

Molecular evolution of mitochondrial introns in the liverwort *Marchantia polymorpha*

By Kanji OHYAMA^{*1,†} and Miho TAKEMURA^{*1}

(Communicated by Yasuyuki YAMADA, M.J.A.)

Abstract: We here describe in detail the characterization and molecular evolution of group II introns in the mitochondrial genome of the liverwort *Marchantia polymorpha*. We find that 18 introns of the 25 group II introns can be assigned by their similarities to six clusters, indicating an intra-genomic propagation of one ancestral intron each into the respective clusters in the liverwort mitochondrial genome. Interestingly, the intra-genomic propagation of some of these introns occurred only after the evolutionary separation of the bryophytes from the other clades of plants. Finally we report that the maturase-like sequences in the liverwort group II introns have further evolved by horizontal and independent transposition and substitution by analogous sequences from other fungal introns.

Keywords: mitochondrial introns, *Marchantia polymorpha*, intron evolution, maturase-like ORFs, intra-genomic propagation of introns

Introduction

On the basis of structural features, in a combination of conserved nucleotide sequences and potential secondary structures, two types of introns can be classified in organelles, namely, group I and group II introns.¹⁾ Group I and group II introns were originally described as two families of introns that are distinguished by unique secondary structures.²⁾ Group I introns are widely distributed over the genomes of bacteriophages,³⁾ prokaryotes,⁴⁾ organelles,⁵⁾ and nuclei.⁶⁾ Group II introns are present in the mitochondrial genomes of fungi⁵⁾ and plants,⁷⁾ and in chloroplast genomes.^{8),9)} The complete nucleotide sequence of the liverwort mitochondrial DNA reveals 94 possible genes in the total length of 186,608 basepairs.¹⁰⁾ Seventeen of these genes are interrupted by a total of 32 introns (Fig. 1). Based on their sequence and structural features, twenty-five of these introns can be assigned to the group II, the remaining seven qualify as *bona fide* group I introns. Here we describe the

detailed characterization of group II introns and derive the molecular evolution of the introns in the mitochondrial genome of the liverwort *Marchantia polymorpha*.

Materials and methods

Computer aided analysis. The complete nucleotide sequence of the liverwort mitochondrial DNA was determined in the laboratory of Plant Molecular Biology, Kyoto University. Computer aided analysis was carried out against the sequence database in the GenBank (accession number M68929) using the Hitachi DNASIS program, and the BLAST, FASTA and ODN programs (DNA Data Bank of Japan, National Institute of Genetics, Japan). Phylogenetic analysis was performed with the CLUSTALW program.

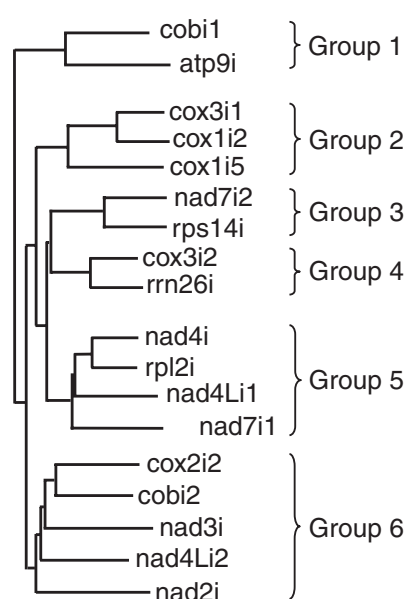
Results and discussion

Six clusters of homologous group II introns.

In the complete nucleotide sequence of the liverwort mitochondrial DNA twenty-five group II introns can be identified by their consensus sequences and secondary structures.¹⁰⁾ Sequence comparison of these group II introns reveals six clusters of highly similar introns.¹¹⁾ In order to identify the relative timing of the evolutionary processes having led to the respective clusters of group II introns, we

^{*1} Research Institute for Bioresources and Biotechnology, Ishikawa Prefectural University, Ishikawa, Japan.

[†] Correspondence should be addressed: K. Ohyama, Ishikawa Prefectural University, 1-308 Suematsu, Nonoichi-machi, Ishikawa 921-8836 Japan (e-mail: kohyama@ishikawa-pu.ac.jp).



Only genes with intron(s) are shown. Black boxes indicate exons. Yellow boxes and green boxes indicate group I and group II introns, respectively. Introns including intact and frame-shifted ORFs are indicated by a single asterisk (*) and a double asterisk (**), respectively. Arrows indicate the direction of transcription.

and direction of this intra-genomic propagation will be discussed below.

Intra-genomic propagation of the liverwort group II introns via an RNA intermediate into genomic sequences with splice site similarity. The high similarity of these introns suggests that they are derived from a common ancestor by duplication and insertion into another site. Similar to retrotransposons, group II introns multiply by reverse transcription of the RNA, in this case the excised intron, and subsequent insertion of the cDNA into a new genomic locus. To insert the DNA fragment generated from the RNA intermediate by a reverse splicing reaction, base pairing interactions between the insertion site which interacts as an intron binding sequence (IBS) in the exon and the exon binding sequence (EBS) in the intron are required.¹³⁾ This necessary compatibility between the EBS sequence of the moving intron and the IBS-like sequence of a novel insertion locus in a different exon will subsequently ensure the correct insertion of the complete new intron into the previously intron-less mRNA sequence during the reverse splicing step.

