RESEARCH PAPER

Mobile Genetic Elements 4, e29384; January–December 2014; © 2014 Landes Bioscience

An intronic open reading frame was released from one of group II introns in the mitochondrial genome of the haptophyte *Chrysochromulina* sp. NIES-1333

Yuki Nishimura[1,2], Ryoma Kamikawa[3,4], Tetsuo Hashimoto[1,5], and Yuji Inagaki[1,5,*]

[1]Graduate School of Life and Environmental Sciences; University of Tsukuba; Tsukuba, Japan; [2]Graduate School of Systems and Information Engineering; University of Tsukuba; Tsukuba, Japan; [3]Graduate School of Human and Environmental Studies; Kyoto University; Kyoto, Japan; [4]Graduate School of Global Environmental Studies; Kyoto University; Kyoto, Japan; [5]Center for Computational Sciences; University of Tsukuba; Tsukuba, Japan

Keywords: group II intron, mitochondrial genome, intron encoded protein, lateral intron transfer, *Chrysochromulina* sp.

Abbreviations: mt; mitochondria, gII intron; group II intron, IEP; intron encoded protein

Mitochondrial (mt) genome sequences, which often bear introns, have been sampled from phylogenetically diverse eukaryotes. Thus, we can anticipate novel insights into intron evolution from previously unstudied mt genomes. We here investigated the origins and evolution of three introns in the mt genome of the haptophyte *Chrysochromulina* sp. NIES-1333, which was sequenced completely in this study. All the three introns were characterized as group II, on the basis of predicted secondary structure, and the conserved sequence motifs at the 5′ and 3′ termini. Our comparative studies on diverse mt genomes prompt us to propose that the *Chrysochromulina* mt genome laterally acquired the introns from mt genomes in distantly related eukaryotes. Many group II introns harbor intronic open reading frames for the proteins (intron-encoded proteins or IEPs), which likely facilitate the splicing of their host introns. However, we propose that a "free-standing," IEP-like protein, which is not encoded within any introns in the *Chrysochromulina* mt genome, is involved in the splicing of the first *cox1* intron that lacks any open reading frames.

Introduction

Mitochondrial (mt) genomes can be regarded as a model system for studying intron evolution, as a massive amount of mt genome data including group I and/or group II (gII) introns have been accumulated (for instance, see NCBI Organelle Genome Resource: http://www.ncbi.nlm.nih.gov/genomes/OrganelleResource.cgi?opt=organelle&taxid=2759). Group II introns are found in the genomes of prokaryotes (bacteria and archaebacteria), as well as mitochondria and plastids,[1,2] which are derived from an α-proteobacterium[3] and a cyanobacterium,[4] respectively. So far, gII introns have been identified in mt genomes from members of phylogenetically diverse eukaryotic assemblages such as Metazoa,[5,6] Jakobida,[7,8] Archaeplastida,[9-11] Fungi,[12,13] Cryptophyta,[14] Haptophyta,[15,16] and Stramenopiles.[17-19] These gII introns possess features at both the primary and secondary structure levels. At the primary structure level, gII introns possess highly conserved sequence motifs at the 5′ and 3′ ends (i.e., 5′-GTGYG…AY-3′; Y for T or C).[20] At the secondary structure level, we anticipate the transcripts of typical gII introns (intron RNAs) to form a characteristic bulge structure with six stems, so-called domains I to VI.[21] Both primary and secondary structures of intron RNAs are most likely critical for the splicing reaction.[22]

Group II introns can be regarded as mobile genetic elements, which are transmittable between an intron-containing and an intron-lacking loci (intron homing), regardless of their evolutionary distance. The mobility of gII introns are most likely conferred by the proteins encoded within gII introns (intron-encoded proteins or IEPs). Typical IEPs comprise three functionally distinct domains, namely i) reverse transcriptases (RT), ii) domain X, which is also referred to as maturases, and iii) endonucleases (En),[23] although some IEPs were reported to lack En domain.[20,22] Among the three domains in IEP, RT and En domains are predicted to catalyze reverse transcription of intron RNA and digest the target (intron-lacking) locus, respectively.[20,23] Domain X may not be responsible for intron mobility, but assists splicing by stabilizing the conformation of intron RNA. Nevertheless, we have known of many "IEP-free" gII introns, and it is difficult to predict the protein factors, which cooperate with a particular IEP-free intron in *trans*. To our knowledge, there is only a single report that successfully identified the organellar

*Correspondence to: Yuji Inagaki; Email: yuji@ccs.tsukuba.ac.jp
Submitted: 04/21/2014; Revised: 05/27/2014; Accepted: 05/27/2014; Published Online: 05/27/2014
Citation: Nishimura Y, Kamikawa R, Hashimoto T, Inagaki Y. An intronic open reading frame was released from one of group II introns in the mitochondrial genome of a haptophyte Chrysochromulina sp. NIES-1333. Mobile Genetic Elements 2014; 4:e29384; http://dx.doi.org/10.4161/mge.29384

