Title: Organization and expression of mitochondrial genes for ribosomal proteins and NADH dehydrogenase subunits from a liverwort, Marchantia polymorpha

Author(s): Takemura, Miho

Citation: Kyoto University (京都大学)

Issue Date: 1995-03-23

URL: https://doi.org/10.11501/3080950

Type: Thesis or Dissertation

Textversion: author

Kyoto University
Organization and expression of mitochondrial genes for ribosomal proteins and NADH dehydrogenase subunits from a liverwort, *Marchantia polymorpha*.

Miho Takemura

1995
Introduction

Mitochondria, which are present in all eukaryotic cells, are the energy-converting organelles. They contain multiple copies of the mitochondrial DNA (mtDNA) and consequently their own genetic systems. Although the mitochondrial genomes of mammals and fungi have been well investigated in detail, the studies of those from plants have been made little progress because of their large size and complex structure. The mammalian mitochondrial genome is a circular molecule of about 16 kb (16,569 bp in human) (Anderson et al., 1981) and extraordinarily information is tightly packed. The fungi mtDNA is larger than that of mammal (about 80 kb in yeast or 94,192 bp in Podospora) (de Zamaroczy and Bernardi, 1986; Cummings et al., 1990). On the other hand, plant mitochondrial genomes, which vary in size from about 200 kb in Brassica (Labacq and Vedel, 1981) to approximately 2,500 kb in muskmelon (Ward et al., 1981), are much larger than those from mammals and fungi. They are usually organized multiple circular molecules, with conversion of circular forms mediated by frequent homologous recombination between repeated sequences. For example, the mitochondrial genome of Brassica campestris is found to be organized as a tripartite structure, a "master" circle of 218 kb and two subgenomic circles of 135 kb and 83 kb which are formed through a directly repeated 2 kb sequence (Palmer and Shields, 1984). In addition to the multiple partite structures, sequences highly homologous to chloroplast DNA are generally present in the plant mitochondrial genomes, and thus this may complicate structures of them. Different chloroplast sequences are found at various locations in the mitochondrial genomes of different species. These findings demonstrate that DNA transfer from chloroplast to mitochondria is common in higher plants and that most of the events might happen recently.

The molecular mechanisms that regulate mitochondrial gene expression have been most thoroughly investigated in animals (especially human) and fungi (particularly yeast). In vertebrate, each strand of the circular mitochondrial DNA is divergently transcribed from a single major promoter near the primary origin of
replication (reviewed in Clayton, 1984; Clayton, 1991). On the other hand, there are several sites of transcription initiation on the mitochondrial genome of yeast, *Saccharomyces cerevisiae* (Christianson and Rabinowitz, 1983). Moreover, it is demonstrated that not only mitochondrial RNA polymerase but also transcription factor are required for promoter-specific transcription in human and yeast (reviewed in Schinkel and Tabak, 1989). Studies of gene expression in plant mitochondria are considerably less well advanced than animal and fungal mitochondria. To date, several characteristics of transcription in plant mitochondria are almost clear (reviewed in Gray et al., 1992). Like the situation in yeast, transcription of plant mtDNA is initiated at multiple sites, although many separate initiation sites are present for a single gene unlike the case of yeast. The most highly conserved consensus of transcription initiation region have been identified in the monocot and also in the dicot. However, these consensus motifs seem to differ between the various plant species, indicating the flexibility of promoter sequences in plant mitochondria. Molecular mechanisms of other events occurred in plant mitochondria, for example RNA processing, trans-splicing and RNA editing, are still unknown.

Recently, mtDNA of a liverwort, *Marchantia polymorpha*, was found to be a single circular molecule of about 184 kb in size by electron microscopic observation and restriction endonuclease mapping in this laboratory (Oda et al., 1992b). Then, the complete nucleotide sequences of the liverwort mtDNA was determined and its entire gene organization was identified (Fig. 1, Oda et al., 1992a; Oda et al., 1992c). In the sequence of 186,608 bp, 96 possible genes were detected. These included genes for three species of ribosomal RNA, 29 genes for 27 species of tRNA, 31 open reading frames for functionally known proteins, five for functionally unknown open reading frames which showed similarity to those of other organisms, and 28 open reading frames predicted as possible genes. Thirty-two introns were found in the coding regions of 17 genes. Twenty-five of them belonged to group II introns and remaining seven were group I introns. At present, liverwort is the only plant species whose complete nucleotide sequences of mtDNA are available. In contrast to the heterogeneity of most plant mtDNA, no recombination was detected in liverwort (Oda et al., 1992a). It was also found that no sequences homologous to chloroplast DNA were present. Since the nucleotide sequences of the liverwort mitochondrial DNA were well-conserved at the DNA level in the course of evolution, RNA editing was apparently lacking in the liverwort mitochondria.

![](fig1.png)

**Fig. 1.** Gene organization of the mitochondrial genome from a liverwort, *Marchantia polymorpha* (Oda et al., 1992a; Oda et al., 1992c). Genes shown outside the map are transcribed anticlockwise, and those inside are transcribed clockwise. Solid and hatched boxes indicate exons and introns, respectively in the coding regions. Asterisks indicate introns having ORF homologous to RNA maturase. Genes for tRNAs are shown as *trn* with the one-letter amino acid codes and their anticodons. Genes encoding the small subunit and large subunit ribosomal proteins are shown as *rps* and *rpl*, respectively. *rrn*, *atp*, *nad*, *cox*, and *cob* represent the genes for ribosomal RNAs, ATPase subunits, NADH dehydrogenase subunits, cytochrome c oxidase subunits, and apocytochrome b, respectively. *orf5* indicate the open reading frames.
In this thesis, the author has intended to make progress in our understanding on the genetic information system of plant mitochondria using a liverwort, *Marchantia polymorpha*. In chapter I, sixteen genes for ribosomal proteins were detected. The genes formed two major clusters, very similar in organization to *E. coli* ribosomal protein operons. Transcription analysis of all the ribosomal protein genes were carried out. In chapter II, eight genes for NADH dehydrogenase subunits were characterized. Transcriptional analysis of these genes were performed. Almost all of them were supposed to be co-transcribed with their neighboring genes. In chapter III, tetrna7 was actively transcribed but the two predicted introns in the gene were not spliced. The Southern blot analysis of the nuclear DNA and the Northern blot analysis of the poly(A)^+ mRNA suggested that the nuclear genome encoded the mitochondrial gene for ND7 polypeptide.
Table 1. Genes for ribosomal proteins found in several organisms.

<table>
<thead>
<tr>
<th>Genome</th>
<th>ribosomal protein genes</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em></td>
<td>rpsA-rpsU</td>
</tr>
<tr>
<td></td>
<td>rplA-rplG, rplL-rplS, rplU-rplY, rpmA-rpmL</td>
</tr>
<tr>
<td>Mitochondria</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td></td>
</tr>
<tr>
<td>Yeast</td>
<td>var1</td>
</tr>
<tr>
<td>Neurospora</td>
<td>rps5</td>
</tr>
<tr>
<td>Angiosperm</td>
<td>rps1, rps3, rps7, rps10, rps12, rps13, rps14, rps19, rpl5, rpl16</td>
</tr>
<tr>
<td>Liverwort</td>
<td>rps1, rps2, rps3, rps4, rps7, rps8, rps10, rps11, rps12, rps13, rps14, rps19, rpl2, rpl5, rpl6, rpl16</td>
</tr>
<tr>
<td>Chloroplast</td>
<td></td>
</tr>
<tr>
<td>Liverwort</td>
<td>rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps15, rps18, rpl2, rpl14, rpl16, rpl20, rpl21, rpl22, rpl23, rpl33, rpl35, orf69</td>
</tr>
</tbody>
</table>

Materials and Methods

Analysis of nucleotide and amino acid sequences

Cloning and sequencing of the liverwort mitochondrial DNA were performed in this laboratory as described previously (Oda et al., 1992a; Oda et al., 1992b). The complete nucleotide sequence has been deposited in GenBank Data Library (accession number M68929). Computer aided analysis of nucleotide and amino acid sequences was carried out using the Hitachi DNASIS program on an NEC-9801 YM computer, and the IDEAS program on a FACOM M-780 computer (Data Processing Center, Kyoto University) using NBRF-PIR Release 25 database.

Isolation of mitochondria RNA from a cell culture of a liverwort, Marchantia polymorpha.

Cells of *M. polymorpha* were cultured in 1-M51C medium as described (Ohyama et al., 1988) on a gyratory shaker under continuous illumination. Liverwort mitochondrial RNA was isolated from 7 or 10-day-old suspension culture of cells. The cells were washed twice with 2% sucrose and suspended in homogenization buffer (0.4M mannitol, 2mM EDTA, 0.1M Hepes-KOH pH 7.5, 0.1% BSA (Fraction-V), 1mM β-mercaptoethanol, 0.6% polyvinylpolypyrrolidone and 1mM aluminon). After disrupting the cells by a French press, the cell homogenates were filtrated by Miracloth (Calbiochem Co.). Nuclei and chloroplasts were removed by two cycles of centrifugation at 1,000 x g for 5 min. Mitochondria in the supernatant were collected by centrifugation at 10,000 x g for 15 min, and washed by homogenization buffer without polyvinylpolypyrrolidone. Mitochondria were precipitated by centrifugation at 10,000 x g for 15 min and suspended gently in dilution buffer (0.4M mannitol, 2mM EDTA, 20mM Hepes-KOH pH 7.5, 0.1% BSA (Fraction-V), 1mM β-mercaptoethanol and 1mM aluminon). Mitochondrial suspension was layered on the top of Percoll stepwise gradient (17% and 28% Percoll in a solution containing 0.4M mannitol, 2mM EDTA, 20mM Hepes-KOH pH 7.5, 0.2% BSA (fatty acid free), 1mM β-mercaptoethanol and 1mM aluminon) and centrifuged at 13,500 rpm for 30 min in a Beckman SW28 rotor. The mitochondrial fraction was obtained from the interface between the two Percoll layers. To remove Percoll, 20 times the volume of the dilution buffer was added, then mitochondria were pelleted by centrifugation at 15,000 x g for 15 min. The mitochondrial pellet was resuspended in lysis buffer (50mM Tris-HCl pH 7.5, 20mM EDTA and 2% sarkosyl) and extracted with phenol, phenol/chloroform, and then with chloroform. After ether extraction and ethanol precipitation, mtRNA was precipitated 4 times in the presence of 2M lithium chloride (Ausubel et al., 1987). After then, the purified RNA (mtRNA) was precipitated with ethanol and dissolved in sterile water just before the use.

