京都大学学術情報リポジトリ
KURENAI 紅
Kyoto University Research Information Repository

京都大学
KYOTO UNIVERSITY

Title	Organization and expression of mitochondrial genes for ribosomal proteins and NADH dehydrogenase subunits from a liverwort, Marchantia polymoropha(Dissertation_全文)
Author(s)	Takemura, Miho
Citation	Kyoto University (京都大学)
Issue Date	1995-03-23
URL	http://dx.doi.org/10.11501/3080950
Right	
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

Contents

Introduction		 1
Chapter I.	Genes for ribosomal proteins in liverwort mitochondrial genome	 5
Chaper II.	Genes for NADH dehydrogenase subunits in liverwort mitchondrial genome	 29
Chaper III.	*ψnad7* gene in liverwort mitchondrial genome	 57
References		 71
Summary		 77
List of publications		 79
Acknowledgments		 81

Abbreviations

ATP	adenosine 5'-triphosphate
dCTP	deoxycytidine 5'-triphosphate
bp	base pair(s)
BSA	bovine serum albumin
Da	dalton
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
Hepes	2-[4-(2-hydroxyethyl)-1-piperazinyl] ethanesulfonic acid
kb	kilobase pair(s)
MOPS	3-(N-morpholino) propanesulfonic acid
mRNA	messenger RNA
mtDNA	mitochondrial DNA
mtRNA	mitochondrial RNA
ORF	open reading frame
PCR	polymerase chain reaction method
PIPES	piperazine-N, N'-bis-(2-ethanesulfonic acid)
RNA	ribonucleic acid
rRNA	ribosomal RNA
SDS	sodium dodecyl sulfate
SSC	standard saline citrate
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA

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.

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, H^+ ATPase subunits, NADH dehydrogenase subunits, cytochrome *c* oxidase subunits, and apocytochrome *b*, respectively. *orfs* 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, *ψnad7* 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.

Chapter I
Genes for ribosomal proteins in liverwort mitochondrial genome

Introduction

Organelles (mitochondria and plastids) contain prokaryotic-type ribosomes, whose constituent proteins are partly encoded by the organelle genome, the remainder being specified by the nuclear genome and imported into the organelle post-translationally. The complete nucleotide sequences of liverwort, tobacco, and rice chloroplast genomes have revealed that each encodes about 20 genes for ribosomal proteins (r-proteins) (Ohyama *et al.*, 1986; Shinozaki *et al.*, 1986; Hiratsuka *et al.*, 1989) (Table 1.). On the other hand, the completely sequenced human mitochondrial genome has no genes for r-proteins (Anderson *et al.*, 1981), while the yeast mitochondrial genome encodes only one species of r-protein (Butow *et al.*, 1985). In these latter cases, all or almost all of the mitochondrial r-proteins must be encoded by the respective nuclear genomes. In fact, some yeast nuclear genes for mitochondrial r-proteins have been cloned and sequenced (Myers *et al.*, 1987; Kitakawa *et al.*, 1990; Kang *et al.*, 1991). To date, only ten r-proteins (S1, S3, S7, S10, S12, S13, S14, S19, L5 and L16) have been described in the mtDNA of several species of angiosperms (Gonzaletz *et al.*, 1993; Hunt and Newton, 1991; Gualberto *et al.*, 1988; Ye *et al.*, 1993; Sutton *et al.*, 1993; Bock *et al.*, 1994; Zhuo and Bonen, 1993; Zanlungo *et al.*, 1994; Suzuki *et al.*, 1991; Kim *et al.*, 1991; Bland *et al.*, 1986; Schuster and Brennicke, 1987a; Bonen, 1987; Wissinger *et al.*, 1990; Wahleithner and Wolstenholme, 1988; Schuster *et al.*, 1990a; Brandt *et al.*, 1993; Conklin and Hanson, 1991; Schuster and Brennicke, 1991; Schuster, 1993). The complete sequence of the liverwort mitochondrial DNA have been determined in this laboratory (Oda *et al.*, 1992a) and genes encoding sixteen different r-proteins (S1, S2, S3, S4, S7, S8, S10, S11, S12, S13, S14, S19, L2, L5, L6, and L16) have been identified. In this chapter, the author described the gene organization and the characteristics of the deduced amino acid sequences of these mitochondrially encoded r-proteins. Furthermore, the author reported the expressions of these genes at RNA levels.