Consequently, we searched the splice sites of

Group 1

	cobi1
*atp9i	424

Group 2

	*cox1i2	*cox1i5
cox3i1	196	724
*cox1i2		935

Group 3

	rps14i
*nad7i2	260

Group 4

	cox3i2
rrn26i	277

Group 5

	nad4i	rpl2i	*nad7i1
nad4Li	467	506	485
nad4i		277	400
rpl2i			400

Group 6

	nad3i	nad4Li2	*cox2i2	cobi2	O.b.
nad2i	573	550	620	620	400
nad3i		573	528	620	903
nad4Li2			528	596	1006
*cox2i2				573	870
cobi2					809

Fig. 3. Divergence times of different pairs of homologous introns in millions of years.

Search for intron sequence similarities within the liverwort mitochondrial sequence was performed by using the FASTA and ODEN programs (gap penalty: $a = 4$) on a FACOM-M-77/10UTS computer (DNA Data Bank of Japan, National Institute of Genetics, Mishima, Japan). Introns which encode RNA maturase like ORFs are indicated by asterisks. Intron *O. berteriana* is the third intron in the *nad2* gene of *O. berteriana*.¹²⁾ The rate of substitution per site (K) is calculated by the Jukes-Cantor method: $k = -3/4 \ln(1 - 4p/3)$ where p is the fraction of observed substitutions.¹⁶⁾ The rate of substitutions per site per year is calculated as follows: $k = K/(2t)$, with $t = 400$ Myr. Dates of divergence are calculated as follows: $T = (K/2)/(0.395 \times 10^{-9})$. 0.395×10^{-9} is the rate of substitution per site per year for introns, estimated from the comparison of the *nad2i* sequence in the liverwort and in the flowering plant *Oenothera*.

the various group II introns in the mitochondrial mRNAs for similarities in the surrounding exon sequences for compatible IBS-like motifs. Indeed, evaluation of these IBS-similarities does indicate the direction of the intra-genomic propagation of various intron pairs in different liverwort mitochondrial genes: Intron *cox1i2* has duplicated and homed into the *cox3* mRNA to become *cox3i1* (Fig. 4), *rrn26i* amplified into *cox3i2*, and *nad4i* has duplicated and evolved one copy to *rpl2i*. This trend is especially prominent in those introns which propagated intra-genomically after the evolutionary separation from the line of the higher plants. Between these intron pairs, the potential base-pairings of the deduced EBS-IBS sequences are particularly high probably because of the short time scale since the duplication and insertion of these intron copies from the corresponding ancestral introns.¹¹⁾ The direction of the propagation of the *nad7i2* and *rps14i* introns is not clear, because the *pseudonad7* gene has several stop codons in its coding region and the IBS-EBS sequences show the same degree of matching base-pairs in either direction.

Intron propagation via an RNA intermediate requires the activity of a reverse transcriptase, which is often supplied by the ORFs sometimes encoded in group II introns.¹⁴⁾ As the final step of

amplification and transposition, homologous recombination with an endonuclease or integrase activity is needed to insert the DNA copy of the intron into the new genomic locus.¹⁵⁾ However, these endonucleases or integrases can also act in *trans* and can thus be encoded by other introns. For example, as is common for group II intron ORFs, some of the ORFs encoded by the group II introns in the *M. polymorpha* mitochondrial genome do not code for proteins with endonuclease or integrase domains. The required enzymatic activity may be supplied by one or more of the proteins encoded by group I introns in the liverwort mitochondrial genome, such as the *cox1i4* or *cox1i8* introns. The ORFs encoded by these two introns do contain such motifs typical for an endonuclease activity,¹⁶⁾ suggesting that they might enable or enhance homologous recombination. Thus the ORFs encoding proteins in the group I introns in liverwort mitochondria might participate in the intra-genomic propagation of the group II introns in the genome. Interestingly, these intra-genomic propagations in the liverwort mitochondrial genome are only seen for group II introns, but not for group I introns. The reason for this bias remains unclear at present. Possibly one of the three main requirements for frequent intra-genomic propagation events (reverse splicing, reverse tran-

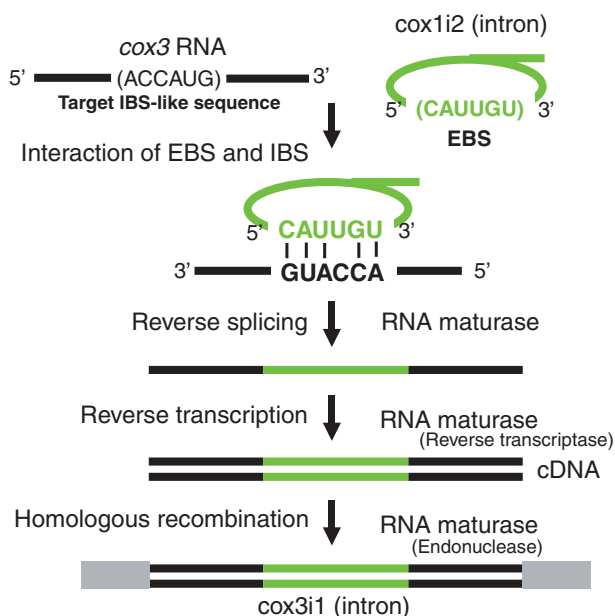


Fig. 4. Schematic presentation of the proposed process of intra-genomic propagation of group II introns.