Northern Hybridization

RNA samples were denatured, loaded on 0.8% agarose gel containing 2.2M formaldehyde, 20mM MOPS pH 7.0, 5mM Sodium acetate, 1mM EDTA, and capillar-blotted onto Nylon membrane (Biodyne™ A, Pall, Tokyo). Hybridization was done at 45°C in a solution containing 6 x SSC, 0.1% SDS, 200 µg/ml calf thymus DNA, 1
x Denhardt's and 20% formamide. After hybridization, the membranes were washed successively in 6x SSC, 0.1% SDS at 42°C. Oligonucleotides were synthesized by automated DNA synthesizer (Applied Biosystems, USA) and have been designated according to the exon that they specify as follows. Oligonucleotides were labeled by [γ-32P]ATP (5,000 Ci/mmol, Amersham) using a MEGALABEL kit (TAKARA, Kyoto).

*rps*2: S'-GGCCTTTGGCACTAATGATAGATCCAATC-3' (Fig. 5A, 1)
*rps*4: S'-AGTTGCTTCTGAAACATAG-3' (Fig. 5A, 2)
*rps*12: S'-CGAACAAGCTGTATCGTGC-3' (Fig. 6A, 1)
*rps*7: S'-GAAAATACGATGGCTGCACG-3' (Fig. 6A, 2)
atp6: S'-GGGAAAAACGTTGTTCCACCG-3' (Fig. 6A, 3)
*ycb*1: S'-TGCTCCATGAAAGAGGATCCC-3' (Fig. 6A, 4)
*rps*10: S'-GCACGAGCTTCGCTTTGATTCTCAG-3' (Fig. 7A, 1)
rp12: S'-GCTTCGATCCACCTCCCGTGAAA-3' (Fig. 7A, 2)
rps19: S'-AGGGACATTCCATGAGCGCTC-3' (Fig. 7A, 3)
rps5: S'-GAGCTCGTGAATTCGATTAC-3' (Fig. 7A, 4)
rpl16: S'-GCTTATAGCAGCACGCGAGTGCTCCTCAA-3' (Fig. 7A, 5)
rpl5: S'-CATGATATGGCTCCTGAGTGC-3' (Fig. 7A, 6)
rps14: S'-CTTATATACGGAAACAGGGCGGCCATGC-3' (Fig. 7A, 7)
rps8: S'-CTTGGCATTGGGACATTCCACGCC-3' (Fig. 7A, 8)
*rpl6*: S'-GGGAGCTGCTGAAATGCATC-3' (Fig. 7A, 9)
rpl9: S'-AGGCATTTCTAGCTGCATGCTCAA-3' (Fig. 7A, 10)
*rpl15*: S'-AGGGACATTCCATGAGCGCTC-3' (Fig. 7A, 11)
rpl1: S'-GGAGTTTTTACGCTGTACATCCACC-3' (Fig. 7A, 12)

**Results and Discussion**

**Amino acid sequences of r-proteins encoded by liverwort mtDNA**

Amino acid sequences of r-protein genes detected in the liverwort mitochondrial genome were compared with their counterparts from *E. coli*, liverwort chloroplast, and the mitochondria of angiosperms (Fig. 1). The degree of sequence identity of the liverwort mitochondrial r-proteins with their homologues in other systems ranged from 23.7% to 62.1% (*E. coli*), 22.4% to 64.2% (liverwort chloroplast), and 43.6% to 80.8% (angiosperm mitochondria) (Table 2). The low values in liverwort mitochondria vs chloroplast amino acid sequence comparisons indicate that inter-organellar gene transfer does not occur between the liverwort chloroplast and mitochondrial genomes as observed in *Oenothera rps*4 gene (Schuster and Brennicke, 1987b). The mitochondrial RPS12 is encoded in the mitochondrial genome of not only the liverwort but also most higher plants investigated to date, while in *Oenothera* only small part of the reading frames is retained by the mitochondrial genome and a complete copy is encoded by the nuclear genome (Grohmann et al., 1992). This nuclear-encoded S12 and the liverwort mitochondrial S12 showed 79.2% identity.

**Table 2.** Amino acid sequence homology (%) of liverwort mitochondrial ribosomal proteins to those of *E. coli*, angiosperm mitochondria, and liverwort chloroplast, and of liverwort chloroplast to that of *E. coli*.

<table>
<thead>
<tr>
<th>Protein</th>
<th><em>E. coli</em></th>
<th>Angiosperm mt</th>
<th>Liverwort cp</th>
<th>Liverwort cp/E. coli</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>23.7</td>
<td>43.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>27.1</td>
<td>47.9</td>
<td>22.8</td>
<td>44.3</td>
</tr>
<tr>
<td>3</td>
<td>25.4</td>
<td>46.9-50.4</td>
<td>24.1</td>
<td>40.6</td>
</tr>
<tr>
<td>4</td>
<td>25.0</td>
<td>58.5</td>
<td>22.4</td>
<td>40.1</td>
</tr>
<tr>
<td>7</td>
<td>35.8</td>
<td>58.5</td>
<td>29.9</td>
<td>43.8</td>
</tr>
<tr>
<td>8</td>
<td>35.1</td>
<td>58.3</td>
<td>26.7</td>
<td>45.5</td>
</tr>
<tr>
<td>10</td>
<td>31.5</td>
<td>58.3</td>
<td>48.8</td>
<td>51.5</td>
</tr>
<tr>
<td>11</td>
<td>48.0</td>
<td>80.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>62.1</td>
<td>80.8</td>
<td>64.2</td>
<td>78.2</td>
</tr>
<tr>
<td>13</td>
<td>38.3</td>
<td>64.6-70.4</td>
<td>38.4</td>
<td>45.0</td>
</tr>
<tr>
<td>14</td>
<td>43.3</td>
<td>55.0-61.2</td>
<td>42.9</td>
<td>63.0</td>
</tr>
<tr>
<td>19</td>
<td>41.7</td>
<td>54.1</td>
<td>42.9</td>
<td>63.0</td>
</tr>
<tr>
<td>L2</td>
<td>44.8</td>
<td>43.2</td>
<td>48.4</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>28.6</td>
<td>60.9-68.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>36.6</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>50.4</td>
<td>71.9-78.5</td>
<td>45.9</td>
<td>53.8</td>
</tr>
</tbody>
</table>
Fig. 1. (cont.)

---

Fig. 1. (cont.)
Fig. 1. Amino acid alignments of r-proteins deduced from the liverwort mitochondrial DNA sequence, with counterparts from *E. coli*, the liverwort chloroplasts, and angiosperm mitochondria. Amino acids are denoted by their 1-letter symbols. Numbers at the ends of sequences indicate the numbers of amino acid residues. Identical amino acids are designated by a colon. Dashes are assumed with counterparts from £.

Organization of liverwort mitochondrial r-protein genes

Most of the genes for r-proteins in the liverwort mitochondrial genome were organized into a cluster (rps10-rpl2-rps19-rps3-rpl6-rps5-rps14-rps8-rpl6-rps3-rps11-rps1) similar to that seen in *E. coli* t-protein operons S10 (S10-L3-L4-L23-L2-S19-L22-S3-L16-L29-S17) (Zurawski and Zurawski, 1985), *s*cp (L14-L24-L5-S14-S8-L6-L18-S5-L30-L15-secY-X) (Cerretti et al., 1983) and α (S13-S11-S4-ropA-L17) (Bedwell et al., 1985). An additional cluster (rps12-rps7) had the same order as the homologous genes in the *E. coli* str operon (S12-S7-fus) (Post and Nomura, 1980). Genes for rps4 and rps2 were located elsewhere in the liverwort mitochondrial genome. A large cluster of r-protein genes has not been found in the
mitochondrial genomes of the other organisms, whereas a very similar clustered organization of r-protein genes exists in chloroplast genomes (Ohyama et al., 1986; Shinozaki et al., 1986; Hiratsuka et al., 1989). The organization of the r-protein gene clusters in liverwort mtDNA was compared with those of the liverwort chloroplast and E. coli genomes (Fig. 2). Several r-protein genes that are present in E. coli operons were not found in the liverwort mitochondrial genome, whereas the rps1 gene (which is not located in the E. coli S10-spα or str operons) was found in the liverwort cluster. Nevertheless, organization and order of respective genes were very similar in these three genomes. This finding strongly supports the endosymbiotic hypothesis, which postulated that the organelles of eukaryotes originated from prokaryotic (specifically eubacterial) ancestors in evolution (Gray et al., 1989).

In the mitochondrial genomes of angiosperms, ten genes for r-proteins are identified so far. Differed from the case of the liverwort mitochondria, genes for rps12, rps13 and rps14 are closely linked to non-ribosomal protein genes (Fig. 3) (Gualberto et al., 1988; Suzuki et al., 1991; Bland et al., 1986; Schuster and Brennicke 1987a; Bonen, 1987; Wissing, et al., 1990; Wahleithner and Wolstenholme, 1988; Schuster et al., 1990a). For example, ntdS and rps12 genes are co-transcribed in the wheat, maize and rice mitochondrial genomes (Gualberto et al., 1988; Suzuki et al., 1991). On the other hand, maize mitochondrial rps3 and rpl16 genes are not only closely linked but even overlap, as did the liverwort rps3 and rpl16 genes. Such rps3-rpl16 gene clusters are also found in the mitochondria genomes of Brassica, Oenothera, and Petunia; although the Oenothera rps3 and Petunia rpl16 genes are thought to be pseudo genes. Especially in Brassica (Ye et al., 1993), more two r-protein genes, rpl5 and rps14, are clustered and their organization is similar to those within the E. coli S10-spα operon and the liverwort mitochondrial cluster. In Oenothera and Petunia, rps19 genes are also found in the clusters (rps19-rps3-rps16) which have quite similar organizations to those of the liverwort mitochondrial genome and E. coli S10 operon (Bock et al., 1994; Sutton et al., 1993).

While the liverwort rps3 gene analyzed here shows a classic ATG codon, no ATG is encoded in the rpl16 gene at the beginning of the sequence similarity with other organisms. There is a termination codon (TAA) at 24 bp upstream of GTG (valine) at the beginning of the homology (Fig. 4). This finding raises the possibility that the GUG (valine) codon rather than an internal ATG is used for translation initiation of the rpl16 gene in the liverwort. The maize (Hunt and Newton, 1991) and Brassica (Ye et al., 1993) rpl16 genes encode three and two in-frame ATG codons further upstream that could also serve as initiation codons, respectively. On the other hand, in the case of Oenothera (Bock et al., 1994), a termination codon (TAA) is found at 9 bp upstream and a in-frame ATG codon is absent. Since GTG at that position is conserved in the other plants maize, Oenothera and Brassica, this GTG codon is also considered as translation initiation codon in the liverwort.
Fig. 3. Organization of r-protein genes in angiosperm mitochondrial genomes. Solid boxes and open boxes indicated r-protein genes and non r-protein genes, respectively. Hatched boxes are genes having introns.

Fig. 4. Overlapping regions of rps3 and rpl16 genes in liverwort and maize mitochondrial genomes. Amino acids are denoted by their 1-letter symbols. Black boxes indicate initiation codons (GUG) in both liverwort and maize rpl16 genes. Three inframe ATG codons in maize are boxed. Termination codon (TAA) 24 bp upstream from the initiation codon in liverwort rpl16 gene is underlined.