Table 1. Genes for ribosomal proteins found in several organisms.

genome	ribosomal protein genes
E. coli	*rpsA-rpsU* *rplA-rplG, rplI-rplS, rplU-rplY, rpmA-rpmI*
Mitochondria	
Human	-
Yeast	*var1*
Neurospora	*rps5*
Angiosperm	*rps1, rps3, rps7, rps10, rps12, rps13, rps14, rps19, rpl5, rpl16*
Liverwort	*rps1, rps2, rps3, rps4, rps7, rps8, rps10, rps11, rps12, rps13, rps14, rps19, rpl2, rpl5, rpl6, rpl16*
Chloroplast	
Liverwort	*rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps15, rps18, rpl2, rpl14, rpl16, rpl20, rpl21, rpl22, rpl23, rpl33, rpl35, orf69*

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-9801VM 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 gradients (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 capillary-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 6 x 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^{32}-P]ATP (5,000 Ci/mmol, Amersham) using a MEGALABEL kit (TAKARA, Kyoto).

rps2 : 5'-GGCCTTTTGCACTAATGATAGATCCAATC-3' (Fig. 5A, 1)
rps4 : 5'-AGTTGCCTTGCTTGAAACATAG-3' (Fig. 5A, 2)
rps12 : 5'-CGAACAAGCTGATTCATTGTTGGC-3' (Fig. 6A, 1)
rps7 : 5'-GAAACTACGATTGGCTTCAGC-3' (Fig. 6A, 2)
atp6 : 5'-GGGGAAAAACCGTTGTTTCACCG-3' (Fig. 6A, 3)
ψcob : 5'-TGGCTCCATGAAGAGGATCCC-3' (Fig. 6A, 4)
rps10 : 5'-GCAGAAGCCCTGACCTTTGATTCTCAAAAG-3' (Fig. 7A, 1)
rpl2 : 5'- GCTTCGATCCACCTCCTCGGTGAAA-3' (Fig. 7A, 2)
rps19 : 5'-AGGACCTTTCCATATAGAGCGTG-3' (Fig. 7A, 3)
rps3 : 5'-GAGTCTGACTGAAATCGGATTTAC-3' (Fig. 7A, 4)
rpl16 : 5'-GCTTATAGCACGACGCGCTGCTTCAA-3' (Fig. 7A, 5)
rpl5 : 5'-CATGATATGCCCTCGTAGAGTGC-3' (Fig. 7A, 6)
rps14 : 5'-CTTATATACGGAACGAGGGCGCCCTGTG-3' (Fig. 7A, 7)
rps8 : 5'-CTTGGCATAGAATCTCACCGCC-3' (Fig. 7A, 8)
rpl6 : 5'-CCCTTTATAAACTTCAGGAGG-3' (Fig. 7A, 9)
rps13 : 5'-CGTTGGCCGCGTAAGGGCAATCC-3' (Fig. 7A, 10)
rps11 : 5'-GGTCGGCATCCATTATGTGGCG-3' (Fig. 7A, 11)
rps1 : 5'-GGAGTTTTCAGTCCTGTATCCACCAA-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 rps4* 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*.