As an example, the EBS of the *cox1i2* intron (lariat green lines) binds to the IBS-like sequence of the *cox3* mRNA and inserts by reverse splicing into the *cox3* mRNA and then the cDNA was synthesized by the reverse transcriptase activity of the intron-encoded RNA maturase. Finally, in separate and independent steps, the cDNA with the *cox3i1* intron (copied and propagated from the original *cox1i2*) eventually replaced by homologous recombination all the copies of the genomic *cox3* gene without intron on other mitochondrial DNA molecules.

scription, and homologous recombination) as depicted in Fig. 4 is not met by the group I introns and consequently no successful amplifications have survived the evolutionary selection to be still detectable today.

Evolutionary origin of five *cox1* introns inserted at the same sites as those of their fungal counterpart. While the *cox1* genes of higher plants contain no introns at all, there are nine introns in the liverwort mitochondrial *cox1* gene coding for cytochrome *c* oxidase subunit 1.¹⁰⁾ We have previously described that six of these introns, the 3rd, 4th, and 6th to 9th introns show all the characteristics of group I introns, while the rest of the *cox1* introns, the 1st, 2nd, and 5th introns, can be clearly classified as group II introns. Five of these *cox1* introns, *cox1i2*, *cox1i4*, *cox1i6*, *cox1i7*, and *cox1i8* are inserted at the same sites where introns have been reported in the genes of fungal mitochondrial *cox1* genes.^{10),16)}

To analyze the timing of the evolutionary events leading to this distribution of introns in the mitochondrial *cox1* gene of the liverwort relative to the intron evolution in fungi, we constructed a phylogenetic tree of the mitochondrial *cox1* gene from the liverwort and four species of fungi. As a reference time scale and evolutionary marker we used a nuclear gene, the ribosomal 5.8S rDNA sequences^{17)–20)} from the same species for an analogous phylogenetic tree. Comparison of the two derived trees shows that the mitochondrial *cox1* genes of *Saccharomyces cerevisiae*,⁵⁾ *Schizosaccharomyces pombe*,²¹⁾ *Neurospora crassa*²²⁾ and *Podospora anserina*,²³⁾ respectively, show less sequence divergence and have thus evolved somewhat slower between *M. polymorpha* and the four fungi than have the nuclear encoded 5.8S rDNA gene sequences (Fig. 5A and 5B).

In order to analyze the possibility of the liverwort *cox1* introns inserted at the same sites as the respective *cox1* introns in the fungal genes being derived from the same ancestral intron, we compared the amino acid sequences of the ORFs encoded by the liverwort *cox1i4* and *cox1i8* group I introns to the respective ORFs in the analogous introns in fungi. Total database searches with the putative proteins encoded in the *cox1i4* and *cox1i8* introns in the liverwort indeed identified as closest matches the intron counterparts inserted at the same sites in the fungal mitochondrial genes.¹⁶⁾ The *S. pombe* ORF in *cox1i1* intron is most similar to

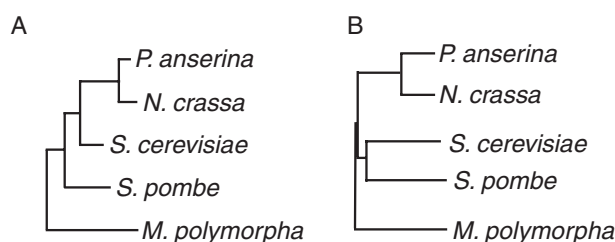


Fig. 5. Phylogenetic trees of the liverwort and fungal nuclear-coded ribosomal 5.8S rDNA genes and the mitochondrial *cox1* genes.

(A) Phylogenetic tree of the nuclear-encoded ribosomal 5.8S rDNA sequences. GenBank accession numbers are: *P. anserina* (AE388930); *N. crassa* (X02447); *S. cerevisiae* (D89866); *S. pombe* (AB054041, unpublished); *M. polymorpha* (AB021684). (B) Phylogenetic tree of the mitochondrial *cox1* nucleotide sequences from the same species. GenBank accession numbers are: *P. anserina* (X55026); *N. crassa* (X14669); *S. cerevisiae* (YSCMTOC1); *S. pombe* (X54221).

the liverwort ORF in *coxli4* intron, and the liverwort ORF in *coxli8* intron identified as closest relatives the ORFs in the introns *coxli15* in *P. anserina*, *coxli5B* in *S. cerevisiae*, and *coxli3* in *S. pombe*, respectively, as described previously.¹⁶⁾ For these introns, the relationships and thus evolutionary origins and pathways correlate for the three features splice site (i.e. EBS/IBS), intron sequence and structure, and the encoded ORF.