The liverwort rpl2 gene has one group II intron of 775 bp, while the counterparts of higher plants have no introns. The rps14 gene is also interrupted by a 896 bp group II intron in the liverwort mitochondria, though no introns have been reported for those of angiosperms. In contrast, the rps3 genes of higher plants studied so far contain one intron and the insertion sites of these introns are the same among them (Hunt and Newton, 1991; Ye et al., 1993; Sutton et al., 1993; Bock et al., 1994). There may be two possibilities for such differences of the intron contents. One possibility is that these introns had been present in the last common ancestor of bryophytes and angiosperms and then were lost from one or the other. The other is that each gene has acquired introns independently after the divergence of them.

Overall, the organization of r-protein genes is much different in liverwort and angiosperms mtDNAs. Indications are that r-protein gene organization has undergone drastic changes in the mitochondrial genome of angiosperms in the course of evolution, probably as a result of recombination events (Palmer and Shields,
as well as gene transfer into nuclear DNA (Stern and Lonsdale, 1982). On the other hand, there is apparently no homologous recombination through directly repeated sequences in the liverwort mitochondrial genome, suggesting that this genome retains the primitive form (Oda et al., 1992a). It is possible that the mitochondrial genomes of angiosperms do not encode as many r-protein genes as the liverwort mitochondrial genome, in spite of the much larger average size of the former.

**Inferred characteristics of liverwort mitochondrial r-proteins**

Interestingly, whereas liverwort chloroplast r-proteins were similar in size to their *E. coli* counterparts, liverwort mitochondrial r-proteins L2, S3, S7 and S8 were larger than their counterparts in *E. coli* (Fig. 1m, 1c, 1e, and 1f, respectively). Moreover, r-protein S3 in angiosperm mitochondria appeared to be much larger than that in its liverwort mitochondrial homologue (Table 3.) (Hunt and Newton, 1991; Ye et al., 1993; Sutton et al., 1993). However, liverwort and angiosperm S3 amino acid sequences deduced from the corresponding mtDNA sequences showed a high degree of similarity with each other in the N-terminal and C-terminal regions (Fig. 1c). On the other hand, the wheat S7 protein is much smaller than its counterparts from liverwort mitochondria and slightly shorter than those of *E. coli* and liverwort chloroplast. In the case of yeast mitochondrial r-protein (L8) (encoded by nuclear genome), the N-terminal region is homologous to *E. coli* r-protein L17 while the C-terminal region shows similarity to that of *E. coli* S13 r-protein (Kitakawa et al., 1990). It has been postulated that the yeast L8 protein gene might have arisen as a result of fusion of genes for L17 and S13 proteins (Kitakawa et al., 1990). However, the extra portions of liverwort mitochondrial L2, S3, S7, and S8 proteins showed no similarity to any other known r-proteins. Therefore, it is possible that their genes may be products of fusion with genes for uncharacterized r-proteins, or they may simply be unusually large as a consequence of insertions. In either case, extra segments of the proteins may be removed by post-translational processing during the assembly of ribosome particles.

**Table 3. Sizes of ribosomal proteins from liverwort mitochondria. *E. coli*, angiosperm mitochondria, and liverwort chloroplast genomes.**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Liverwort mt</th>
<th>E. coli</th>
<th>Angiosperm mt</th>
<th>Liverwort cp</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2</td>
<td>501</td>
<td>273</td>
<td>93</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>188</td>
<td>179</td>
<td>185-192</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>101</td>
<td>177</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>135</td>
<td>136</td>
<td>144-185</td>
<td>143</td>
</tr>
</tbody>
</table>

Numbers indicate amino acid residues.

Ribosomal proteins S1 and L6 in liverwort mitochondria appeared to be smaller than their counterparts in *E. coli*, lacking the C and N terminal portions of *E. coli* S1 and L6 r-proteins (Fig. 1a and 1o, respectively). Wheat S1 is much smaller than the liverwort S1 and is only about one-third the size of the *E. coli* counterparts (Gonzalez et al., 1993). The missing portions of these proteins may not play an important role in ribosome assembly and function. However, the presence of "extra" and "missing" portions of liverwort mitochondrial r-proteins must remain an inference until direct sequencing of the mitochondrial r-proteins themselves has been performed.

**Evolutionary events of organelle gene transfer into the nuclear genome**

It has been shown that ribosomes in *E. coli* contain over 50 distinct r-proteins.
Genes for 16 and 20 species of r-proteins have now been detected in the liverwort mitochondrial and chloroplast genomes, respectively. The remainder are assumed to be encoded by the nuclear genome. It is of interest that 11 genes (rps12, rps7, rpl2, rps19, rps3, rps16, rps14, rps8, rpl11, rps4, and rps2) were found to be encoded by both organelle genomes. Similarly homologous genes are known to exist in chloroplast and mitochondrial genomes for subunits of NADH dehydrogenase (nad genes in mitochondria, ndh genes in chloroplast) and ATP synthase (atp genes) (Ohyama et al., 1991). It is unlikely that such common genes are maintained in the two organelle genomes by chance. In the plant kingdom, endosymbiosis of a chloroplast ancestor is thought to have followed that of a mitochondrial ancestor. Thus many genes of the mitochondrial genome must already have been transferred into the nuclear genome by the time of the endosymbiotic event that gave rise to the chloroplast ancestor. Since then, additional migration of both chloroplast and mitochondrial genes to nuclear genome is presumed to have taken place. It is conceivable that there may have been duplication of mitochondrial genes already encoded by the nuclear genome at the time the chloroplast genome was being established, with one copy subsequently acquiring the signal peptide sequence necessary to transport the encoded r-protein into the chloroplast. In that case, the homologous chloroplast gene could simply have been lost, rather than being transferred to the nucleus. However, this is only a speculation for the present.

Transcriptions of the rps2 and rps4 genes in liverwort mitochondria

RNA blot analysis was carried out using oligonucleotide probes specific to either rps2 or rps4 as shown in Fig. 5A. In the case of rps2, multiple transcripts of 6.0 kb, 3.5 kb, 3.0 kb and 1.8 kb were detected (Fig. 5B, lane 1). Three smaller RNA species, 3.5 kb, 3.0 kb and 1.8 kb, hybridized with the many other ribosomal protein sequences (see below). They are identical in size to either 18S rRNA or 26S rRNA and thus were most likely the rRNAs, which were identified due to spurious cross-hybridization of either rps2 or the other r-protein gene sequences. Nevertheless, at least the 6.0 kb transcript was specific for rps2. This transcript could contain the...
coding region of \textit{rps2} (714 bp) and was too large to cover those of \textit{orf277} and \textit{orf228} located approximately 0.3 kb and 2.4 kb downstream from \textit{rps2}, respectively. However, Northern analysis using the probes specific to both \textit{orfs} revealed that there were no signals (data not shown), suggesting that these \textit{orfs} may not be transcribed or may be transcribed at very low levels, and that \textit{rps2} is transcribed independently of them.

When probed with an oligonucleotide complementary to \textit{rps4}, two major bands of 2.4 kb and 1.4 kb were observed (Fig. 5B, lane 2). Both transcripts are able to cover its coding region (591 bp), but could not contain \textit{rrn18} located about 4.6 kb downstream from \textit{rps4}. In addition, such signals as 2.4 kb and 1.4 kb were not detected using a probe specific for \textit{trnG} located 30 bp upstream of \textit{rps4} (data not shown). These findings indicate that \textit{rps4} is transcribed individually. The heterogeneity of transcript size might be resulted from RNA processing and/or multiple transcription initiation sites as reported in maize mitochondria (Mulligan \textit{et al.}, 1988a; Mulligan \textit{et al.}, 1988b).

\textbf{Co-transcription of the \textit{rps12} and \textit{rps7} in liverwort mitochondria}

To study the expression of the \textit{rps12} and \textit{rps7} which were organized into a cluster, oligonucleotide probes specific for them were prepared (Fig. 6A). Northern blot analysis showed these genes to be transcribed in liverwort mitochondria (Fig. 6B). A large transcript of about 7.0 kb was detected with probes from both of them. This indicated that \textit{rps12} and \textit{rps7} were co-transcribed in a primary transcript of 7.0 kb. Three smaller RNA, 3.5 kb, 3.0 kb and 1.8 kb hybridized with \textit{rps12} probe, but they were probably tRNAs which cross-hybridized fortuitously as described above (Fig. 6B, lane 1). On the other hand, one more major band of 3.0 kb was found with a probe for \textit{rps7} (Fig. 6B, lane 2). The 7.0 kb transcripts were much larger than the coding regions of \textit{rps12} and \textit{rps7} (1,070 bp in total), suggesting that they were co-transcribed with additional genes downstream and/or upstream. To examine this possibility, Northern blot analysis using probes complementary to \textit{atp6} and \textit{ycob} which were located 3.1 kb and 2.4 kb downstream from \textit{rps7}, respectively, were carried out. Four bands of 7.0 kb, 3.0 kb, 2.2 kb and 1.8 kb were detected in the case of \textit{atp6} (Fig. 6B, lane 3), while no signal was observed in the case of \textit{ycob} (Fig. 6B, lane 4). These results suggested that \textit{rps12}, \textit{rps7} and \textit{atp6} but not \textit{ycob}
were transcribed in a single primary transcript of 7.0 kb and that the 3.0 kb transcript probably contained only rps7 and atp6 genes. It has not been clear whether this 3.0 kb RNA molecule was produced by processing of the 7.0 kb transcript or was transcribed from the region upstream of rps7 gene independently of rps12. The rps12-rps7 region of liverwort mitochondria possibly functions as a ribosomal protein gene operon like the str operon of E. coli.