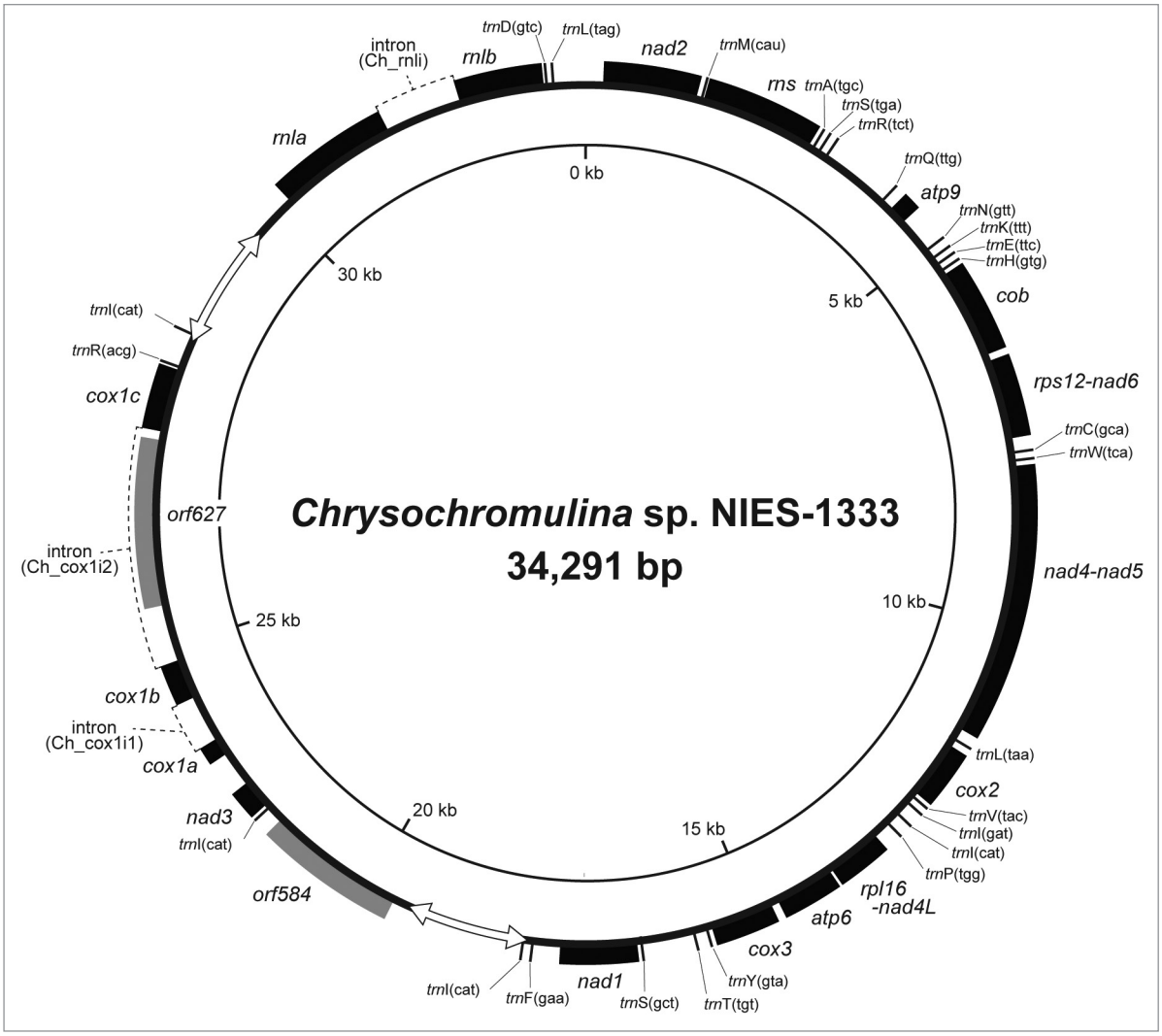

Figure 1. Mitochondrial Genome of *Chrysochromulina* sp. NIES-1333. Protein-coding genes and rRNA genes are represented by boxes. Gray boxes represent two open reading frames, of which amino acid sequences showed significant sequence similarity to intron-encoded proteins. Transfer RNA genes are represented by lines. Introns are shown in dotted lines. Arrows represent duplicated regions.

genome-encoded *trans* factor involved in the splicing of a IEP-free intron.[24]

In this study, we completely sequenced the mt genome of a haptophyte *Chrysochromulina* sp. NIES-1333, and identified three introns in total, two of those are found in the *cox1* gene encoding cytochrome *c* oxidase subunit 1, and the last one is found in the *rnl* gene encoding large subunit rRNA. Analyses of the intron sequences suggest that the three introns in the *Chrysochromulina* mt genome belonged to group II. We identified two open reading frames encoding putative IEPs. Both showed significant sequence similarity to gII intron-hosted IEPs in the mt genome; one is *orf627* encoded in the second *cox1* intron, and the other is *orf584*, which is free-standing. Phylogenetic analyses of IEPs and comparisons of intron positions across phylogenetically diverse mt genomes revealed that the *Chrysochromulina* mt genome shares homologous introns with distantly related mt genomes.