Protein	*E. coli*	Angiosperm mt	Liverwort cp	Liverwort cp/*E. coli*
S 1	23.7	43.6	-	-
2	27.1		22.8	44.3
3	25.4	46.9-50.4	24.1	40.6
4	25.0		22.4	40.1
7	35.8	58.5	29.9	43.8
8	35.1		26.7	45.5
10	31.5	58.3	-	-
11	48.0		48.8	51.5
12	62.1	80.8	64.2	70.2
13	38.3	55.0-61.2	-	-
14	43.3	64.6-70.4	38.4	45.0
19	41.7	54.1	42.9	63.0
L 2	44.8		43.2	48.4
5	28.6	60.9-68.7	-	-
6	36.6		-	-
16	50.4	71.9-78.5	45.9	53.8

(a) *rps1*

```
Liverwort  mt  M--SFSQLFPKYNSSFNPLRGSAIQCSVIQLQQNKVLVDTGLKT----PIICFQHELKRVPITK---------------
Wheat      mt                                  MF:::A:iG:----:K::M:D::TGL::KRATRFENKVGFKNVVAG
E. coli        :TE::A:::EESLKEIETRP::IVRGV:VAIDKDV::::A:::SESAI:AEQ:KNAQGELE:QV---------------

               --------------------------QARFHFGIEDVEV-FGEPKMLLPKPLEIKCKRKLVWIELTKIWRSDQNLVKGF
               ESLIKKRIFERPFIDLVAGESLIKERAA:::NDFVGSLD:AA:::L-:::QRP----RQNRA::::K::::T-KKK::::
               --------------------------GDEVDVALDA::DG:::T--::SR--:-:A::HEA::T:E:AY-E:AET:T:V

               ILNSVKGGYAVAIAGYIAFLPKS--LLRS-RKVFYSQWRI--FSILNMKPKISNIVVKEIGDGKIDYFSPTKSHQKQTKY
               :IKKI::::L:::::FTT:::FKKA::KK-:IANDR------:T:DSIN::RRD::IIAAD*-----------------
               :NGK::::FT:ELD:IR::::G:LVDV:PV:DTLHLEGKELE:KVIKLDQ:RN:V::SRRAVIESENSAERDQLLENLQE

               LGAKLKHWRNMKKNTNVKKKYIFSEKVPTTKKTKQGFKHLGPKPLAYTEKKRETTKQSTKNNVFQLKDQGQGKSLVFVDV
               --------------------------------------------------------------------------------
               GMEVKGIVK:LTDYGAFVDLGGVDGLLHI:DMAWKRV::PSEIVNVGD:ITVKVL:FDRERTRVS:GLKQL:EDPWVAIA

               LTQSS---------------------------------------------------------------------------
               --------------------------------------------------------------------------------
               KRYPEGTKLTGRVTNLTDYGCFVEIEEGVEGLVHVSEMDWTNKNIHPSKVVNVGDVVEVMVLDIDEERRRISLGLKQCKA

               --------------------------------------------------------------------------------
               --------------------------------------------------------------------------------
               NPWQQFAETHNKGDRVEGKIKSITDFGIFIGLDGGIDGLVHLSDISWNVAGEEAVREYKKGDEIAAVVLQVDAERERISL

               --------------------------------------------------------------------------------
               --------------------------------------------------------------------------------
               GVKQLAEDPFNNWVALNKKGAIVTGKVTAVDAKGATVELADGVEGYLRASEASRDRVEDATLVLSVGDEVEAKFTGVDRK

               ----------------------------------------------  270
               ----------------------------------------------  170   43.6%(58/133)
               NRAISLSVRAKDEADEKDAIATVNKQEDANFSNNAMAEAFKAAKGE   557   23.7%(64/270)
```