Transcription analysis of the twelve genes organized into a large ribosomal protein gene cluster

Next, expressions of the twelve genes forming the large ribosomal protein gene cluster (rps10-rpl2-rps19-rps3-rpl16-rpl5-rps14-rpl6-rps13-rps11-rps1) were analyzed. Using oligonucleotide probes complementary to each gene as shown in Fig. 7A, Northern hybridizations were performed (Fig. 7B). In several cases, three bands of 3.5 kb, 3.0 kb and 1.8 kb were detected, but they were supposed to be resulted from cross-hybridizations with rRNAs as mentioned above. Four probes complementary to rpl2, rps19, rps3 and rpl16 mainly hybridized a common 7.3 kb transcript (Fig. 7B, lanes 2 to 5), while one major band of about 9.6 kb was found when probed for rps10 (Fig. 7B, lane 1). In contrast, no major signal was detected when hybridized with each probe specific for rpl5, rps14, rps8, rpl6, rps13, rps11 and rps1 (Fig. 7B, lanes 6 to 12). The distance between 3' end of rps10 and 5' end of rpl5 is 7.2 kb, therefore it is likely that the 7.3 kb transcript initiates from the region between rps10 and rpl2 and terminates between rpl16 and rpl5. These results suggest that only four genes, rpl2, rps19, rps3 and rpl16 are co-transcribed and expressed as a single transcription unit similar to the S10 operon of E. coli. On the other hand, it is supposed that rpl5, rps14, rps8, rpl6, rps13, rps11 and rps1 are not transcribed or are transcribed at very low levels. When probed for ymad7 located at 1.3 kb upstream of rps10, the 9.6 kb transcript was also found, demonstrating that ymad7 and rps10 were probably co-transcribed (see also Chapter III).
Chapter II
Genes for NADH dehydrogenase subunits in liverwort mitochondrial genome

Introduction

NADH dehydrogenase, which is also called as NADH:ubiquinone oxidoreductase (EC 1.6.99.3), is the respiratory chain complex I. This is the largest complex of respiratory enzyme complexes in mitochondrial inner membrane and the first enzyme in respiratory chain that accepts electrons from NADH and transfers them to ubiquinone. It consists of approximately 30-40 subunits, and also contains one FMN (flavin mononucleotide) and iron-sulfur clusters as redox groups (Weiss et al., 1991; Walker, 1992). In mammals, seven subunits of this enzyme are encoded by mitochondrial genomes and synthesized in mitochondria. These mitochondrial genes for the subunits are identified and designated as ND1-4, ND4L, and ND5-6 (Chomyn et al., 1985; Chomyn et al., 1986) (Table 1.). Podospora anserina mitochondrial genome also contains seven genes for subunits of the enzyme complex, ND1-4, ND4L, ND5-6 (Cummings et al., 1990). On the other hand, the mitochondrial genome of yeast has not been reported to have a gene for any subunit of NADH dehydrogenase. It is assumed that genes for other subunits of this enzyme are nuclear-encoded and that their gene products are imported from cytoplasm into mitochondria.

Nine kinds of genes for subunits of this enzyme, nad1 (Bland et al., 1986; Stern et al., 1986; Wahlleithner et al., 1990; Wissinger et al., 1991; Chapdelaine and Bonen, 1991), nad2 (Xue et al., 1990), nad3 (Gualberto et al., 1988; Rasmussen and Hanson, 1989; Schuster et al., 1990b; Suzuki et al., 1991; Kim et al., 1991), nad4 (Lamattina et al., 1989; Wintz et al., 1989; Lamattina and Gregor, 1991; Gass et al., 1992; Geiss et al., 1994), nad4L (Brandt et al., 1992), nad5 (Wissinger et al., 1988; Ecke et al., 1990; de Souza et al., 1991; Knoop et al., 1991; Park and Breitenberger, unpublished data in Genbank, accession no. M57478), nad6
(Haouazine et al., 1992; Nugent and Palmer, 1993). *nad7* (Bonen et al., 1994; Herz et al., 1994; Gälber et al., 1994), *nad9* (Kubo et al., 1993; Lamattine et al., 1993; Grohmann et al., 1994) have been identified so far on mitochondrial genomes from higher plants.

Seven coding regions whose deduced amino acid sequences have significant similarities with those of subunits of human mitochondrial respiratory NADH dehydrogenase have been found in the chloroplast genomes of *Marchantia polymorpha* (Ohyama et al., 1986) and *Nicotiana tabacum* (Shinozaki et al., 1986) and named *ndh* genes (*ndh1-4, ndh4L, and ndh5-6*) in *M. polymorpha*, or *ndhA-ndhG* in *N. tabacum*, respectively. In addition, three gene, *ndhH, ndhl* and *ndhJ* (previously named ORF392, *frxB* and ORF169 in liverwort chloroplast genome, respectively) have been identified on chloroplast genomes by comparing bovine nuclear-encoded subunits of this enzyme (Dupuis et al., 1991).

Recently a complete nucleotide sequence of the mitochondrial genome from a liverwort, *Marchantia polymorpha*, and deduced its gene organization have been determined in this laboratory (Oda et al., 1992a; Oda et al., 1992c). In this chapter, the author describe detailed characterization of the eight subunits of NADH dehydrogenase encoded by liverwort mitochondrial genes, *nad1, nad2, nad3, nad4, nad4L, nad5, nad6*, and *nad9* and also showed transcriptions of these genes.

### Materials and Methods

#### Analysis of nucleotide and amino acid sequences

Computer analysis of nucleotide and amino acid sequences was carried out as described in Chapter I.

#### Isolation of mitochondria RNA from and Northern Hybridization

The liverwort mitochondrial RNA was isolated by the methods as described in Chapter I. For Northern hybridization, a 664 bp *BglII-PstI* restriction fragment was isolated from a plasmid pLB104 (Oda et al., 1992c) and labeled by [α-32P]dCTP (3,000 Ci/mmol, Amersham) using Random Primer DNA labeling kit (Boehringer Mannheim). The following oligonucleotides were also used as probes specific for each regions.

* nad1: 5'GATCATAACGATATCGTGAAATGCTGCGC-3' (Fig. 6A, 1)*
* nad5: 5'CATTCTAGATACGCAAAATTTCCTCTGATAGCC-3' (Fig. 6A, 2)*
* nad4L: 5'AATAGCGGATTCCGCAGCAGCCACCGTTAA-3' (Fig. 6A, 3)*
* nad6: 5'TATGGTAGATAGGTGACTCATCTAATTTTCGTGTTGTAAT-3' (Fig. 6A, 4)*
* nad9: 5'ACCAACATATCAGATTTTCGTTTC-3' (Fig. 6A, 5)*
* nad5 exon1: 5'GCCCAAGCAGCTCCGGAATAATAGACCTCCGGG-3' (Fig. 7A, 1)*
* spacer between: 5'GAGGGGATGTGCGTTAAATAGACCTCCGGGGG-3' (Fig. 7A, 2)*
* nad5 and nad4: 5'TCTGTAGACCCGGTGTGTATTGTGGGCGAAT-3' (Fig. 7A, 3)*
* nad4 intron: 5'GGTTTCAATCTAATTTGCGCAGGGCCGCGTGG-3' (Fig. 7A, 4)*
* nad4 exon2: 5'TGAGCGATTCCCTCCATAGAGGACATCCACC-3' (Fig. 7A, 5)*
* nad2 intron: 5'GCCGATCCCGCGCTTAGTCGACTCCGCAGC-3' (Fig. 7A, 7)*
* nad2 exon2: 5'CGCCCGATGCTCCTAAGATCATAGAAGCAAAGCA-3' (Fig. 7A, 8)*

### Results and Discussion

#### Organization of nad genes in the liverwort mitochondrial genome

The coding regions for the *nad* genes of the liverwort mitochondrial genome were predicted by comparing them with amino acid sequences of known *nad* genes from other organisms, and their exon-intron junctions were assumed by the conserved
secondary structures of their introns. All of the liverwort nad genes, including pseudo-nad7, were located on the same DNA strand (Oda et al., 1992a). The liverwort nad1, nad3, nad4L, nad6 and nad9 genes are scattered throughout the genome while the liverwort nad5, nad4, and nad2 genes form a cluster (Oda et al., 1992a; Nozato et al., 1993). In the case of the liverwort chloroplasts, ndh7 (previously assigned ORF392), ndh1, ndh8 (previously assigned frx5), ndh6, ndh4L, and ndh4 form a cluster in the small single copy (SSC) region (Kohchi et al., 1988), and ndh3, ndh9 (previously named as ORF169), and ndh10 (previously assigned psbG) in the large single copy (LSC) region (Umesono et al., 1988). All but ndh2 are on the same strand of the chloroplast DNA. Therefore, the organization of the mitochondrial nad and the chloroplast ndh genes is not similar. This indicates the independent gene rearrangements in the individual organelle genomes during the evolution.

Among the liverwort mitochondrial nad genes, nad2, nad4, and nad5, are tandemly clustered as in the following order: nad5 - 1334 bp spacer - nad4 - 27 bp spacer - nad2. This closely related gene arrangement suggests co-transcriptional expression of them. On the other hand, these three genes of the angiosperms analyzed to date, are not found to be clustered. In contrast, it is reported that the nad1 exon d and nad6 gene are closely linked and co-transcribed in wheat mitochondria (Haouazine et al., 1993).

Characterization of the liverwort nad genes

The liverwort nad1 gene has no intron, though the nad1 genes of wheat, Oenothera, and Petunia have four group II introns in their coding regions (Fig. 1A and Table 2.). All of the introns in higher plant nad1 genes are inserted at identical sites (Fig. 2a). The mitochondrial genome of Podospora anserina has a gene equivalent to the nad1, i.e., ND1, which is also split by the four group I introns. These facts indicate the possibility that the ancestors of plant and fungi originally had group I introns in their nad1 genes, but the plant species have lost them all after the divergence of plants from fungi; the angiosperms then having acquired the group II introns since divergence from the bryophytes (Ohta et al., unpublished data).

The nad2 gene in the liverwort mitochondria is interrupted by 1,418 bp group II intron. It is also reported that the nad2 gene in Oenothera mitochondria have four group II introns (Binder et al., 1992) (Fig. 1B and Table 2.). Especially, the liverwort nad2 intron was inserted at the identical site to the Oenothera nad2 intron c/d and showed sequence similarity with that. Liverwort chloroplast ndh2 is also split by a group II intron of 536 bp and specifies 501 amino acid residues (Ohyama et al., 1986). Though the amino acid sequence of liverwort mitochondrial NAD2 shows significant similarity (32.9% amino acid identity) with liverwort chloroplast ndh2 gene product, the insertion site of the intron is not conserved, suggesting that the insertional event of introns into the chloroplast ndh2 and the mitochondrial nad2 would have occurred at least after divergence of a prototype of a gene for NADH dehydrogenase into chloroplastic and mitochondrial genes.

The liverwort nad3 gene is interrupted by a 1,485 bp group II intron although those of Oenothera (Schruster et al., 1990), Petunia (Rasmussen and Hanson, 1989), maize (Gualberto et al., 1988), wheat (Gualberto et al., 1988; Gualberto et al., 1989), and liverwort chloroplasts (Kohchi et al., 1988) do not have any introns (Fig. 1C and Table 2.).

Liverwort mitochondrial nad4 contains one group II intron of 899 bp, although nad4 genes of angiosperm mitochondria have up to three introns (Fig. 1D and Table 2.). The insertion sites of introns in nad4 genes from higher plant mitochondria are conserved among them, but different from those of the liverwort. This suggests that the origin of introns in liverwort mitochondrial nad4 genes would be different from those in higher plant mitochondrial genomes. Interestingly, the liverwort nad4 intron has partly sequence similarity with introns in the liverwort nad4L, rpl2 and pseudo-nad7, suggesting that these introns have been evolved from a common ancestor and that these introns have moved in the liverwort mitochondrial genome in the course of evolution (Ohta et al., unpublished data).