Results and Discussion

Overview of the *Chrysochromulina* mt genome

The mt genome of *Chrysochromulina* sp. NIES-1333 was assembled into a circular molecule of 34,291 bp in length with an A + T content of 70.0% (**Fig. 1**). We identified 16 functionally assignable open reading frames (including those for two IEPs; see below). UGA codons are most likely assigned for tryptophan instead of termination signal (**Table S1**), as reported previously.[25] This type of mt genetic code was reported in other members of Prymnesiophyceae, one of the two classes of Haptophyta.[26-28] We detected 26 tRNA genes and a set of small and large subunits of rRNAs; No 5S rRNA gene was identified. A set of tRNA genes identified in the mt genome is sufficient to translate all amino acid codons except for GGN (N = A, C, G or U) codons for glycine (**Table S1**). All genes mentioned above were encoded on a single strand. A region with an approximate length of 1.6 Kbp,

which contained a single tRNA gene for isoleucine, was found to be duplicated (shown by open arrows in **Fig. 1**).

In terms of gene repertoire, the *Chrysochromulina* mt genome is fundamentally similar to those of other haptophytes, namely *Emiliania huxleyi*,[29] *Diacronema lutheri* (www.bch.umontreal.ca/ogmp/projects/pluth/gen.html), *Phaeocystis* spp.,[12] as shown in **Table 1**.

General features of the intron in the *Chrysochromulina rnl* gene

The *rnl* gene hosts a single intron encoding no apparent open reading frame (designated as Ch_rnli; **Fig. 1**). The intron was found to be inserted at the position between the 837th and 838th bases in the *Homo sapiens* homolog (GeneID: 4550 in NC_012920). Ch_rnli starts with 5′-GTGCG… and ends with …CT-3′, which is similar to the consensus motifs shared among typical gII introns (5′-GTGYG…AY-3′; Y for T or C). Although the intron sequence was too divergent to predict the entire secondary structure, we successfully identified the typical domains, V and VI, which are characteristic secondary structures of gII introns (**Fig. S1A**), with the aid of MFannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl). Thus, we characterized Ch_rnli as group II.

The homing position of Ch_rnli was found to be identical to those of *rnl* introns found in a member of Archaeplastida (the red alga, *Pyropia haitanensis*; NC_017751), two members of Stramenopiles (the brown alga, *Pylaiella littoralis*, and the diatom, *Phaeodactylum tricornutum*; NC_003055 and HQ840789, respectively), and a member of Opisthokonta (the fungus, *Gigaspora rosea*; NC_016985) (**Fig. S1A**: Note that none of the *rnl* introns found in the mt genomes of other haptophytes, *D. lutheri* and *Phaeocystis globosa*, shared the homing position with Ch_rnli). We predicted the secondary structures of domains V and VI in Ch_rnli and the four introns described above, but detected no apparent homology at the nucleotide sequence level among them (**Fig. S1B**). Furthermore, Ch_rnli hosts no IEP, which is considered a key aspect to inferring intron evolution.[22] Thus, we hesitate to discuss the evolutionary relationship among Ch_rnli and the introns listed above, solely based on their homing positions.

General features of two introns in the *Chrysochromulina cox1* gene

Two introns were found in the *Chrysochromulina cox1* gene (**Fig. 1**). We designate the first and second introns in the *cox1* gene as Ch_cox1i1 and Ch_cox1i2, respectively. The two introns commonly start with 5′-GTGCG… and end with …AC-3′, being consistent with the consensus motifs shared among typical gII introns (5′-GTGYG…AY-3′). Ch_cox1i1 was inserted between the second and third letters (phase-2) of the codon corresponding to Phe[68] in the *H. sapiens cox1* gene, sharing the homing position with the gII introns in *cox1* genes of the cryptophyte, *Rhodomonas salina*, and the diatom, *Phaeodactylum tricornatum* (**Fig. S2C**). Ch_cox1i2 was found at phase-2 of the codon corresponding to Phe[237] in the *H. sapiens cox1* gene, being homologous to those of the gII introns in *cox1* genes of the haptophyte *D. lutheri* and the diatom *Ulnaria acus* (**Fig. S2C**). Ch_cox1i2 hosts an intronic open reading frame for an IEP, while Ch_cox1i1 encodes no apparent open reading frame. Both Ch_cox1i1 and Ch_cox1i2 can be folded into the characteristic secondary structures shared among gII introns, albeit with some ambiguity remaining in domain I (indicated as "DI" in **Fig. S1B and S1C**). All together, we concluded that the two introns belong to group II.

Evolution of Ch_cox1i2 and its IEP

The IEP encoded in the intronic open reading frame of Ch_cox1i2, ORF627, most likely facilitates splicing of the host intron. The ORF627 amino acid sequence showed similarity to other gII intron-hosted IEP sequences deposited in the GenBank database; the top blastp hit was an IEP encoded in the first gII intron of the *cox1* gene in the haptophyte *D. lutheri* (Dl_cox1i), with a 49% sequence similarity and an *E*-value of 0.0. In both maximum-likelihood (ML) and Bayesian analyses of an IEP alignment (**Fig. 2**), *Chrysochromulina* ORF627 formed a clade with two IEPs encoded in *cox1* gII introns, namely Dl_cox1i and that of the diatom *U. acus* (Ua_cox1i) with a ML bootstrap support value (MLBP) of 96% and a Bayesian posterior probability (BPP) of 1.00 (**Fig. 2**). As we generally believe that gII introns and their IEPs have coevolved,[22] the intimate relationship among the IEPs encoded in Ch_cox1i2, Dl_cox1i, and Ua_cox1i suggests that their host introns are derived from a single ancestral intron bearing an IEP. The single origin of Ch_cox1i2, Dl_cox1i, and Ua_cox1i discussed above is consistent with the fact that the three introns share a homing position (**Fig. S2C**).