(b) *rps2*

```
Liverwort  mt     MYNSNLLVIQKLLSTNAYLGHRIPTSDFQGYLYGFRNEMAIIDLEKTLICLRRTCNLIGSIISAKGH-LLLVNTNPE
E. coli            MATVSMRDMLKAGVHFGHQT:YWNPKMKPFIP:A::KVH::N::::VPMFNEALAELNK:A:R::K-I:F:G:KRA
Liverwort  cp  MKQKSWNIH:EEMMEAGVHFGHQA:KWNPKMAP:IFTE:KGIH::N:TQ:ARF:SEA:-DLVANA:S::KQF:I:G:KYQ

               YNKIIQQMAKKTNQSYINHKWIGGFLTNWKHMKKVKKHFQDFSAHPNLKDAFTSSPFDYFPRFKKMQKCFE---GIMTHN
               ASEAVKDA:LSCD:FFV::R:L::M:::::TVRQSI:RLK:LETQSQDGTPDKLTKKEALM:TEELE:LENSLG::KDMG
               AADL:ESS:L:ARCH:V:Q::L::M::::STIETRLQK:K:LENKKKTGTINRLPKKEAANLKRQLDHLQKYLG::KYMT

               I-PDCLVIINANQNSMAILEANQLQIPIVALVDSNIPNRLHKLITYPVPVNDDSIKFVYLFCNLITKTVILSKRSQRPKV
               GL::A:FV:D:DHEHI::K:::N:G::VF:I::T:SDPDGVDF---VI:G:::A:RA:T:YLGAVAA::RE-G:::DLAS
               SL::IVI::DQQKEFT::Q:CIT:G::TIC:::TDCDPDMTDI---:I:A:::ARASIRWIL:KL:LAICE-G:YNSI:N

               KVKRL       237
               QAEESFVEAE  241   27.1%(64/236)
                           235   22.8%(54/237)
```

(c) *rps3*

```
Liverwort  mt  MAQKVNPISVRLNLNRSSDSSWFSDYYYGKLLYQ-DLNFRDYFGSIRPPTGNTFGFRLGRCIIHHPPKRTPIHVFFLDRL
Maize      mt  ::R:G:::::::D::::::P:i::::::::::i::::-:V:L:S::S::::::RL:::::::::::L:::::::::F:LPR:P
Oenothera  ps  ::*:G:::::::D:::::::::::::::::::iV::-:V:L:S:i::::::::RL:::::::::::L:::::::::F:LPR:P
Petunia    mt  ::R:G:::::::D::::::::::::::::::iV::-:V:L:S::::::::R:RL:::::::::::L:::::::::F:LPR:P
Brassica   mt  ::R:G:::::::GK:::::::R::::::::FV::-:V:L:S:::::::::::RL:::::::::::LL:::::::::F:LPR:P
E. coli        :G:::H:NGI::GIVKPWN:T::ANTKEFADNLDS:FKV:Q:LTKELAKASV----------SRIVIE:PAKSIRV----
Liverwort  cp  :G::I::LGF::GITQNHR:Y::ANKK:S:VFEE-:KKI::CI--ELYVQKHIKNSSNYGL:ARVEI::KTDLIQV----
```