The liverwort nad4L gene has two group II introns, 1,720 bp and 1,151 bp in size. On the other hand, no introns were found in nad4L genes of angiosperms mitochondria so far (Fig. 1E and Table 2.).
Fig. 1. Gene structures of *nad* genes in liverwort and higher plants mitochondria. The exons of the *nad* genes are indicated by solid boxes. A; *nad1*, B; *nad2*, C; *nad3*, D; *nad4*, E; *nad4L*, F; *nad5*, G; *nad6*, H; *nad9*. 

Fig. 1 (Cont.)
Table 2. Numbers of introns in nad genes

<table>
<thead>
<tr>
<th>Genes for subunits</th>
<th>nad1</th>
<th>nad2</th>
<th>nad3</th>
<th>nad4</th>
<th>nad5</th>
<th>nad6</th>
<th>nad7</th>
<th>nad8</th>
<th>nad9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mitochondria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liverwort</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>(2)</td>
<td>0</td>
</tr>
<tr>
<td>Angiosperms</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>1-3</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Human</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yeast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Podospora</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Chloroplast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liverwort</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The mitochondrial nad5 gene of liverwort consist of two exons separated by a 672 bp group I intron, while those of higher plants have four introns, two cis-spliced and two trans-spliced which are inserted at almost identical sites among these plants (Fig. 1F and Table 2). Like the situation of the nad1 gene, the insertion site of the liverwort nad1 intron was different from those of angiosperms.

There is no intron in the nad6 and nad9 genes of the liverwort mitochondria, nor are introns present in angiosperm mitochondria (Fig. 1G and 1H, respectively and Table 2).

Amino acid sequence comparison of NADH dehydrogenase subunits encoded by liverwort mitochondrial DNA

Amino acid sequences of the eight liverwort mitochondrial nad gene products were compared with their counterparts from higher plant mitochondria, Podospora mitochondria, human mitochondria, and liverwort chloroplast ndh gene products (Fig. 2) and amino acid homologies (%) between them are summarized in Table 3.

The product of the liverwort nad1 gene is of a polypeptide of 328 amino acid residues. An alignment of the deduced amino acid sequences of the NADH dehydrogenase subunit 1 genes from several organisms are given in Fig. 2a. Full nucleotide sequences of the nad1 genes are available from wheat, Oenothera, and Petunia. These mRNAs undergo trans-splicing and RNA editing. Partial nucleotide sequences of the nad1 gene have also been determined for tobacco (Blanc et al., 1986), maize (Blanc et al., 1986), watermelon (Stern et al., 1986), and broad bean (Wahleithner et al., 1990). The existence of sequences homologous to those of the nad1 genes have been found in the mitochondria of spinach (Stern and Palmer, 1986), sunflower (Siculella and Palmer, 1988), several species of Brassica (Makaroff and Palmer, 1987; Palmer and Herbon, 1988), sugarbeet (Beers and Lonsdale, 1988), and rice (Yamato et al., 1992). The amino acid sequences of the mitochondrial nad1 gene products are highly (>80%) conserved between a liverwort and other plants.

The subunit 2 of NADH dehydrogenase (NAD2) from liverwort mitochondria which is encoded by the nad2 gene is a polypeptide of 489 amino acid residues. This protein shows significant levels of amino acid sequence identities with counterparts from the other organisms as shown in Fig. 2b and Table 3. Liverwort mitochondrial NAD2 is 142-amino acids longer than that of the human mitochondrial ND2 of 347 amino acids. This difference in length is mainly due to an additional stretch of amino-terminal amino acid sequences in the liverwort mitochondrial NAD2 (Fig. 2b). However, this amino-terminal region missing in human ND2 product has significant sequence similarity with its counterpart of the liverwort chloroplast NDH2.

The number of amino acid residues in the putative product of the liverwort mitochondrial nad3 gene is 118, as same as those of wheat, Oenothera and Petunia mitochondria. The alignments of the amino acid sequences of the nad3 homologues are shown in Fig. 2c. The relatively low homology between liverwort and Petunia mitochondria is probably due to the amino acid sequence being deduced from the DNA sequence of Petunia whereas that derived from the edited RNA sequence is given for wheat and Oenothera. This indicates the possible RNA editing in the Petunia mitochondrial transcripts.

The subunit 4 of NADH dehydrogenase (NAD4) from liverwort mitochondria which is encoded by the nad4 gene is a polypeptide of 495 amino acid residues.
This protein shows a high level of amino acid sequence identity with the counterparts from wheat, *Brassica* and lettuce mitochondria, and also significant levels of sequence similarity with those from another organisms as shown in Fig. 2d and Table 3.

The liverwort *nad4L* gene putative product is a polypeptide of 100 amino acid residues. The liverwort *nad4L* gene product has a 36.7% homology with the corresponding product of the liverwort chloroplasts (Fig. 2e), and is a 52.8% homology to that of the ND4L gene of *P. anserina* mitochondria. The nucleotide sequence of the *nad4L* gene from Arabidopsis is only available as higher plants, and its gene product shows 85.0% similarity to that of the liverwort.

The subunit 5 of NADH dehydrogenase (NAD5) from liverwort mitochondria which is encoded by the *nad5* gene is a polypeptide of 669 amino acid residues. This subunit shows a high level of sequence identity with the counterparts from the other plant species, *Oenothera*, wheat, *Arabidopsis*, and also significant levels of sequence similarity with counterparts from *Podospora* mitochondria, human mitochondria, and with liverwort chloroplast *ndh5* gene product as shown in Fig. 2f and Table 3.

The liverwort *nad6* gene product would be a polypeptide of 199 amino acid residues. The deduced amino acid sequences of the *nad6* genes from several organisms are aligned as shown in Fig. 1g. The liverwort *nad6* gene product shares a homology of 15.2% with the human ND6, a value is much lower than those of the other *nad* genes. Recently, wheat and *Brassica* mitochondrial *nad6* genes have been isolated and sequenced (Haouazine et al., 1992; Nugent and Palmer, 1993). As compared the amino acid sequences of the liverwort *nad6* gene with those of them, high sequence similarities (75.3% in both cases) were found.

It was recently reported that wheat, sugar beet, and potato mitochondria encode homologues of the nuclear-encoded 30 kDa subunit of bovine mitochondrial complex I (Lamattina et al., 1993; Kubo et al., 1993; Grohman et al., 1994, respectively). These genes also showed similarities with the liverwort chloroplast gene *ndh7* (Ohyama et al., 1986) and were named as *nad9*. The liverwort mitochondrial genome contains an ORF (*orf212*) sharing sequence homology with these *nad9* genes, so this ORF

![Fig. 2 (Cont.)](image-url)
Fig. 2. Amino acid sequence comparison of liverwort mitochondrial nad 9 gene products with their counterparts from other organisms. Identical amino acid residues with liverwort mitochondrial nad 9 gene products are marked with colons. Bars indicate artificial shifting to maximized homology and absence of corresponding amino acid residues. Solid arrow heads show positions of insertion sites of introns in their genes. The results of RNA editing were introduced into the sequences and the edited sites reported in higher plant mitochondria to date are underlined. Length of gene products and amino acid residues is not known. Taking together with the fact that these three nad genes are tandemly clustered, this finding suggests that these three genes may have evolved from a common ancestor in the course of the evolution through gene rearrangement in mitochondrial genome.

Repetitive sequences of nad genes in the liverwort mitochondrial genome

The repeated sequences of higher plant mitochondrial genome play a role in the recombination of mitochondrial DNA (reviewed in Lonsdale, 1989). Some of the liverwort repeated sequences contain parts of the nad genes (Nozato et al., 1993). The S'-half portion of the nad6 gene is duplicated on the opposite strand (from position 153,438 to 153,109, Oda et al., 1992c) and the length of the duplicated...
sequence is 336 bp, 299 bp of which is identical (89.0% identity) to the original copy of the nad6 gene. The alignment of these two sequences is shown in Fig. 4A. Curiously, in spite of several deletions and substitutions in the nucleotide sequence, a reading frame can be deduced in the sequence of the duplicated nad6 fragment. This ORF starts at the same initiation codon and extends to the boundary of the duplicated region, resulting in the formation of orf100 with some alterations relative to the original copy. Incidentally, this orf100 is located between a 18S ribosomal RNA gene (rrn18) and an initiator tRNA gene (trnM-CAU), and these three genes are on the same DNA strand, although the possibility of its expression and its function are unknown (Fig. 4B).

Numerous abnormal ORFs, which consist of DNA fragments derived from the common genes and from unknown sources, reside on the mitochondrial genome of higher plants. They have been studied vigorously in the connection with cytoplasmic male sterility and have been found to be expressed (reviewed in Lonsdale, 1989).
is of interest whether the liverwort orf100, which contains the altered 5'-half portion of the nad6 gene is or is not transcribed. The hydrophobic nature of the orf100 product suggests that it could be integrated into a mitochondrial membrane.

In addition, several parts of the nad5-nad4-nad2 gene cluster were duplicated in the liverwort mitochondrial genome. (i) 3'-terminal region of the second exon of nad2 (364 bp) and its following non-coding region of 187 bp (total 551 bp) were directly repeated at 673 bp upstream from nad3 which is approximately 27 kb downstream from nad2 on the same strand, although 11 bases were altered and one base was deleted in the repeated region. (ii) A part of the first exon of nad5, which is 172 bp long, starting from the second nucleotide of the ATG translation initiation codon, was repeated at 58 bp upstream from a gene for threonyl tRNA with the anticodon GGU and which is located at a distance of approximately 50 kb on the opposite strand. This region had 87.2% identity with the corresponding region of nad5. (iii) The most striking repeated sequence in this nad gene cluster is an 800 bp segment, which is located in the spacer region between nad5 and nad4. This segment starts 206 bp downstream from the stop codon of nad5 and is duplicated on the opposite strand in the second group II intron of the cob gene, which encodes apocytochrome b protein as shown in Fig. 5. The repeated region extends from the 5' end of the intron to the end of the fifth stem structure, which is typical in group II introns (Michel and Dujon, 1983). Inverted repeat sequences of eight nucleotides (GAGTACC and GGTCCTC) were detected near the ends of the repeated sequence on the spacer region. In the Northern hybridization analysis, only a 9.6 kb premature RNA band was detected using oligonucleotide probe B, which specifically hybridizes RNA molecules including the spacer region between nad5 and nad4. If the cob gene is actively transcribed in the liverwort mitochondria, the repeated region in the cob intron should hybridize with the 9.6 kb premature RNA molecules generated from this nad gene cluster. This suggests that anti-sense RNA molecules complement to part of the spacer region between nad5 and nad4 are present in the liverwort mitochondria. It is noteworthy that no small RNA transcript is detected in this repeated spacer region. This finding suggests that an anti-sense RNA controls RNA stability in the mitochondria, but functional tests need to be carried out before this can be substantiated. The molecular mechanisms which lead to the generation of repeated sequences on the liverwort mitochondrial genome and their functions are also not known. Furthermore, the liverwort mitochondrial DNA is a single circular molecule as determined by electron microscopy and restriction mapping (Oda et al., 1992b), although many repeated segments are detected on the DNA sequence of this mitochondrial DNA. This suggests that liverwort mitochondria have lost recombinational system or have not acquired one during its evolution. It is possible that some DNA fragments generated from the mRNA transcript by reverse transcriptase
could have been integrated into distant regions of the liverwort mitochondrial DNA.

