carboxylase/oxygenase: a novel splicing process for unusual intervening sequences? *J Mol Biol.* 245:22–33.

Triemer RE, Linton E, Shin W, Nudelman A, Monfils A, Bennett M, Brosnan S. 2006. Phylogeny of the Euglenales based upon combined SSU and LSU rDNA sequence comparisons and description of *Discoplastis* gen. nov. (Euglenophyta). *J Phycol.* 42:731–740.

Turmel M, Gagnon M-C, O'Kelly CJ, Otis C, Lemieux C. 2009. The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Mol Biol Evol.* 26:631–648.

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

Vesteg M, Sándorová Z, Krajčovič J. 2012. Selective forces for the origin of spliceosomes. *J Mol Evol.* 74:226–231.

Vesteg M, Vacula R, Steiner JM, Mateásiková B, Löffelhardt W, Brejová B, Krajcovic J. 2010. A possible role for short introns in the acquisition of stroma-targeting peptides in the flagellate *Euglena gracilis*. *DNA Res.* 17:223–231.

Wessler SR. 1998. Transposable elements associated with normal plant genes. *Physiol Plant.* 103:581–586.

Wiegert KE, Bennett MS, Triemer RE. 2012. Evolution of the chloroplast genome in photosynthetic euglenoids: a comparison of *Eutreptia viridis* and *Euglena gracilis* (Euglenophyta). *Protist* 163:832–843.

Wiegert KE, Bennett MS, Triemer RE. 2013. Tracing patterns of chloroplast evolution in euglenoids: contributions from *Colacium vesiculosum* and *Strombomonas acuminata* (Euglenophyta). *J Eukaryot Microbiol.* 60:214–221.

Yamaguchi A, Yubuki N, Leander BS. 2012. Morphostasis in a novel eukaryote illuminates the evolutionary transition from phagotrophy to phototrophy: description of *Rapaza viridis* n. gen. et sp. (Euglenozoa, Euglenida). *BMC Evol Biol.* 12:29.

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

Yoshida H, Matsui T, Yamamoto A, Okada T, Mori K. 2001. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell* 107:881–891.