An analogous observation is made for the structures and EBS/IBS interaction sites of the liverwort group II intron *coxli2* and the *S. cerevisiae* intron *coxli1*, which is inserted at the same site. However, when analyzing the ORF encoded by this intron *coxli2* in the liverwort *cox1* gene, similarity was found to be higher to the ORF in an intron in the fungus *N. crassa* intron *coxli1* (41% similarity) than to the *coxli1* intron in *S. cerevisiae*, which is inserted at the genomic site homologous to liverwort (35% similarity; Fig. 6). This finding suggests that the ORF in the liverwort *coxli2* intron has been horizontally replaced by the respective ORF from *N. crassa*. On the other hand, the lower sequence similarity observed between the respective ORFs in the liverwort *coxli2* and *coxli5* introns is probably due to frameshift of the ORF in *coxli5* (Fig. 6). Since the respective introns were classified to same group (group 2 in Fig. 2), the intron *coxli2* (or part of it) subsequently propagated intragenomically and invaded the liverwort *coxli5* intron where it also replaced the previous ORF during the evolution (Fig. 7).

A somewhat different scenario is deduced for the origin and evolution of the introns *nad7i1* and *rnr18i* in the liverwort mitochondrial genome. These two introns most likely diverged independently and at different times from the same intron in fungi (see Fig. 2). Although intron *rnr18i*, unlike *nad7i1*, does not belong to any of the six groups by its overall sequence similarity pattern, the two ORFs encoded by the liverwort *nad7i1* and *rnr18i* introns show higher similarity with the ORF encoded in the *P. anserina* intron *coxli1* (42% and 44%, respectively) (Fig. 8). This observation implies that this ORF transposed horizontally and separately into each of the two liverwort introns from the ancestral fungal intron (Fig. 9). We have previously identified an analogous evolutionary pathway for the liverwort *coxli6* and *coxli7* introns and their encoded frame-shifted ORFs.¹⁶⁾ As de-