**Transcriptions of the nad1, nad3, nad4L, nad6 and nad9 genes in liverwort mitochondria**

To examine the viabilities of the liverwort nad genes, nad1, nad3, nad4L, nad6 and nad9, their mRNA transcription was analyzed by Northern hybridization using the appropriate synthetic oligonucleotides as probes (Fig. 6A). These five genes were found to have transcripts whose lengths were long enough to be their mature mRNAs (Fig. 6B). Two major transcripts (7.6 kb and 5.7 kb) of the nad1 gene were observed when probed with a synthetic oligonucleotide complementary to the 5'end portion of the gene (Fig. 6B, lane 1). The coding region of the nad1 gene has a length of 987 bp. Therefore, both of the transcripts are able to cover its coding region, and their large sizes suggest co-transcription with a putative orf154 at approximately 1.9 kb downstream from this gene (Fig. 6B, lane 1).

The transcripts of the nad3 gene were probed with a sequence complementary to the exon 1, producing three major signals of 4.8 kb, 3.2 kb, and 2.5 kb (Fig. 6B, lane 2). As the nad3 gene has an intron (1,485 bp in length), its removal may provide an explanation for the difference in size between 4.8 kb and 3.2 kb transcripts. The nad3 gene has a reading frame of only 357 bp, so there is the possibility that co-transcription occurs with the trnV-UAC gene at 607 bp downstream from the nad3 gene (Fig. 6A, lane 2). However, using a probe for trnV-UAC, it was demonstrated that none of the nad3 transcripts are of the same size as the trnV-UAC transcript (data not shown).

Using a synthetic probe complementary sequence to exon 2, the nad4L gene was demonstrated to have a rather complex pattern of transcription. An major discrete transcripts of 4.5 kb and heterogeneous transcripts of approximately 3.2 kb and 1.8 kb were detected (Fig. 6B, lane 3). Excision of the two introns in the nad4L mRNA did not offer a clear explanation for the identity of each transcript or provide an indication of simple splicing events. The lengths, more than 1.8 kb of the nad4L gene transcripts are sufficient for covering the coding region of the nad4L (0.3 kb), and possibly include the coding regions of two putative open reading frames (ORFs) upstream of (orf86a) and downstream from (orf244), the nad4L gene within a region of approximately 5 kb (Fig. 6A, lane 3).

The two major transcripts, 2.9 kb and 1.7 kb, of the nad6 gene were detected with the probe sequence complementary to a region exclusive to the nad6 gene, but not to orf100, which, as described above, has a 5'-half portion of the nad6 gene. Therefore, cross-hybridization to any orf100 transcript could be excluded. The major transcripts detected (2.9 kb and 1.7 kb) were much larger than the size (0.6 kb) of the nad6 gene (Fig. 6B, lane 4). Therefore, nad6 gene produced larger transcript sizes than the predicted for this gene indicating its co-transcription with neighboring genes upstream and/or downstream (Fig. 6A, lane 4).

In case of the nad9 gene, one major transcript (2.5 kb) was observed (Fig. 6B, lane 5). Since the coding region of the nad9 has a length of 639 bp, this transcript was thought to contain orf732 and/or atpA genes. Using the probes specific for exon 1 of the atpA gene, one band was found at the similar size of about 2.5 kb. On the other hand, no transcript having the same size was detected when probed for orf732 (data not shown), suggesting that the nad9 gene is co-transcribed with at least 5'-portion of the atpA gene.

All of the nad genes analyzed above were shown to be transcribed; this strongly suggests that they do encode proteins, which is further supported by the fact that their putative products and their counterparts from different organisms have conserved amino acid sequences. All of the transcripts are much longer in length than would be expected from the lengths of the coding region. This implies that they could be transcribed with their neighboring genes. The multiple sizes of transcripts for a single gene can be attributed to their representing different stages in RNA processing and to the occurrence of multiple initiation and termination sites for transcription. The transcripts of some maize mitochondrial genes have been demonstrated to have numerous transcription initiation sites (Mulligan et al., 1988a; Mulligan et al., 1988b), whereas a single transcription initiation site has been identified in other higher plant mitochondria (Rothenberg and Hanson, 1987; Young et al.,
Co-transcriptional expression of the three nad genes, nad5, nad4, and nad2

To study the expression of the clustered nad genes (nad5, nad4, and nad2), exon and intron specific oligonucleotide probes and a 664 bp BglII-PstI restriction fragment were prepared (Fig. 7A). The total mitochondrial RNA isolated from liverwort cells was hybridized with probes as shown in Fig. 7B. All the probes hybridized with an RNA band of 9.6 kb. This indicates that these three nad genes are actively transcribed in a single primary transcript. The RNA transcripts from the nad5 gene, which is located at the 5' end of the nad gene cluster were detected as three hybridizing bands of 9.6 kb, 2.8 kb, and 2.1 kb, in Northern blot that was hybridized with an oligonucleotide probe specific for the nad5 exon I (Fig. 7B, lane 1). A probe for the nad4 intron hybridized with bands of 9.6 kb, 5.4 kb, and 3.9 kb (Fig. 7B, lane 4), indicating that premature mRNA molecules containing the nad4 intron sequence are accumulated in the liverwort mitochondria as processed RNA molecules of 5.4 kb and 3.9 kb. The accumulation of mRNA molecules of 9.6 kb, 5.4 kb, 3.0 kb, and 2.2 kb, all including the nad2 intron was also observed in a Northern blot, which was hybridized with a probe for the nad2 intron (Fig. 7B, lane 7). On the other hand, putative processed RNA transcripts without intron sequences were detected as a 2.1 kb band coding for the nad5 by probing with oligonucleotide

---

**Fig. 6A, B.** Transcription analysis of the liverwort nad1, nad3, nad4L, nad6 and nad9. **A** Gene organization of the liverwort nad genes. The location of each gene is illustrated with a box, filled for exon and open for intron, respectively. Lane 1, nad1; lane 2, nad3; lane 3, nad4L; lane 4, nad6; lane 5, nad9. Each probe is shown by an asterisk under the gene organization. **B** Northern hybridization was performed by the probes specific to lane 1, nad1; lane 2, nad3; lane 3, nad4L; lane 4, nad6; lane 5, nad9 genes. Size of each transcript is indicated by a number in kilobases (kb).
Fig. 7A & B. Transcription analysis of the liverwort nad2, nad4 and nad5. A. Organization of the liverwort nad2, nad4 and nad5. Coding regions and introns are indicated by filled and open boxes, respectively. Locations of oligonucleotides and a 664-bp Bgelli-PstI DNA fragment used as probes are indicated by asterisks (1-5, 7 and 8) and a bar with arrows in both sides (6), respectively. Possible mRNA transcripts are shown with molecular sizes in kilobases (kb). A large repeated region is shown as a broken line between nad5 and nad4.

Fig. 7A & B. Northern hybridization of total mitochondrial RNA was carried out. Molecular sizes are indicated in kilobases (kb).
Chapter III

ynad7 gene in liverwort mitochondrial genome

Introduction

NADH dehydrogenase (NADH:ubiquinone oxidoreductase or complex I, EC 1.6.99.3) is the first enzyme in respiratory chain and consists of approximately 30-40 subunits (Weiss et al., 1991; Walker, 1992). As described in Chapter II, seven subunits are encoded by the mitochondrial genomes in mammals (Chomyn et al., 1985; Chomyn et al., 1986) and in Podospora anserina (Cummings et al., 1990). The corresponding mitochondrial genes are designated as ND1, ND2, ND3, ND4, ND4L, ND5 and ND6. On the other hand, other subunits are assumed to be nuclear-encoded. No genes for any subunits are found in yeast mitochondrial genome (de Zamaroczy and Bernardi, 1986).

Genes for seven subunits 1, 2, 3, 4, 4L, 5 and 6 of the complex (nad1, nad2, nad3, nad4, nad4L, nad5 and nad6) are identified on the liverwort mitochondrial genome (Oda et al., 1992a; Oda et al., 1992c) and their expressions at RNA levels are reported (Nozato et al., 1993; Yamato et al., 1993). Recently, ORFs homologous to the genes for the 30 kilodalton (kDa) subunits of bovine mitochondrial complex I are found in higher plant mitochondrial genomes and designated as nad9 genes (Kubo et al., 1993; Lamattina et al., 1993). The liverwort mitochondrial genome also contains its counterpart which was previously named orf212. Plant chloroplast genomes encode 11 genes homologous to those for the components of mitochondrial complex I and they have been named as ndhs (Ohyama et al., 1986; Shinozaki et al., 1986; Nixon et al., 1989; Dupuis et al., 1991), although their functions in chloroplasts have not been elucidated. Recently, however, it has been reported that their homologues, ndhB and ndhL, are essential to inorganic carbon transport in cyanobacteria, Synechocystis PCC6803 (Ogawa, 1991a; Ogawa, 1991b; Ogawa, 1992) and that NADH dehydrogenase is involved in the cyclic electron flow through PS I as well as the respiratory flow to the intersystem chain in Synechocystis PCC6803 (Mi et al., 1992).
It is reported that genes for the eighth 49-kDa subunits of the complex I are encoded by the nuclear genomes of bovine (Fearnley et al., 1989) and Neurospora crassa (Preis et al., 1990), and they are designated as ND7. Moreover, ND7-homologues (nad7) have been found in the mitochondrial genomes of wheat (Bonen et al., 1994), potato (Gäbler et al., 1994) and also in kinetoplastid in protozoa (Koslowsky et al., 1990). On the other hand, ORFs having sequence similarity with ND7 are identified on chloroplast genomes and named as ndhH (Fearnley et al., 1989).

Although reading frames homologous to portion of bovine and Neurospora crassa mitochondrial ND7 are detected in the liverwort mitochondrial DNA, those are supposed not to be functional, namely parts of a pseudogene, pseudo-nad7 (y nad7) (Oda et al., 1992a; Oda et al., 1992c). In this chapter, the author described the detail structure of y nad7, showed active expression of this pseudogene in a liverwort, Marchantia polymorpha, and discussed about the gene transfer into nuclear genome.

Materials and Methods

Analysis of nucleotide and amino acid sequences

Computer analysis of nucleotide and amino acid sequences was carried out as described in Chapter I.