The ancestral haptophyte species likely possessed a *cox1* gene with a particular gII intron, as *Chrysochromulina* sp. and *D. lutheri* are representatives of two major classes, Prymnesiophyceae and Pavlovaphyceae, in Haptophyta, respectively. This scenario suggests that multiple intron losses occurred in the *cox1* genes of *Emiliania huxleyi*,[28,29] members of the genus *Phaeocystis*,[12,26] and *Isochrysis galbana*.[30]

The IEP phylogeny and comparison of homing positions imply that the homologous introns are present in two distantly related branches (haptophytes and diatoms) in the tree of eukaryotes. This sporadic intron distribution can be explained by a scenario incorporating lateral intron transfer. There is an alternative, but less plausible scenario assuming that the *cox1* gene in an ancestral organism, which has existed prior to the divergence of major eukaryotic assemblages including diatoms and haptophytes, may have already possessed a gII intron at phase-2 of the codon corresponding to Phe[237] in the *H. sapiens cox1* gene, and would have been (secondarily) lost in multiple descendants (i.e., ancestral co-occurrence followed by multiple secondary losses). We prefer the scenario incorporating lateral intron transfer to the alternative one, but these scenarios should be reexamined by future studies based on a broader diversity of gII introns (and their IEPs) compared with those considered in the current study.

Link between a free-standing *orf584* and an IEP-free Ch_cox1i1

Most IEPs are encoded in intronic open reading frames (as observed in Ch_cox1i2; see above), but a few of those are free-standing in genomes (e.g., *orf732* in the *Marchantia polymorpha* mt genome; highlighted by a star in **Fig. 2**). Our blast search showed that *Chrysochromulina* ORF584, which is encoded in a

Table 1. Gene repertoires in haptophyte mitochondrial genomes

	Chrysochromulina sp.	*Diacronema lutheri*	*Emiliania huxleyi*	*Phaeocystis globosa*	*Phaeocystis antarica*
rnl	Y [1]	Y [1]	Y	Y [1]	Y
rns	Y	Y [1]	Y	Y	Y
rrn5	N	Y	Y	N	N
trn	23 species	22 species	23 species	23 species	23 species
nad1	Y	Y	Y	Y	Y
nad2	N	N	N	Y	Y
nad3	N	N	N	Y	Y
nad4	Y	Y	Y	Y	Y
nad4L	Y	Y	Y	Y	Y
nad5	Y	Y	Y	Y	Y
nad6	Y	Y	Y	Y	Y
cob	Y	Y	Y	Y	Y
cox1	Y [2]	Y [1]	Y	Y	Y
cox2	Y	Y	Y	Y	Y
cox3	Y	Y [1]	Y	Y	Y
atp4	N	Y	Y	Y	Y
atp6	Y	Y [1]	Y	Y	Y
atp8	N	Y	N	Y	Y
atp9	Y	Y [1]	Y	Y	Y
rps3	N	N	Y	Y	Y
rps8	N	N	Y	N	N
rps12	Y	Y	Y	Y	Y
rps14	N	Y	Y	N	Y
rps19	N	Y	N	N	N
rpl14	N	Y	N	N	N
rpl16	Y	Y	Y	Y	Y
dam	N	N	Y	N	N
Others	*orf627*[a] *orf538*[b]	*orf636*[c] *orf105*[d]	*orf104*[d]	N	N

Y, yes; N, no. Numbers of introns are shown in brackets. [a]Encoded in the second *cox1* intron. [b]Free-standing open reading frame encoding a protein with amino acid sequence similarity to group II intron-encoded proteins. [c]Encoded in the *cox1* intron. [d]Encodes an uncharacterized protein.

free-standing open reading frame, bore a significant sequence similarity to gII intron-hosted IEPs; The top blastp hit of ORF584 amino acid sequence was an IEP (ORF724) encoded in the first intron of the *cox1* gene in the diatom, *P. tricornutum* (Pt_cox1i1) with a 53% sequence similarity and an *E*-value of 0.0. ORF584 equips En, RT, and domain X, implying that this protein assists intron splicing. The phylogenetic analyses of the IEP alignment (**Fig. 2**) recovered a robust affinity between *Chrysochromulina* ORF584 and ORF724 encoded in Pt_cox1i1 with a MLBP of 100% and a BPP of 1.00 (**Fig. 2**). This indicates the two proteins were derived from the single ancestral IEP encoded in a gII intron, which is homologous to Pt_cox1i1, the host intron of ORF724. Curiously, Pt_cox1i1 and Ch_cox1i1 appeared to share a homing position (see **Fig. S2C**). We also noticed that the nucleotide sequence of domain VI in Ch_cox1i1 and that in Pt_cox1i1 are similar to one another (**Fig. S1D**), although this domain sequences are generally variable among gII introns.[21] The homing position and sequence similarity in domain VI, between Ch_cox1i1 and Pt_cox1i1, suggests that the two introns are homologous to each other. All together, we here propose that ORF584 used to be encoded in Ch_cox1i1, and still assists the splicing of the host intron even after being free-standing secondarily in the current *Chrysochromulina* mt genome. To the best of our knowledge, co-relation between a particular pair of free-standing IEP and IEP-free introns has been reported only one time prior to this work.[24]