Fig. 1. (cont.)


```
SQSRHQGLGAIPSVKLIRRI-------------------------------NDNTVKQR---NEVGIWPKK--RYEYH
LRLKRRDKSRPGKD:GRWWA-FGKVGPIGCLHSSEGTEEERNEVRGRGAGKRVESIDREKQ---::IR::::MQ::G::
RRLKRGKKSRPGKE:ARWWE-FGKVGLIGCLHSNDDTEEERNEVGGRGAGKRVESIRLDD:EKQ::IR::::KQG:G::
RRLKRRKSRPVKE:GRWGA-FGKVGPIGCLHSSDGTEEERNEVRGRGAGKRVESIRLDD:EKQ::IR::::KQG:G::
RRLKRREKTRPGKE:GRWWTTFGKAGPIGCLR--DDTEEERNEVRGRGARKRVESIRLDD:KKQ::IRG::::KQ::G::
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------

DRLPSIQK-IDQLLRVSDWMADIHSTFQS--IWPKDENDDRRASEERYAFSRFAPSILVAVRAEKKKAIFGSEGDPFGPT
::T::RK:NFSKS::::RAFKHPKYAGVVND:AFLI::::SFIKTKLFK:FFLPKKSRSDG:TSHLLKRTLPAV------
::S:::K:NLSKi::::GAFKHPKYAGVLND:AFLI::::SFRKTKLFK:FF-PKKSRSDGPTSHLLKRTPPAV------
::S:::K:NLSKi::::GAFKHPKYAGIEND:AFLI::::SFRKTNLFK:FF-PKKSRSDRPTSHLLKRTLPAV------
::T:::K:NLSKS::I:GAFKHPKYGGVVND:AFLI::::SFRKTKFFK:FF-PKKSRSDGPTSYL--RTLPAV------
-------------------------------------S------------------------------------------
--------------------------------------------------------------------------------

GRAFLDYFVMQYFFNLKNQIQFDPMVN-RSPVAQGVAKTSMIGKAI---PAKTEQGTQSGESICQPRSTLYFDAI-----
-:PS:N:S:::::::T::KMH:::V:VLNHF::P:::EP:TM:G:K---GGSLDKRIR:RIAFFVES::SDKKCLARAKK
-:PS:N:L::::YL:T:::MH:::L:VLNHF::P:::EP:TM:G:N-RQGRSN:LRIR:RIAFFVESL:SEKKSLAEDKK
-:PS:N:S::::LL:T::K:H:::V:VLNHF::P:::EP:TN:G:N-AQGRSLDKRIA:CIAFFVEN::SEKKCLAEAKK
-GPS:NFL::::::T:::MN:::V:VLNHF::P:A:EP:TM:R:NGTGDRSLQKRIR:RIAFFVES::SEKKCLAEAKN
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------

--------------------------------------------------------------------------------
RLIHFIRQANDLRFAGTTKTTISLFPFFGATFFFSRDGVGVYNNFFYEYAREQLLGQLRIKCRNLMGKDKVMELIEKFID
GLPHFIRQENDLRFAGRTKTTISLFPFFGATFFPRDGVGVYKHLFFEDAREQLLGQLRIKCWKLMGKDKVMELIEKFID
RVTHFIRQANDLRFAGTTKTTISLFPFFGATFFFLRDGVGVYNNLFFEDAREQLLGQLRRKCWNLMGKDKVMELIEKFID
RLTHLIRLANDLGFAGTTKTTISLFPFFGATFFFLRDGVGVYNNL---DAREQLLNQLRVKCWNLLGKDKVMELIEKFKD
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------

---------------IFLRYARFRKATSLSSRYYYLKKMQSLFSNQTKTNTLIQPVKIASVYQSASLIAQEISWKLEQK
LGRIGKLIKGIEMMIE:I::KR:IPYGYN-:YLNEV-Q::R:FL::R:N::::ES:::K::::::::::D::FQ:RNN
LGGIGELIKGIEIMIE:I::NR:IPYGYN-:YLNEV-:::R::L::R:::::I:ES:::K::::::::i:::D::FQPRN:
LNRIGELIRGIEMMIE:I::NR:IPYGYN-YYLNEV-:::R::LY:R:N::::ES:::K:::::::i:::D::FQiRN:
LGGIEELIKVIDMMIE:I::KRGIPYGYN-:YFNEV-Q::R:FL::R:N:K::ES:::K::::::::::D::FQ:KN:
---------------TIHTARPGIVIGKKG--EDVE:L-RKVVADIAGVPAQ:NIAEVRKPELD:K:V:DS:TSQ::RR
---------------EIYTGFPALLVE:RGQGIEQ::LNVQNILSSEDRRLRMTLIE::KP:GEPKIL:KK:AL:::SR

K-SFRQICRSIFKQIKKC-P-YVKGIRIGCSGRL-NGAEIAKTECKKYGETSLHVFSDQIDYAKTQASTPYGILGVKVWV
PI:::S:FSK:V:D:PLIM:KG:E::::C:::::-G:::::R:::G:::K::CN::NQK::::PAEV::RD::S::::RI
RR:::S:FSQ:V:D:PLVMKKG:E::::C::::i-K:::::R:::G:::K::CN::NQK::::SAEV::R::::::::I
TR:::S:FSK:V:D:PLVMKKG:E::::C:::::-E:::::R:::G:::K::CN::NQK:::iLAEV::R:::S:::::I
RR::HS:FAK:V:E:P:----R:E::::CF::::KDA::K:Q:K:Y:HRK::RN::NQK::::PAEV::R:::S:::::I
V-M::RAMKRAV-:NAMR-L-GA:::KVEV:::-G:::::R::WYRE:RVP::TLRAD:::NTSE:H:T::VI:::::I
V-A::RTMKKAI-ELA:K-G-NI:::K:QIA:::-::::::RV:WARE:RVP:QTIRAR:N:CYYA:Q:I::V::I:::I

SYFLTQKKGTSCAISKTYKIS             430
::SQ-N:::R--:::E::E:              559   49.6%(207/417)
::SK                              550   50.4%(203/403)
::SK-K:::R--:::E::E:              562   50.1%(208/415)
::SQ-K:G:R--:::E::E:              555   46.9%(194/414)
PKGEILGGMAAVEQPEKPAAQPKKQQRKGRK   233   25.4%(60/236)
PQDEE                             217   24.1%(53/220)
```