Nucleic acids preparation

Liverwort mitochondrial RNA of the liverwort was isolated from 7-10 day old suspension cultured cells as described in Chapter I. Liverwort mitochondrial DNA was obtained from 7-10 day old suspended cultured cells as described by Oda et al. (1992b).

Liverwort nuclear DNA was isolated from 7-day-old suspension cultured cells. Cells were homogenized in isolation buffer [1M hexylene glycol, 10mM PIPES-KOH pH 7.0, 2mM MgCl2, 10mM β-mercaptoethanol, 10mM EDTA, and 0.5% Triton X-100] using French press, and filtered through two layers of miracloths.

The crude nuclear fraction was precipitated and washed with isolation buffer. The nuclei were fractionated by Percoll stepwise gradients with 60% and 90% Percoll in a solution containing 10mM PIPES-KOH pH 7.0, 2mM MgCl2, 10mM β-mercaptoethanol, 10mM EDTA, and 1M sucrose. The nuclear fraction between 60% and 90% Percoll was resuspended in resuspension buffer [1M hexylene glycol, 10mM PIPES-KOH pH 7.0, 2mM MgCl2, 10mM EDTA, 10mM β-mercaptoethanol, and 20% glycerol] and centrifuged at 1,000 x g for 10 min. The nuclei pellet was washed with resuspension buffer and resuspended in 20mM Tris-HCl and 10mM EDTA. Then nuclei were lysed by addition of 0.1 volume of 10% SDS and proteinase K (at final concentration 0.012%). Nucleic acid was extracted with phenol/chloroform and precipitated by ethanol.

Total cellular RNA was isolated using the guanidinium isothiocyanate procedure (Chomczynski and Sacchi, 1987). Poly(A) mRNA was purified by use of oligo(dT)-latex (OligolatecTM-dt30, Daisichi Pure Chemicals, Tokyo) according to the protocol of the manufacturer.

Southern and Northern blot analysis

Nuclear and mitochondrial DNA samples were digested with a restriction enzyme Xhol. They were electrophoresed in 0.8% agarose gels and transferred onto nylon membranes (Biodyne™ A, Pall, Tokyo). The membranes were prehybridized and hybridized at 42°C in a solution [0.5M NaPO4 pH 7.2, 1% BSA, 1mM EDTA, and 7% SDS]. After hybridization, filters were washed in 2 x SSC containing 0.1% SDS several times at room temperature and then in 1 x SSC containing 0.1% SDS at 42°C.

Denatured mitochondrial RNA and poly(A) mRNA samples were loaded on 0.8% agarose gels containing 2.2M formaldehyde, 20mM MOPS-KOH pH 7.0, 5mM sodium acetate, and 1mM EDTA, and blotted onto nylon membrane. Hybridization for mitochondrial RNA was performed at 45°C in a solution containing 6 x SSC, 0.1% SDS, 200 μg/ml calf thymus DNA, 1 x Denhardt's solution and 20% formamide. After hybridization, membranes were washed several times in 6 x SSC,
0.1% SDS at 42°C. On the other hand, hybridization for poly(A) mRNA was carried out at 42°C in a solution [0.5M NaPO₄ pH7.2, 1% BSA, 1mM EDTA, and 7% SDS]. The membrane was washed in 6 x SSC, 0.1% SDS at room temperature and then in 2 x SSC, 0.1% SDS at 42°C.

Oligonucleotide probes were synthesized by DNA synthesizer (Applied Biosystems, USA) as followed and then end-labeled by [γ³²-P]ATP (5,000 Ci/mmol, Amersham) using a polynucleotide kinase (Takara, Kyoto). An RNA ladder (BRL) was used as a size standard. A 800 bp DNA fragment was amplified by PCR using 5'-end and 3'-end primers specific to the hypothetical exon 2 and labeled with [α-³²P]dCTP (3,000 Ci/mmol, Amersham) using a Random primed DNA labeling kit (Boehringer Mannheim). The oligonucleotide sequences are:

- exon 1 : 5'-ACGGCATATTGGATTCTCATAGAGGCAC-3' (Fig. 4A, 1)
- intron 1 : 5'-AAATCGGATCTGTCGACCCTCTATGTA-3' (Fig. 4A, 2)
- exon 2 : 5'-GGACTGGCGGTGTAATGTTAAGGC-3' (Fig. 4A, 3)
- intron 2 : 5'-GTCCAAACAAGCGTGAGTGACATGACATCTGTCG-3' (Fig. 4A, 4)
- exon 3 : 5'-CATAGTACTCAAGATTTGAGTGAAGGAGGAG-3' (Fig. 4A, 5)
- exon 1-exon 2 : 5'-ACGGAATTCAATCAATGCTGTT-3' (Fig. 4A, 6)
- exon 2-exon 3 : 5'-ATGCGTAAACATGATAATGTTTGGAGAGGA-3' (Fig. 4A, 7)

Results and Discussion

Structure of ynad7 gene corresponding to the bovine ND7 subunit of NADH dehydrogenase

Previously, reading frames which showed significant amino acid sequence similarities with the eighth 49 kDa subunit of NADH dehydrogenase (ND7) from bovine heart (Fearnley et al., 1989) and Neurospora crassa (Preis et al., 1990) were detected between a transfer RNA gene cluster (trnD-trnS-trnA-trnT) and a ribosome protein gene cluster (rps10-rpl2-rps19-rps3-rpl16-rps5-rps14-rps8-rpl6-rps13-rps11-rps1) (Oda et al., 1992a; Oda et al., 1992c). These reading frames were designated as a pseudogene, ynad7, based on the following observations.

This hypothetical ynad7 gene product showed high amino acid sequence similarities with wheat NAD7 (88.2%), bovine ND7 (70.7%) and Neurospora ND7 (61.2%) (Fig. 1). This also showed 33.5% and 42.8% homologies with the product of Trypanosoma brucei mitochondrial MURF3 gene which were edited by addition and deletion of uridine and with the liverwort chloroplast ORF392, respectively. However, these reading frames in putative exon 1 and exon 2 were interrupted by six translational stop codons (three TGA, one TAG, and two TAA).

Two sets of 5'- and 3'-terminal consensus sequences (GUGYG and AC, respectively) for group II introns (Michel and Dujon, 1983) were located between three putative exons (Oda et al., 1992a). Although assumed intron regions of 3,062 bp and 1,427 bp could form almost typical secondary structures specific to group II introns (Michel and Dujon, 1983), some impairing bases between exon-binding sequences (EBS) and intron-binding sequences (IBS) were found in both of them (Fig. 2). In the first intron of ynad7, a base pairing could not be formed between EBS1 and IBS1 (shown by asterisks in Fig. 2A). On the other hand, in the second intron one base pairing between EBS1 and IBS1, and two base pairings between EBS2 and IBS2 could not be formed (Fig. 2B). It is suggested that both introns have no splicing activity and a mature mRNA is not produced.

Existence of traces of the RNA maturation-like reading frame in introns

In the hypothetical first intron of ynad7, reading frames which showed significant sequence similarities with the RNA maturation encoded in the first intron of mitochondrial coxl gene encoding cytochrome c oxidase in Saccharomyces cerevisiae (Bonitz et al., 1980; Carignani et al., 1983) or a reverse transcriptase-like reading frame in the first intron of Podospora anserina coxl (Cummings et al., 1990) were found (Fig. 3). However, in the liverwort sequences multiple frameshifts and several translational stop codons were detected. In addition, reading frames partly homologous to such intron-coded polypeptides were also found in the putative second intron, though they also contain several stop codons as in the case of the first
**Fig. 1.** Amino acid sequence comparison of ND7 related gene products. Hypothetical product of *vmad7* in the liverwort mitochondria (Liverwort mt), wheat mitochondria-encoded *nad7* gene product (Wheat mt), *Trypanosoma* mitochondria-encoded MURF3 (*Trypanosoma* mt), bovine nuclear-encoded ND7 (Bovine nc), *Neurospora* nuclear-encoded ND7 (*Neurospora* nc) and chloroplast-encoded ORF392 in liverwort (Liverwort cp) are aligned. Identical amino acid residues to those in the assumed *vmad7* product are shown by colons. Bars indicate artificial shifting to maximize sequence similarity and absence of corresponding amino acid residues. Insertion sites of introns are shown by small arrow heads over the sequences. Translational stop codons in *vmad7* gene are indicated by asterisks and filled triangles. Numbers at the ends of sequences indicate the numbers of amino acid residues which include in-frame stop codons. Amino acid sequence similarities with the liverwort mitochondrial *vmad7* gene product are also shown at the ends of sequences.

**Fig. 2.** Secondary structures of the two group II-like introns in *vmad7*. (A the first intron, B the second intron). Nucleotide numbers of loops and junction regions are indicated along the thin circle lines. Putative exon- and intron- binding sequences are highlighted by EBS and IBS with arrows, respectively. Nucleotides which do not make pairing between EBS and IBS are indicated by asterisks. Six stem structures specific for group II intron are indicated by numbers (I-VI).
Expression of \( \text{nad7} \) in the liverwort mitochondria

In order to know whether this pseudogene is expressed at a RNA level, exon and intron specific probes were generated as shown in Fig. 4A. Total mitochondrial RNA isolated from the liverwort cells was hybridized with \(^{32}\text{P}\)-labeled oligonucleotide probes (Fig. 4B). Probes for the putative three exons and two introns hybridized with bands of 16 kb and 9.6 kb (Fig. 4B, lanes 1 to 5). This indicates that these regions are actively transcribed in continuous mRNA transcripts of 16 kb or 9.6 kb which could cover not only \( \text{nad7} \) itself (5,668 bp) but also regions upstream and/or downstream. Northern hybridization probed for \( \text{rps10} \) which is located 1.3 kb downstream from \( \text{nad7} \) demonstrates the existence of the same 9.6 kb band as shown in Chapter I, supporting that \( \text{nad7} \) is co-transcribed with at least \( \text{rps10} \).

In addition to the two common bands, only one additional band was detected in all five cases. Namely, a 2.6 kb band was detected as probed only for the first exon (exon 1), while a 5.6 kb RNA molecule was found as probed for exon 2, exon 3, intron 1 and intron 2. This indicates that the 5.6 kb RNA molecules still contain these two introns and that these regions are not spliced out in the liverwort mitochondria.

To detect RNA molecules which contain joint sequences between exon 1 and exon 2 or between exon 2 and exon 3, synthetic oligonucleotide probes of 30 mers were used for RNA blot analysis (Fig. 4B, lanes 6 and 7). However, no significant hybridization signal was detected, confirming that joint molecules, namely spliced RNA molecules, do not exist in the liverwort mitochondria. No splicing event of the introns in \( \text{nad7} \) gene may be caused by impairing bases between EBS and IBS as mentioned above.