The first intron in the *R. salina cox1* gene (Rs_cox1i1) is unlikely to be homologous to Pt_cox1i1 or Ch_cox1i1, although

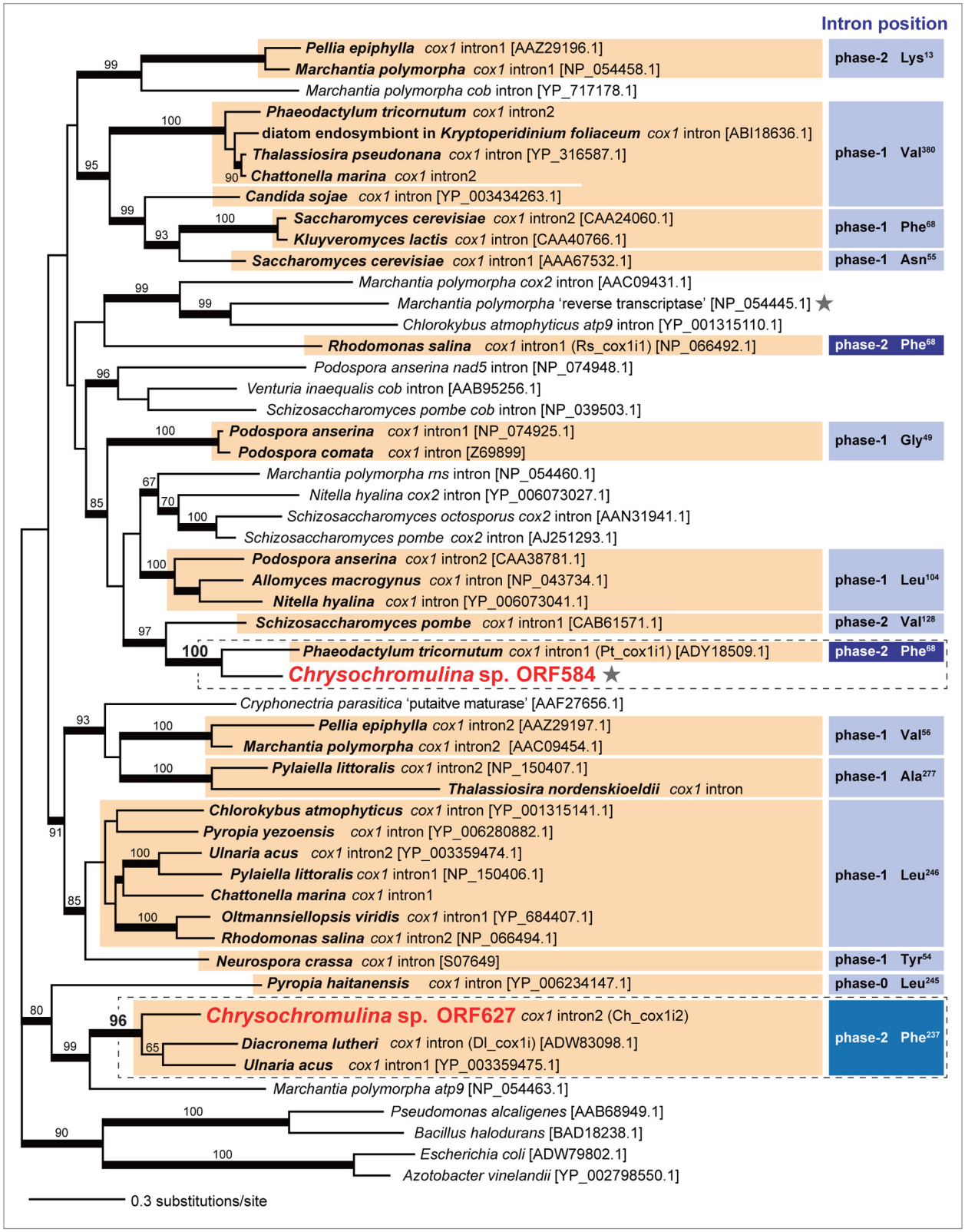

Figure 2. Phylogeny Inferred from 52 Intron-encoded Protein (IEP) Amino Acid Sequences. The IEP alignment was subjected to both maximum-likelihood (ML) and Bayesian methods. As the two methods reconstructed very similar trees, only ML tree is shown here. The tree is rooted by the bacterial sequences. Values at nodes represent ML bootstrap support values greater than 50%. The nodes supported by Bayesian posterior probabilities equal to or greater than 0.95 are highlighted by thick lines. The IEPs encoded in *cox1* introns are shaded in orange. The detailed homing positions of *cox1* introns are given on the right side of the tree. Codon numbers are based on the *Homo sapiens cox1* gene (GenBank accession number, YP_003024028). Free-standing IEPs are highlighted with stars.

the three introns share the homing position (**Fig. S1C**). The IEP phylogeny placed the IEP encoded in Rs_cox1i1 in a remote position from the clade of ORF584 and ORF724 (**Fig. 2**), strongly arguing against the homology between Rs_cox1i1 and Pt_cox1i1/Ch_cox1i1. The homology between Pt_cox1i1 and Ch_cox1i1, which were found in two phylogenetically distantly related species (i.e., a haptophyte and a diatom), can be explained by lateral intron transfer. Nonetheless, we cannot exclude the alternative possibility assuming ancestral co-occurrence followed by multiple secondary losses. We prefer the simplicity of the first scenario incorporating lateral intron transfer, but the alternative scenario should not be ignored before mt genome diversity is sufficiently covered.