(d) *rps4*

```
Liverwort  mt  MFASRFKVCRQILENVWQTKKLTLKQKFLISELQKQKKN--KKQSDFSI--QLQTIKKLSLFYGNLPIKK---MQRAKTH
E. coli           MARYLGPKLKLSRREGTD:F::SGVRAID-T:C:IEQAPG:HGARKPRLSDYGVQ:REKQKVRR:YGVLER:FRNYY
Liverwort  cp     MSRYRGPRVKIIRRLGA:P-GLTNKTLK:K-SGYI-NQS-TSNKK--VS:YRIR:EEKQKLRFHYGLTER:LL:YV

               TYIDKKNS-----LLFNIEKRLDVILVRLNFCSTMFQARQLISHKNICVNYKKVNIPGFQVSNGDLISIQENSLDFFKSN
               KEAARLKGNTGEN::ALL:G:::NVVY:MG:GA:RAE::::V:::A:M::GRV:::ASY:::PN:VV::R:KAKKQ--:R
               RIAR:AKGSTGQV::QLL:M:::N:IF::GMAP:IPG::::VN:RH:LI:NNT:D::SYNCKPK:V:T:KDR:KSQ--:I
```

Fig. 1. (cont.)

```
IRKNFQTNRIRRMKPNHLEVNYKTLKAVVLYEPQQIQFPYKIDLDLLD         196
V-:AALELAEQ:E::TW:::DAGKMEGTFKRK:ERSDLSAD:NEH:IVELYSK     206   25.0%(49/196)
:I::-LNSFQKQKI::::TFDLMQI:GL:NQIIDREWIYL::NEL:VVEYYSRQV   202   22.4%(44/196)
```

(e) *rps7*

```
Liverwort  mt  MNLFGKSNFNCVFSSSLDWFHSSRLSEKVGTKKNRICRETESFYALCLSHRRYLCYALEGLLPSRPRGRRASTYNCSDNL
Wheat      mt
E. coli                                                                                  MPRRRVIGQ
Liverwort  cp                                                                            MSRKSIAEK

               GYIRGLNGKQKQLIKKLVHICMIDGKKTRSRAIVYKTFHRLAPHGDVIKLLVN---AIENVKPICEVKKVRISGTTRLVP
                 MGDFD:E::E:::::NFR::::R::V::::::::::RTERDVIK:MVD--:VD:I:::::V::GVA::IYD::
               RK:LPDPKFGSE:LA:F:N:L:V::::STAES:::SALET::QRSGKSE:EMAFEV:L:::R:TV:::SR:VG:S:YQ::
               QVAKPDPIYRNR:VNM::NRILKN:::SLAYR:L::AMKNIKQKTKKNP:F:LR-Q:VRK:T:NVT::AR::D:S:YQ::

               SIIATNRQETLAIRWMLESAAKRRMGKKSISLDQCLYAEILEASQKMGIARKKRDDLHKLAEANRSFSHYRWW
               G:V:RD::Q::::::I:GA:F:::-ISYR:::EK::F::::D:YR:R::S::R:EN::G::ST::::A:F:::
               VEV-PV:RNA::M::IV:A:R::GDKSMALR:ANE:SD-AA:N---K:T:V:::E:V:RM::::KA:A::::LSLRSFSH
               LE:KSTQGKA:::::L:GASR::SGQNMAFK:SYE:ID-AARD---N:::IR:KEET::M:::::A:A:F:

                                 230
                                 148   58.5%(86/147)
               QAGASSKQPALGYLN   178   35.8%(57/159)
                                 155   29.9%(47/157)
```