Mp <i>coxli2</i> ORF	MNNFAQRWLFSTNHKIDGTYLIFGAIAGVMGTCFSLVIRMLAQPGNQILGGNHQL
Mp <i>coxli5</i> ORF	VRRARLVFGKEXYPSCAYVXLLXGNTCSRMRHLELIHIVHCQQLIKGDAKDC
Nc <i>coxli1</i> ORF	MSSISITWTERWFLSTNAKIDGVLYLIFALFSGLLGTAFAVLIRMLSGVGVYIADNQLY
Sc <i>coxli1</i> ORF	MVQRWLYSTNAKIDAVLYFMLAIFSGMAGTAMSLIIRLELAAPGSQYLHNSQL
Mp <i>coxli2</i> ORF	YNGAPGYTSSDKKSYPLFPVPSRLANKLGHRLNLRHNSMAYSAASFNRTAGLPLRG
Mp <i>coxli5</i> ORF	PSWWCLLAGVVKVRLTNLIRTHCLIRGYISQAEYVMCAEXPRDLGSRKLLMLLAVRA
Nc <i>coxli1</i> ORF	N-----AIITAHAILMSAPMCFGVGLFIRPLMFQGVKWPAPAGELISLGETQTAKRRL
Sc <i>coxli1</i> ORF	FN-----VLVVGHAFLMIFCAPFRLIYHCIEVLIDKHISVYSINENPTVSPFWLLVVTYM
Mp <i>coxli2</i> ORF	RPERAEGRARITDLTNRARITVDLLSQAHHYKSTISKPLRSELCTIHCVCTVPGKMWG
Mp <i>coxli5</i> ORF	VQSQGIPITXELSAIXEVRKDLTPLG-----ALRMVRAKTRKXSNHGX
Nc <i>coxli1</i> ORF	KGRNLINIAIIGLLPQLSYTTIRLTN-----LFTFDLSWSG
Sc <i>coxli1</i> ORF	VFRVNHMAYFVGANSTGTMACHKASG-----VKQPAQGNKX
Mp <i>coxli2</i> ORF	TLHRYQLSNQLAQPNFSGKIFLGFTK-KGSKCLPDTYKNDYVAGAKITRP---NRD
Mp <i>coxli5</i> ORF	PLHVVQTSRGLRGPVVVKERENPARPRRLVTRRILQLMSRANPLASLERIVGRAT
Nc <i>coxli1</i> ORF	RLRYIKSRRLSGHFIISLRITQMTLGNTHDDCINNTKCRDPSTITALAKE---LRD
Sc <i>coxli1</i> ORF	PMARLTNSCKECLGFSLTPSHLGIVIAHVLEEEVHELTYKESLALSKEHLEGTSSNG
Mp <i>coxli2</i> ORF	GFRARNSCGFLQSSLAVKPESSRSYCSIPESVDNETSSRNGSARPAKAAWSDVRMS
Mp <i>coxli5</i> ORF	PVHMLRKWXTIVKTRDTIMDLXELYIQCFWPMRVKVLRLAIXLQDMQVKPKPTGX
Nc <i>coxli1</i> ORF	HLRVALGNLGVPKYHGIAAGLNLRLGGGAIIVRGKLGK--VAVPSQMTIRTISSKAGQS
Sc <i>coxli1</i> ORF	KLRNTGLSERGNPDNGVFMVFKNLNKNVRYFSTLSKLNARKEDSLAYLTINTITDFSEL
Mp <i>coxli2</i> ORF	VSEIQAYLGP---DNRYGLIHIISDPTFLALCYESIRGKPG--TSGSDAKPLDGP--
Mp <i>coxli5</i> ORF	TGILYNLRNVXAKGRPKLNLPBEYLYIQCFWPMRVKVLRLAIXLQDMQVKPKPTGX
Nc <i>coxli1</i> ORF	AKRSDDIIVS---TSKPDINMKAANMGLIIVAYELLKSNPNMGTGKANTITDGLNL
Sc <i>coxli1</i> ORF	NKLMENHNKT---ETINIRILKLMDSIRMLLIYVKNKISKSGKMSKGSNNITFDGLNI
Mp <i>coxli2</i> ORF	EMFVQVGEKLLKGGQFEFSPAR--RITPKGKKEKRLPGINSVVKQKCYGEKIVQKALQL-
Mp <i>coxli5</i> ORF	PKTKLSDRARDQTKKQKNSQKNIHETLGRGITCSTCRMLSPFQEKXKXKHSNX
Nc <i>coxli1</i> ORF	KFLEKIQRLDRDGHKEFFPARRIQIPKPKKKEKRLPLTIASPR---DKVVQKAIQL-
Sc <i>coxli1</i> ORF	SYLNLKSKDINTNMFKFSPVRRVEIPKTSGG-FRPLSVGNFR-----EKIVQESMRM-
Mp <i>coxli2</i> ORF	VLEAIYEPFLDCSHGFRHRSCHTALKRLCLEGHYPVVEGNIRKFFDSIPHKVILHK
Mp <i>coxli5</i> ORF	SWKPYKNDFTALMDIQTGXIMSLCPTPLFDRQSPFMGVLRLGSLNCFDKIPHSTIMKR
Nc <i>coxli1</i> ORF	VMEPVFEKIFLDCSHGFRPHRGTKTAIQYVDAKFQSSHPIEADFSKAFDSIAHSLMEF
Sc <i>coxli1</i> ORF	MLEIYYNNSFSYSHGFRPNLSCLTAIQCNNYMQCNVFIKVDLNKCFDTTIPHMLINV