There are two possibilities in the production of a functional ND7 protein in the liverwort mitochondria as follows; (i) the functional ND7 gene encoded by the mitochondrial genome may have been transported to the nuclear genome and original \( \text{nad7} \) on the mitochondrial genome would have become a pseudogene. And biological active polypeptide of ND7 subunit may be transported from cytoplasm into mitochondria. Actually, in mammal or fungi ND7 is encoded by the nuclear genome.
and its translation products are imported into mitochondria (Walker, 1992). (ii) The chloroplast-encoded ND7 homologue, ORF392 products, might be transported from chloroplast into mitochondria by unknown mechanisms.

**Detection of nad7-like DNA segment in liverwort nuclear genome**

To know the possibility of transfer of nad7 coding region to the nuclear genome, liverwort genomic DNA digested with XhoI was blotted to a membrane filter and probed by the $^{32}$P-labeled exon 2 specific fragment (Fig. 5A). As a result, the exon 2 specific probe hybridized with one major band of 7.2 kb (Fig. 5B, lane 1). When the liverwort mitochondrial DNA digested with XhoI was probed with the same probe, a single 26 kb band was detected which corresponded to mitochondrial $\psi$nad7 (Fig. 5B, lane 2). In contrast, the exon 2 specific probe hybridized with a 42 kb XhoI chloroplast DNA fragment which corresponded to the nad7 homologue (Ohyama et al., 1986, data not shown). These results indicate that this 7.2 kb DNA fragment was not derived from organellar, mitochondrial and chloroplast genomes but from nuclear genome in liverwort cells and suggest that gene(s) for the mitochondrial ND7 polypeptide is encode by the nuclear genome in liverwort.

**Detection of poly(A)$^+$ mRNA corresponding nad7 gene in liverwort cells**

To detect RNA molecules derived from putative nuclear nad7, poly(A)$^+$ mRNA isolated from liverwort cells was probed by $^{32}$P-labeled DNA fragment specific to exon 2 as used in genomic Southern blot analysis. One major band was detected at the size of 2.2 kb in RNA blot (Fig. 5C). This indicates that poly(A)$^+$ mRNA molecules which has sequence similarity with mitochondrial $\psi$nad7 (at least exon 2 region) present in liverwort cells. Therefore, it is strongly suggested that the gene for the subunit 7 of the complex I is encoded by nuclear DNA and the translational products are transported into mitochondria in liverwort cells. Cloning and structural analysis of cDNA and genomic DNA encoding subunit 7 would be elucidated this assumption.

---

**Fig. 4A, B.** Northern blot analysis of mitochondrial RNA. A Gene organization of the liverwort $\psi$nad7 gene. Filled and open boxes indicate hypothetical exons and introns, respectively. Asterisks with numbers show positions of oligonucleotide probes used for RNA blot analysis. B Total mitochondrial RNA blots were probed by oligonucleotides specific for exon 1, lane 1; intron 1, lane 2; exon 2, lane 3; intron 2, lane 4; exon 3, lane 5; by an overlapping oligonucleotide between exon 1 and exon 2 (lane 6), and by an overlapping oligonucleotide between exon 2 and exon 3 (lane 7). Molecular sizes of transcripts are shown in kilobases (kb).
transferred from the mitochondrion to the nucleus via a C-to-U edited RNA intermediate. On the other hand, in cowpea, cox2 is not encoded by mitochondrial genome but by nuclear genome (Nugent and Palmer, 1991). These findings suggest that after the transfer of mitochondrial cox2 to the nucleus, the original mitochondrial gene has been lost in cowpea, while in soybean that was not lost but inactivated. In liverwort, mitochondrial nad7 was possibly transferred to the nuclear genome as in the cases of higher plant cox2 genes. Similarly to the cowpea cox2, liverwort nad7 was not lost but retained in mitochondrial genome. However, unlike cowpea mitochondrial cox2, liverwort mitochondrial nad7 is actively transcribed, although its transcript is not apparently functional. Recently, it has been reported that functional nad7 genes are encoded by the mitochondrial genomes in higher plants (Gülbet et al., 1994; Bonen et al. 1994). Therefore, a gene transfer event of nad7 has been presumably occurred in an ancestor of liverwort after the split of bryophyte. Since liverwort \( \psi \text{nad7} \) reading frames show high levels of amino acid sequence similarities with parts of wheat mitochondrial nad7, loss of function of the original nad7 on the liverwort mitochondrial genome possibly occurred more recently in evolution.

Figure 5A, B and C. A Gene organization of the liverwort mitochondrial \( \psi \text{nad7} \) gene. Bar with arrows shows an exon 2 specific DNA probe generated by PCR amplification using cloned mitochondrial DNA as a template. B Southern hybridization analysis of nuclear DNA (lane 1) and mitochondrial DNA (lane 2) from liverwort cells with an exon 2 specific probe. Molecular sizes are indicated as kilobase (kb). C Northern hybridization analysis of liverwort poly(A)\(^+\) mRNA with an exon 2 specific DNA probe. Two minor bands of 3.3 kb and 1.6 kb correspond to ribosomal RNAs.

It is reported that cytochrome c oxidase subunit 2 gene (cox2) encoded by soybean mitochondria is silent and that its functional counterpart is encoded by the nuclear genome (Covello and Gray, 1992). By comparison of mitochondrial and nuclear cox2 sequences, it is supposed that in an ancestor of soybean, cox2 was
References


Chomyn, A., Mariottini, P., Cleeter, M.W.J., Ragan, C.L., Matsuno-Yagi, A., Hatefi,


Summary

Chapter I

Sixteen genes for ribosomal proteins, rps1, rps2, rps3, rps4, rps7, rps8, rps10, rps11, rps12, rps13, rps14, rps19, rpl2, rpl5, rpl6 and rpl16, were detected in the complete sequence (186,608 bp) of the mitochondrial DNA from a liverwort, Marchantia polymorpha. The genes formed two major clusters, rps12-rps7 and rps10-rpl2-rps19-rps3-rpl16-rps5-rps4-rps8-rpl6-rps13-rps11-rps1, very similar in organization to E. coli ribosomal protein operons (str and S10-sp-α operons, respectively). In contrast, rps2 and rps4 genes were located separately in the liverwort mitochondrial genome (the latter was part of the α operon in E. coli). This finding supports the endosymbiont hypothesis, which postulated that organelles of eukaryotes originated from prokaryotic ancestors in the course of evolution. Furthermore, several ribosomal proteins encoded by the liverwort mitochondrial genome differed substantially in size from their counterparts in E. coli and liverwort chloroplast. The Northern hybridization analysis showed that rps2 and rps4 genes were transcribed in liverwort mitochondria. The rps12 and rps7 genes organized into the cluster were possibly co-transcribed. Additionally, it was suggested that the four genes, rpl2, rps19, rps3 and rpl16 which included in the large cluster were expressed as a single transcriptional unit and that rps10 were co-transcribed with ynad7. The remainder seven genes in the large cluster were supposed to be silent or to be transcribed at low levels.

Chapter II

The genes encoding subunits, 1, 2, 3, 4, 4L, 5, 6 and 9 of the NADH dehydrogenase (nad1, nad2, nad3, nad4, nad4L, nad5, nad6, and nad9) were identified in the mitochondrial genome of a liverwort, Marchantia polymorpha. Three genes nad5, nad4, and nad2 were tandemly clustered whereas nad1, nad3, nad4L, nad6, and nad9 genes were located separately on the liverwort mitochondrial genome. Their gene products showed high levels of amino acid sequence identity with the
correspondings from higher plants, and significant levels of similarity with those from liverwort chloroplast, *Podospora anserina* mitochondria, and human mitochondria. In addition, three clustered genes, *nad2*, *nad4* and *nad5*, have conserved amino acid residues in their central regions. Several regions of the *nad* genes were repeated in the liverwort mitochondrial genome. The Northern hybridization analysis using either exon or intron specific probes showed that all *nad* genes were transcribed in the liverwort mitochondria. It was also indicated that five genes *nad1*, *nad3*, *nad4L*, *nad6*, and *nad9* produced transcripts larger in length than would be predicted for the respective genes and thus were possibly co-transcribed with their neighboring genes upstream and/or downstream. On the other hand, three clustered genes were transcribed as a single precursor mRNA and were processed into mature mRNA molecules in the liverwort mitochondria.

Chapter III

A pseudogene, *vmad7*, which had a significant amino acid sequence similarity with the bovine nuclear-encoded gene for the eighth 49 kDa subunit of NADH dehydrogenase has been identified on the mitochondrial genome from a liverwort, *Marchantia polymorpha*. The predicted coding region, which included six termination codons, was actively transcribed into RNA molecules of 16 kb and 9.6 kb, but RNA splicing products were not detected in the liverwort mitochondria. This may be caused by the incomplete structures of the two hypothetical introns of this gene. Genomic DNA hybridization analysis and RNA hybridization analysis using poly(A)*mRNA* suggested that a structurally related nuclear gene encoded the mitochondrial ND7 polypeptide. These results imply that this *vmad7*, is a relic of a gene transfer event from mitochondrial genome into nuclear genome during mitochondrial evolution in *M. polymorpha*.

**List of Publications**


Acknowledgements

The author wishes to express her sincere gratitude to Professor Dr. Kanji Ohyama, Laboratory of Plant Molecular Biology, Department of Agricultural Chemistry, Faculty of Agriculture, Kyoto University, for his courteous guidance and continuous encouragement throughout the course of this study.

The author greatly appreciates to Dr. Hideya Fukuzawa, Laboratory of Plant Molecular Biology, Faculty of Agriculture, Kyoto University, and Dr. Takayuki Kobchi, Nara Institute of Science and Technology, for their invaluable suggestion and discussion and technical advice.

The author wishes to express her grateful acknowledgment to Professor Dr. Yasuyuki Yamada, Laboratory of Molecular and Cellular Biology, Faculty of Agriculture, Kyoto University, for his kind encouragement.

The author thanks Professor Dr. M.W. Gray, Dalhouse University, Canada for critical reading of the manuscript.

The author wishes to express her greatly appreciation to the author's colleagues in the laboratory, Dr. Yutaka Ogura, presently at the Research Institute for Biosources, Okayama University; Dr. Takeshi Kanegae, presently at Laboratory of Plant Morphogenesis, Faculty of Science, Tokyo Metropolitan University; Dr. Kenji Oda, presently at Department of Plant Biology, University of California, USA; Dr. Katsuysuki Yamato, presently at Division of Biological Sciences, University of Missouri-Columbia, USA; Dr. Eiji Ohta, presently at Kubota Ltd.; Mr. Yasukazu Nakamura, presently at National Institute of Genetics and Ms. Naoko Nozato, presently at Asahi Breweries Ltd. for their substantial collaboration.

The author thanks Mr. Kinya Akashi and Mr. Yoshihide Kobayashi, and the all other members of the laboratory.

Lastly, I am much indebted to my father, my mother and my brother for their affectionate supports.

March 1995

Miho Takemura