Materials and Methods

Cell culture and mt genome sequencing

The haptophyte *Chrysochromulina* sp. NIES-1333 was purchased from the National Institute for Environmental Study (NIES; 16–2 Onogawa, Tsukuba, Ibaraki 305–8506, Japan). Haptophyte cells were grown in f/2 medium (http://mcc.nies.go.jp/02medium.html#f2) at 20 °C under 14 h light/10 h dark cycles. The cultured cells were harvested by centrifugation. Total DNA and total RNA were extracted from the harvested cells by CTAB buffer[31] and TRIzol (Invitrogen), respectively. Total RNA was used to synthesize cDNA with random hexamers and Superscript II reverse transcriptase (Invitrogen). RNA extraction and cDNA synthesis were conducted following manufacturers' protocols.

We amplified the entire mt genome by combination of LA PCR with TaKaRa *LA Taq* DNA polymerase (TaKaRa), genome walking with the GenomeWalker Universal kit (Clonthech), and rolling circle amplification (RCA) with the illustra TempliPhi 100 Amplification kit (GE Healthcare Life Sciences). Amplified DNA fragments of < 3 Kbp-long and those of < 10 Kbp-long were cloned to pGEM T-easy vector (Promega) and pCR-TOPO-XL (Invitrogen), respectively. The short amplicons (< 3 Kbp-long) were sequenced by the Sanger method. 454 pyro-sequencing by the GS-Jr system (454 Sequencing, Roche) was performed on the long amplicons (≥ 10 Kbp). Newbler (454 Sequencing, Roche) was used for de novo assembly of the pyro-sequencing reads. The DNA amplifications and sequencing described above were conducted by following manufactures' instructions, except the RCA with custom primers instead of random hexamers supplied in the kit, as described in Kamikawa et al. (2014).[32] The custom primers used for the RCA were designed based on the *cob* and *cox3* sequences determined previously.[25] All the DNA sequences obtained were finally assembled into a circular molecule, with an approximate length of 34 Kbp. The complete mt genome sequence was deposited to DDBJ/EMBL/GenBank accession number AB930144.

Genome analyses

Genes encoding proteins and rRNAs were identified by blastx and blastn searches,[33] respectively, against the non-redundant database in National Center for Biotechnology Information (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Transfer RNA-encoding genes were found by using tRNAscan-SE.[34] Independent from the analyses described above, we re-annotated the genome by MFannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl).

During the annotation of the *Chrysochromulina* mt genome, we noticed that both *cox1* and *rnl* genes are intervened by introns. The precise boundaries of the introns in the mt genome were confirmed by sequencing the corresponding transcripts (cDNAs). The secondary structures of the introns identified in the mt genome were predicted by MFOLD,[35] followed by manual refinement by referring to Toor et al. (2001)[21] and GOBASE database.[36]

Phylogenetic analyses

We found that *orf627* and *orf584* in the *Chrysochromulina* mt genome encode IEPs (see above). The conceptual amino acid sequences of the two IEPs were aligned with 46 IEPs encoded in other mt genomes and four bacterial homologs by Muscle.[37] The IEP sequences were retrieved from the GenBank database by referring to pioneering phylogenetic studies (e.g., refs. 17 and 18). After manual refinements and exclusion of ambiguously aligned positions, the final alignment, including 52 IEP sequences with 453 amino acid positions, was used for phylogenetic analyses.

The alignment was subjected to both ML and Bayesian methods using RAxML7.2.6[38] and PhyloBayes3.3,[39] respectively. We applied LG amino acid substitution model[40] with among-site rate variation approximated by a discrete gamma distribution with four rate categories to ML analyses (LG + Γ + F) of the original alignment and 100 bootstrap replicates. The ML tree was selected by heuristic searches from 10 randomized maximum-parsimony (MP) starting trees. In ML bootstrap analysis, heuristic tree search was performed from a single MP starting tree per replicate. Bayesian analysis was conducted with the LG + Γ+ F model. Two independent Monte Carlo chains were run for 5,800–5,850 cycles, reaching maxdiff value of 0.08353. The first 100 cycles were discarded as "burn-in"; the consensus tree, branch lengths, and BPPs was calculated from remaining trees.

Disclosure of Potential Conflicts of Interest

No conflict of interest to declare was disclosed.

Acknowledgments

Sequencing of the *Chrysochromulina* sp. mt genome was supported by the Cooperative Research Grant of the Gene Research Center, University of Tsukuba. We thank Dr. Kevin C. Wakeman (Okinawa Institute for Science and Technology, Japan) for English-language corrections. YN is supported by a research fellowship from the Japanese Society for the Promotion of Science (JSPS) for Young Scientists (No. 25789). This work was supported by the JSPS grants awarded to YI (No. 23117006) and TH (Nos. 23405013 and 23247038). RK was supported by a grant from the Institute for Fermentation, Osaka, Japan.