Mp <i>coxli2</i> ORF	ISQKVKCHRTLELLQRLRAGYKDPSTGQVVISLDEGTSQGSVLSPLCNIIHLHDEPVM
Mp <i>coxli5</i> ORF	LTALIKCQRTLELVKPLKAGYIDLENGKVIHSHSGTQGSVLSPLCNIVLHLDQFML
Nc <i>coxli1</i> ORF	LKETITCEKTLKIRSLGKAGYIEFGLHNN-LDITGPQGSILSPLCNIFLHRLDPMF
Sc <i>coxli1</i> ORF	LNRIKDKGFMDLLYKLLRAGYVKNNNYNN-TTIGTPQGSVSPILCNITFLDKLDKYLE
Mp <i>coxli2</i> ORF	K-LRDRFNKG--KSRINPEYKLLT---RRHMANRQDRS-----LLIKRRLIPSKDPLD
Mp <i>coxli5</i> ORF	S-MXNRPKKG--INRRENPTYRPFRRKKRYSNPAIRRA-----MLMEMRRNPKYDVID
Nc <i>coxli1</i> ORF	S-IAEFNIG--VKKRSKEYMALMMKCYMRSGQSDISNPELVHAIIRNNMLTTPSPVTKD
Sc <i>coxli1</i> ORF	NKFENEFNTGNMNRGRNPIYNSLSSIIYRCKLSSEKLF-----LRLRDHYQRNMGSD
Mp <i>coxli2</i> ORF	PYFRRLIYVRAYDDFVILVSGTRLETFAIQASQLNPLHRSRLLESLKTVVSHLANKGF
Mp <i>coxli5</i> ORF	PNFRRLSYIRYVDDFVILVYGTNRBAEKIRAEIQSFLEQACGLNLDNKILITLSEWF
Nc <i>coxli1</i> ORF	DSYVRYVRYVAYDDFIIGVSGSHKTAIVALEIKVQSFVTNQLGLRLNDKGTITKYSDPV
Sc <i>coxli1</i> ORF	KSFKRAYFVRYADDIIIGVMSHNDCKNLINDINFLNKLMSINMDKS-VIKHSKEGV
Mp <i>coxli2</i> ORF	HFLGTGCKRTSR-----HRIFHVRT--VRGKTIQRSTERLRVCAPITKLFYK
Mp <i>coxli5</i> ORF	DFXELANVKKPPEGHSTQSKVPISHGELTPERXCLLPTCCSNRQGSKKKPIXLXCTQLQ
Nc <i>coxli1</i> ORF	KFLGYKMKAPHKMGIVKPMVNLAKYNTLEG--TETRTIARKKIRIRPHMDYKVLKR
Sc <i>coxli1</i> ORF	SFLGYDVKVTPE-----KRPYRMK--KGNDFIRVHHTLSVNVNAPIRSIYMK
Mp <i>coxli2</i> ORF	LKE---KGFVKR-----NEMGKYVPTARRNLTPLDHADILEYNQKVRGTILNYSPAS
Mp <i>coxli5</i> ORF	XEDXWDYRTLISWLPITIRLEGCLTITPQLVIEICKKXYSGLTHALXPLXNRSESE
Nc <i>coxli1</i> ORF	LET---NRFIRKRTSHVHNKLIYRGTFKGNLINDHADIIINYNSVMRGIYNYDYPTS
Sc <i>coxli1</i> ORF	LHK---HGYCSHGILGKPRGVGRILHEEMKTIVLDVPHSKVSFSPSIDFPIRHKMMTDS
Mp <i>coxli2</i> ORF	NRSSIANQIVHVLHMSCALTLALKYKLTASKTFNRPKCLTCFAT-----GMSLFRRPS
Mp <i>coxli5</i> ORF	KKRSINLEIFCDRIIPRAAKTFQRTIRFYSITKXRVQWLLIRLSRFLGLGNFTSLGGR
Nc <i>coxli1</i> ORF	NAPLANVWMLLTSCALTLARKYKLTLSKVRFRKDKLDCITSAKSGKRRISIFKPE
Sc <i>coxli1</i> ORF	N-----YTPDEILDYKYMLPRLSLPSGICQICQSGHDLVHVHRTLNAA
Mp <i>coxli2</i> ORF	AYKAIHLYNPS-----PIARAEQVIDIS
Mp <i>coxli5</i> ORF	FAPFVIRTWSFTXGLFETPEPNLEIXQLTRNGXGTENRPHYVVELTTLKIMRGGLGN
Nc <i>coxli1</i> ORF	DFKKKSIMNGSNPTDRDPAGIDKVNNAKFTSNLFTATCIIIGBTQVDVEMHHVIRKIDRLN
Sc <i>coxli1</i> ORF	NKIKDDYLLGR-----
Mp <i>coxli5</i> ORF	RKFRXYHPIISNPGXPPLSALSIKXPICYRNRLDTINSAISVFXGRAVREIVTYGSEGI
Nc <i>coxli1</i> ORF	QESKLDFFPFRQMAAINRQVPLCKTHHIGLHNTWSEADKATRELAKKPSVKKKKESN
Sc <i>coxli1</i> ORF	-----MIKMNKQITICKTCHFPVHQKYGNGPL (35%)
Mp <i>coxli5</i> ORF	DXPY (19%)
Nc <i>coxli1</i> ORF	WKAKQKNN (41%)