(f) *rps8*

```
Liverwort  mt  MHTLSNLLSSIKNAQKAQKRVLYFSSFKKISKRKKRFVCSACKMMPRVFVSRLCWDFCRILYNEGYIHGFSQEADG-S-L
E. coli                                   MSMQDPIADMLTRI:NGQA:N:AAVTMPS:K:KVAIANV:KE::F:ED:KV:G:TKPE:
Liverwort  cp                             MGNDTIANMITSIRNANLGKI:TVQVPATNITRNIAKILFQEGFIDNFIDNKQNTKDI:

               RIVLKYHSSGIGV---IKKMKTISKPGFRIYSSKNRLSKKREGLGITILSTSKGNLICDREAQKTNFGGGEILCQVF
               ELT:::FQ-:KA:VES:Q---RV:R::L:::KR:DQ:P:VMA::::AVV:::::VMT-::A:RQAGL-::::I:Y:A
               ILN:::-Q-:KKKKSY:TTLRR:::::L::::NHKEIP:VLG:M::V:::::R:IMT-::::RQKKI-:::L::Y:W

               152
               130   35.1%(46/131)
               132   26.7%(35/131)
```

(g) *rps10*

```
Liverwort  mt                          MTAKICIVIKSF-----ENQRSGLLLNTRKIGLPKKQTLYTVLRSPHIDKKSREQFEMRIH
Potato     mt  MRQRRALRRVSQKERPPKVT:T::G:::R::DHPFL::HFW::PPY::::::ESRV::::::::::::::::::F:K:K
E. coli                  MQNQRIRIRLK:FDHRL:DQATAEIV:TAKRTGAQVRGP:P::TRKERF:::I:::VN:DA:D:Y:I:T:

               KQLLV-IETETHKLREKLNWLKLHDLLGVQVKIIFYYQTRLDKVCKS     102
               :EF::-:K::R:E::K:FFRLKRRATRRT                       108   58.3%(49/84)
               L-R::D:VEP:E:TVDA:MR:D:AAGVD::ISLG                  103   31.5%(28/89)
```

(h) *rps11*

```
Liverwort  mt  MQKK-----HGITNMQKKHCITYIQSTFGNTIITLTDYNGNTKTWSSSGSVGFKGSRRSTNYAAQATA-ENAARVAIQLG
E. coli        :A:APIRARKRV-RK:VSDGVAH:HAS:N:::V:I::RQ::ALG:ATA:GS::R:::K::PF:::-V:AGRC:DAVKEY:
Liverwort  cp  :P:SVKKINLRKGKRRLPKGVIH::AS:N:::V:V::IR:QVVS:::A:AC::::TKK::PF:::-::A:::I:IL:DQ:

               FKFVEVRIKGLGYGKESSLRGLKLGGLIITKIRDVTPTPHNGCRPPKKRRV   125
               I:NL::MV::P:P:R::TI:A:NAA:FR::N:T::::I:::::::::::::   129   48.0%(60/125)
               M:QA::M:S:P:P:RDTA::AIRRS:I:LSFV:::::M::::::::R::::   130   48.8%(61/125)
```

Fig. 1. (cont.)