Supplemental Materials

Supplemental materials may be found here:
www.landesbioscience.com/journals/mge/article/29384

References

1. Michel F, Jacquier A, Dujon B. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. Biochimie 1982; 64:867-81; PMID:6817818; http://dx.doi.org/10.1016/S0300-9084(82)80349-0
2. Toro N. Bacteria and Archaea Group II introns: additional mobile genetic elements in the environment. Environ Microbiol 2003; 5:143-51; PMID:12588294; http://dx.doi.org/10.1046/j.1462-2920.2003.00398.x
3. Gray MW, Burger G, Lang BF. Mitochondrial evolution. Science 1999; 283:1476-81; PMID:10066161; http://dx.doi.org/10.1126/science.283.5407.1476
4. Gould SB, Waller RF, McFadden GI. Plastid evolution. Annu Rev Plant Biol 2008; 59:491-517; PMID:18315522; http://dx.doi.org/10.1146/annurev.arplant.59.032607.092915
5. Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW, Schierwater B. Mitochondrial genome of *Trichoplax adhaerens* supports placozoa as the basal lower metazoan phylum. Proc Natl Acad Sci U S A 2006; 103:8751-6; PMID:16731622; http://dx.doi.org/10.1073/pnas.0602076103
6. Vallès Y, Halanych KM, Boore JL. Group II introns break new boundaries: presence in a bilaterian's genome. PLoS One 2008; 3:e1488; PMID:18213396; http://dx.doi.org/10.1371/journal.pone.0001488
7. Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature 1997; 387:493-7; PMID:9168110; http://dx.doi.org/10.1038/387493a0
8. Burger G, Gray MW, Forget L, Lang BF. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biol Evol 2013; 5:418-38; PMID:23335123; http://dx.doi.org/10.1093/gbe/evt008
9. Bégu D, Araya A. The horsetail *Equisetum arvense* mitochondria share two group I introns with the liverwort *Marchantia*, acquired a novel group II intron but lost intron-encoded ORFs. Curr Genet 2009; 55:69-79; PMID:19112563; http://dx.doi.org/10.1007/s00294-008-0225-7
10. Turmel M, Otis C, Lemieux C. An unexpectedly large and loosely packed mitochondrial genome in the charophycean green alga *Chlorokybus atmophyticus*. BMC Genomics 2007; 8:137; PMID:17537252; http://dx.doi.org/10.1186/1471-2164-8-137
11. Mao Y, Zhang B, Kong F, Wang L. The complete mitochondrial genome of *Pyropia haitanensis* Chang et Zheng. Mitochondrial DNA 2012; 23:344-6; PMID:22803707; http://dx.doi.org/10.3109/19401736.2012.690751
12. Paquin B, Lang BF. The mitochondrial DNA of *Allomyces macrogynus*: the complete genomic sequence from an ancestral fungus. J Mol Biol 1996; 255:688-701; PMID:8636971; http://dx.doi.org/10.1006/jmbi.1996.0056
13. Foury F, Roganti T, Lecrenier N, Purnelle B. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae.* FEBS Lett 1998; 440:325-31; PMID:9872396; http://dx.doi.org/10.1016/S0014-5793(98)01467-7
14. Hauth AM, Maier UG, Lang BF, Burger G. The *Rhodomonas salina* mitochondrial genome: bacteria-like operons, compact gene arrangement and complex repeat region. Nucleic Acids Res 2005; 33:4433-42; PMID:16085754; http://dx.doi.org/10.1093/nar/gki757
15. Smith DR, Arrigo KR, Alderkamp AC, Allen AE. Massive difference in synonymous substitution rates among mitochondrial, plastid, and nuclear genes of *Phaeocystis* algae. Mol Phylogenet Evol 2014; 71:36-40; PMID:24216019; http://dx.doi.org/10.1016/j.ympev.2013.10.018
16. Ehara M, Watanabe KI, Ohama T. Distribution of cognates of group II introns detected in mitochondrial *cox1* genes of a diatom and a haptophyte. Gene 2000; 256:157-67; PMID:11054545; http://dx.doi.org/10.1016/S0378-1119(00)00359-0
17. Kamikawa R, Masuda I, Demura M, Oyama K, Yoshimatsu S, Kawachi M, Sako Y. Mitochondrial group II introns in the raphidophycean flagellate *Chattonella* spp. suggest a diatom-to-*Chattonella* lateral group II intron transfer. Protist 2009; 160:364-75; PMID:19346162; http://dx.doi.org/10.1016/j.protis.2009.02.003
18. Ravin NV, Galachyants YP, Mardanov AV, Beletsky AV, Petrova DP, Sherbakova TA, Zakharova YR, Likhoshway YV, Skryabin KG, Grachev MA. Complete sequence of the mitochondrial genome of a diatom alga *Synedra acus* and comparative analysis of diatom mitochondrial genomes. Curr Genet 2010; 56:215-23; PMID:20309551; http://dx.doi.org/10.1007/s00294-010-0293-3
19. Oudot-Le Secq MP, Fontaine JM, Rousvoal S, Kloareg B, Loiseaux-De Goër S. The complete sequence of a brown algal mitochondrial genome, the ectocarpale *Pylaiella littoralis* (L.) Kjellm. J Mol Evol 2001; 53:80-8; PMID:11479679; http://dx.doi.org/10.1007/s002390010196