Fig. 6. Sequence comparison of the ORF sequences in the liverwort *coxli2*, *coxli5*, *N. crassa* *coxli1*, and *S. cerevisiae* *coxli1* introns.

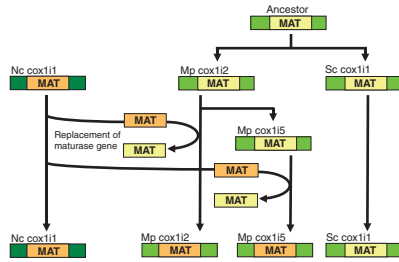


Fig. 7. Possible propagation pathways of ORFs (yellow and orange boxes with MAT) encoded by the liverwort introns cox12, cox15 (light green boxes), and *N. crassa* intron cox11 (dark green boxes).

Pa cox11i ORF	GAPFNVRKSGELDPISYCCCLNLLTYLMTKGLR
Mp nad7i1 ORF	APFKQSMCSRADCLISITLSLKGKPKGYCSMMWKGEGPVKPIALCSTSYSSMDLSSSSG
Pa cox11i ORF	ECMSVNPYLTAIKSVESGEVKASYVLRLLTMVGLCVSIGIKIAIALYWLKISASLI
Mp nad7i1 ORF	VIRLRLVGLTNRGCDWQNHVYVSGHCDLLQATASNLVRKNGISWGQLPYKYGHGSE
Pa cox11i ORF	KNSYSFTTSEYGIYVY-----AKGRRRLNVGNSGLPKGRNSYNGSVVVGTSV-----K
Mp nad7i1 ORF	PTGQMTTHKEKNCLKPPSPKTRAGQRDIPSSPSPDVSLERGGGFAVMAPTGDGSE
Pa cox11i ORF	WIHTKVRKSKVSPKSEARKGQGLGEMLMYNERGQCINAYEVICKLEALYTAYMNIKSE-
Mp rrn18i ORF	MTKCNYSQLLDPEIFRLAYELKSKS-
Mp nad7i1 ORF	QGTSPHASSWPPAAGRRKRATSARTLSTTTIGMRERPNKSLHLKRDREGTQNDLSPS
Pa cox11i ORF	PQNMTFRVDSLETDLGISEKWEFEKISEQLKSE-----
Mp rrn18i ORF	SGNMKPGADKETLDGFSQAYVEKVVRLQKDE-----
Mp nad7i1 ORF	HDKLRFPYTDHHSMAFSPKAGSETQNSKPPGLPMHMGPDVTVRTYGLYAVSPQGPETFN
Pa cox11i ORF	QFRFRPTRRVYIPKANGKMRPLGIASPRDKIVQEVFAILEQVLEPRFHSSSHGFRPGRG
Mp rrn18i ORF	SFQFRPSRREFIPKADGKRLSLGIPSPDKIVQEVVRRILEPFVEPRFLDSSHGFRPHRS
Mp nad7i1 ORF	SRALRGLHRSDFLQGTIKRRLHELPISYKIATHISNRTYIPATKSKSFTIDRSERIILT
Pa cox11i ORF	CHSALATIRYWNIGIKWFIEGDIKGFDDNIDHHLKLLVKHFQ-----
Mp rrn18i ORF	PHTALRQIRRWGTGTSWMIEGDIKGYFDNIDHLLAGFIAELVK-----
Mp nad7i1 ORF	IRHKLTHNLNLTAKGKHDSGTADSDNRMGFETIARLKDKKSPGFKTFLGEFLPFPNQ
Pa cox11i ORF	-----DQRFIDLYWKMVKAGYVEFDKDK-----
Mp rrn18i ORF	-----DQRLALYWKLVFRAGYVNGKAE-----
Mp nad7i1 ORF	MEINVPWVYHYIAIKLYKLYRQLSNPLSNENFCPLVMGVSQAGLLVLHEKPEMECKLG
Pa cox11i ORF	-----SSIIIG-VPOGGIASPILSNLVNLDDEFVQNVINDEFNEKLKGGKHTS
Mp rrn18i ORF	-----PHLLTGVPGGRILSPLLSNILYHQFDLPMEIEIKVYTTTG-----
Mp nad7i1 ORF	NRRRHQRILLQYQPQTSFVPRCGTSRPIAFTTLTTSRGAGYGLKEQPVTVGLRGVLIS
Pa cox11i ORF	KNPAYVIDSRIGKITRLEKRLKSKQELDSGRKLERMKLIKVRATMPSIMP-NPDIAK-
Mp rrn18i ORF	-----ALSKNNPIYLKARKYKLVKSLKASSAEIIRARMDLMKMYGIOTGSR-
Mp nad7i1 ORF	PMLPNILGLPQDFCEBLKIRYKAPTSLITRTIQTQVQKIRGRPESTTGMKINREEGSK
Pa cox11i ORF	--IYVFRYADDWLVGAGSSETARAIKERIAAYLKDILKLELSMEKTLITNASEDKAYFL
Mp rrn18i ORF	--VRYVRYADDWLVGAGSSETARAIKERIAAYLKDILKLELSMEKTLITNASEDKAYFL
Mp nad7i1 ORF	NLLPTICKRLDRRGVVGSKKVALAIRAETASFLKHLHLN---WDNTKITHISSQLALFL
Pa cox11i ORF	GTEIQRISVVKGEIKRPFKNIKGHPQRIPTTSTVMNAPISKLVTKLADKGIWIKSKALNE
Mp rrn18i ORF	GTLSITITRKVYVQSKVG---GGHRRASLGRIRLCIPIDILIGLSQMGACDEK-----
Mp nad7i1 ORF	GTHIKVLRAESIRNHRIL-VVGQRTSATFRLHLAPIERIVKHLHGKGLCTPTG-----
Pa cox11i ORF	DNLIPQPIKWNPLPIDIILRYKMIWNGYINYSFADNKPRLVLIWILRSLAKTLAT
Mp rrn18i ORF	--GTPKAVTKWIFLNVGEIINKYMAVFRGYNYYSFADNKPRLVLIWILRSLAKTLAT
Mp nad7i1 ORF	----KPKFVRWIPLDHHELIIPRYQDIMS GYNNYSFVDNYGMLKRVAYIVRFSAAATLKR
Pa cox11i ORF	KLKLGTVRKVLYLFGVNLRFILGTDNKSIEFTKGNLLPTKPNFGKTNFVDNLKVVES
Mp rrn18i ORF	KLGLN-TAKVIRKFGVDLIFRDHTNEIKHLNFRSLPNKRMNFALSP-----P
Mp nad7i1 ORF	KPKMLSVASVFKSGGTGRGKELAAATFLNLEKRCFAINNKKIMEFSARIR-----
Pa cox11i ORF	LRTVSFPNYVCASGASDNQLVHHVKHIRTIDVKLSGFDKQLAANRQVTLCSICHNV
Mp rrn18i ORF	SDPRVLFDTS CARIQC (44%)
Mp nad7i1 ORF	LRTHPVLTSPPCCICRNQENRMQHVKHLR--KSKANGLTKMLVQLNRKQIPVCRCHLKI
Pa cox11i ORF	HTGKYDGMSLKYMKDIDSKPELNQQ
Mp nad7i1 ORF	HQGYDGMKSLPNRKPVGLPASVVEENPVGLSLDCIDFKLTHPLERSGRQEGKATDSPG
Mp nad7i1 ORF	QKLGPGLRESEDSLGKPYARKSCTSGSEBHYLTGKSLHNWNPISY (42%)

Fig. 8. Sequence comparison of the ORFs encoded by the rrn18i and nad7i1 introns with the *P. anserina* cox11i intron.