20. Bonen L, Vogel J. The ins and outs of group II introns. Trends Genet 2001; 17:322-31; PMID:11377794; http://dx.doi.org/10.1016/S0168-9525(01)02324-1
21. Toor N, Hausner G, Zimmerly S. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. RNA 2001; 7:1142-52; PMID:11497432; http://dx.doi.org/10.1017/S1355838201010251
22. Lambowitz AM, Zimmerly S. Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb Perspect Biol 2011; 3:a003616; PMID:20463000; http://dx.doi.org/10.1101/cshperspect.a003616
23. San Filippo J, Lambowitz AM. Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. J Mol Biol 2002; 324:933-51; PMID:12470950; http://dx.doi.org/10.1016/S0022-2836(02)01147-6
24. Zoschke R, Nakamura M, Liere K, Sugiura M, Börner T, Schmitz-Linneweber C. An organellar maturase associates with multiple group II introns. Proc Natl Acad Sci U S A 2010; 107:3245-50; PMID:20133623; http://dx.doi.org/10.1073/pnas.0909400107
25. Nishimura Y, Kamikawa R, Hashimoto T, Inagaki Y. Separate origins of group I introns in two mitochondrial genes of the katablepharid *Leucocryptos marina*. PLoS One 2012; 7:e37307; PMID:22606358; http://dx.doi.org/10.1371/journal.pone.0037307
26. Hayashi-Ishimaru Y, Ehara M, Inagaki Y, Ohama T. A deviant mitochondrial genetic code in prymnesiophytes (yellow-algae): UGA codon for tryptophan. Curr Genet 1997; 32:296-9; PMID:9342410; http://dx.doi.org/10.1007/s002940050280
27. Inagaki Y, Ehara M, Watanabe KI, Hayashi-Ishimaru Y, Ohama T. Directionally evolving genetic code: the UGA codon from stop to tryptophan in mitochondria. J Mol Evol 1998; 47:378-84; PMID:9767683; http://dx.doi.org/10.1007/PL00006395
28. Sánchez Puerta MV, Bachvaroff TR, Delwiche CF. The complete mitochondrial genome sequence of the haptophyte *Emiliania huxleyi* and its relation to heterokonts. DNA Res 2004; 11:1-10; PMID:15141941; http://dx.doi.org/10.1093/dnares/11.1.1
29. Smith DR, Keeling PJ. Twenty-fold difference in evolutionary rates between the mitochondrial and plastid genomes of species with secondary red plastids. J Eukaryot Microbiol 2012; 59; 181-4; PMID:22236077; http://dx.doi.org/10.1111/j.1550-7408.2011.00601.x
30. Inagaki Y, Hayashi-Ishimaru Y, Ehara M, Igarashi I, Ohama T. Algae or protozoa: phylogenetic position of euglenophytes and dinoflagellates as inferred from mitochondrial sequences. J Mol Evol 1997; 45:295-300; PMID:9302324; http://dx.doi.org/10.1007/PL00006233
31. Kamikawa R, Hosoi-Tanabe S, Nagai S, Itakura S, Sako Y. Development of a quantification assay for the cysts of the toxic dinoflagellate *Alexandrium tamarense* using real-time polymerase chain reaction. Fish Sci 2005; 71:987-91; http://dx.doi.org/10.1111/j.1444-2906.2005.01055.x
32. Kamikawa R, Kolisko M, Nishimura Y, Yabuki A, Brown MW, Ishikawa SA, Ishida K, Roger AJ, Hashimoto T, Inagaki Y. Gene content evolution in Discobid mitochondria deduced from the phylogenetic position and complete mitochondrial genome of *Tsukubamonas globosa*. Genome Biol Evol 2014; 6:306-15; PMID:24448982; http://dx.doi.org/10.1093/gbe/evu015
33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215:403-10; PMID:2231712; http://dx.doi.org/10.1016/S0022-2836(05)80360-2
34. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997; 25:955-64; PMID:9023104; http://dx.doi.org/10.1093/nar/25.5.0955
35. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003; 31:3406-15; PMID:12824337; http://dx.doi.org/10.1093/nar/gkg595
36. Korab-Laskowska M, Rioux P, Brossard N, Littlejohn TG, Gray MW, Lang BF, Burger G. The Organelle Genome Database Project (GOBASE). Nucleic Acids Res 1998; 26:138-44; PMID:9399818; http://dx.doi.org/10.1093/nar/26.1.138
37. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004; 32:1792-7; PMID:15034147; http://dx.doi.org/10.1093/nar/gkh340
38. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006; 22:2688-90; PMID:16928733; http://dx.doi.org/10.1093/bioinformatics/btl446
39. Lartillot N, Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol 2004; 21:1095-109; PMID:15014145; http://dx.doi.org/10.1093/molbev/msh112
40. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol 2008; 25:1307-20; PMID:18367465; http://dx.doi.org/10.1093/molbev/msn067