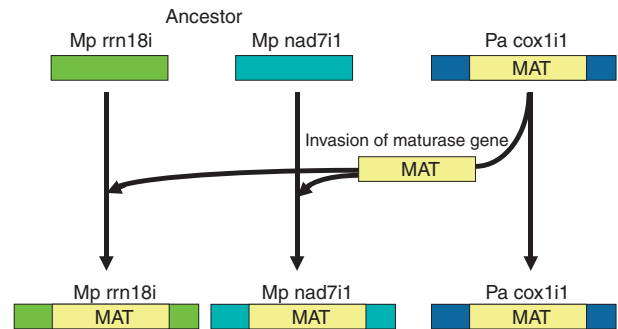


Fig. 9. Possible transposition pathways of the ORFs (yellow boxes) encoded by the rrn18i (green boxes) and nad7i1 (light blue boxes) introns from the *P. anserina* cox11i intron (dark blue boxes).

scribed above for the liverwort introns cox14 and cox18, the liverwort intron cox16 is also a cognate homolog of the *N. crassa* intron cox13 and the *P. anserina* intron cox17A. Likewise, liverwort cox17 is related to the *S. cerevisiae* cox14 and the *P. anserina* cox19 introns, which are apparently all derived from the same ancestral intron and encoded ORF.

Acknowledgements

We would like to thank Professor A. Brennicke, Ulm University, for kindly reading this manuscript, and the members in the laboratory of Plant Molecular Biology, Kyoto University for their great contribution to the work described in this paper. This research was in part supported by a grant-in-aid for Science Research in Priority Areas and by a grant for General Scientific Research from the Ministry of Education, Science and Culture, Japan.

References

- Michel, F. and Dujon, B. (1983) EMBO J. **2**, 33–38.
- Michel, F., Jacquier, A. and Dujon, B. (1982) Biochimie **64**, 867–881.
- Shub, D.A., Gott, J.M., Xu, M.Q., Lang, B.F., Michel, F., Tomaschewski, J., Pedersen-Lane, J. and Belfort, M. (1988) Proc. Natl. Acad. Sci. USA **85**, 1151–1155.
- Kuhse, M.G., Strickland, R. and Palmer, J.D. (1990) Science **250**, 1570–1573.
- Bonitz, S.G., Coruzzi, G., Thalenfeld, B.E., Tzagoloff, A. and Macino, G. (1980) J. Biol. Chem. **255**, 11927–11941.
- Wild, M.A. and Gall, J.G. (1979) Cell **16**, 565–573.
- Fox, T.D. and Leaver, C.J. (1981) Cell **26**, 315–323.
- Ohya, K., Fukuzawa, H., Kohchi, T., Shirai, H.,

- Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z. *et al.* (1986) *Nature* **322**, 572–574.
- 9) Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K. *et al.* (1986) *EMBO J.* **5**, 2043–2049.
 - 10) Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T. *et al.* (1992) *J. Mol. Biol.* **223**, 1–7.
 - 11) Ohyama, K., Oda, K., Ohta, E. and Takemura, M. (1993) *In Plant Mitochondria* (eds. Brennicke, A. and Kuck, U.). Verlagsgesellschaft/VCH Publishers, Weinheim, pp. 115–129.
 - 12) Binder, S., Marchfelder, A., Brennicke, A. and Wissinger, B. (1992) *J. Biol. Chem.* **267**, 7615–7623.
 - 13) Augustin, S., Muller, M.W. and Schweyen, R.J. (1990) *Nature* **343**, 383–386.
 - 14) Michel, F. and Lang, B.F. (1985) *Nature* **316**, 641–643.
 - 15) Dujon, B. (1989) *Gene* **82**, 91–114.
 - 16) Ohta, E., Oda, K., Yamato, Y., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Ohyama, K. and Michel, F. (1993) *Nucl. Acids Res.* **21**, 1297–1305.
 - 17) Sone, T., Fujisawa, M., Takenaka, M., Nakagawa, S., Yamaoka, S., Sakaida, M., Nishiyama, R., Yamato, K.T., Ohmido, N., Fukui, K. *et al.* (1999) *Plant Mol. Biol.* **41**, 679–685.
 - 18) Dettman, J.R., Harbinski, F.M. and Taylor, J.W. (2001) *Fungal Genet. Biol.* **34**, 49–61.
 - 19) Kelly, J.M. and Cox, R.A. (1981) *Nucl. Acids Res.* **9**, 1111–1121.
 - 20) Oda, Y., Yabuki, M., Tonomura, K. and Fukunaga, M. (1997) *Yeast* **13**, 1243–1250.
 - 21) Lang, B.F. (1984) *EMBO J.* **3**, 2129–2136.
 - 22) Field, D.J., Sommerfield, A., Saville, B.J. and Collins, A. (1989) *Nucl. Acids Res.* **17**, 9087–9099.
 - 23) Cummings, D.J., Michel, F. and McNally, K.L. (1989) *Curr. Genet.* **16**, 381–406.

(Received Nov. 12, 2007; accepted Nov. 27, 2007)