(i) *rps12*

```
Liverwort  mt  MPTMNQLVRKGRESKRRTKRTRALNKCPQKQGVCLRVSTRSPKKPNSALRKIAKVRLTNRNEIIAYIPGEGHNLQEHSVV
Maize      mt  :::K:::I:H:::E::::D::::iDQ::::::::::::::T:::::::::::::::::S::HD:F:i::::::::i::::I:
Wheat      mt  :::K:::I:H:::E::::D::::iDQ::::::::::::::T:::::::::::::::::S::HD:F:i::::::::i::::I:
Oenothera  nc  FA:::::I:H:::E:H:SD:K:::G::::::::::::::::T::::::::G:::::::S:::DVF::::::::::::::M:
E. coli        :A:V::::::P:AR:VAKSNVP::EA::::R:::T::Y:TT::::::::::VCR:::::GF:V:S::G::::::::::::I
Liverwort  cp  :::IQ::I:NK:QPIENRTKSP::KG:::RR:::T::Y:TT::::::::::::R::::SGF::T:::::I:::::::::::

               MVRGGRVQDLPGVKYHCIRGVKDLQGIPGRRRGRSKYGTKKPKDYI   126
               L::::::K:i::::F:i::::::::L:::D::K::::::AER::SK   125   80.8%(101/125)
               L::::::K:i::::F:i::::::::L:::D::K::::::AER::SK   125   80.8%(101/125)
               L::::::K::::::F::::::::::L:::D::::::::::AE:::ST   125   79.2%(99/125)
               LI:::::K:::::R::TV::AL:CS:VKD:KQA:::::V:R::A      124   62.1%(77/124)
               L::::::K:::::R::I:::TL:AV:VKD:QQ::::::V::S:       123   64.2%(79/123)
```

(j) *rps13*

```
Liverwort  mt  MSYILGTNLNSNKQVKIALTRIFGIGPKKAIQVCDQLGLSDTIKVNKLTKYQFDQILKIISQNYLVDSELKRVIQRDIKR
Wheat      mt  ::::S:ARSLPDE::R::S:KMD:::::::::LRYR::I:GN::M:E:::::I:::EQM:A:DHV:HW::::GERA::E:
Maize      mt  ::::S:ARSLPDE::R::S:KMD:::::::::LRYR::I:GN::IHE:::::I:::EQM:A:DHV:HW::::GERA::E:
Tobacco    mt  :L::S:AR:VGDE::R::S:K:D::::::::::RYR::I:GN::IKE:::::I:::EQM:G:DHV:HW::::GERA::E:
Oenothera  mt  ::::S:AR:VADE::R::S:KMD::::::::::RSR::--GN::RKE:::::I:::EQMRG:DHV:HW::::GERA::ER
E. coli        VAR:A:I:IPDH:HAV::::S:Y:V:KTRSKAILAAA:IAEDV:ISE:SEG:I:TLRDEVAK-FV:EGD:R:E:SMS:::

               LISIGCYRGFRHNAGLPLRGQRTHTNAKTCRKLRYVSIRS          120
               ::::SR:::I::QD:S::::::::::R:A::QIWK               116   57.8%(67/116)
               ::::SR:::I::QD:S::::::::::R:A::QIWKGNERRLPKEQATD  129   55.0%(66/120)
               ::::S::::I::QD:S:::::::::::R:::::IWK              116   61.2%(71/116)
               F:::S::::I::QD:i::::::S::::R:S::RIRK              114   56.0%(65/116)
               :MDL:::::L::RR:::V:::::K:::R:-::GPRKP:KR          118   38.3%(46/120)
```

(k) *rps14*

```
Liverwort    mt                  MSNQ--IIRDHKRRLLVARYELKRMHYKAICQDRNLPNKIRYEYFFKLSKLPRNSSKTRVRNRCIFT
Broad bean   mt                  ::EKR-N:::::::::A:::::R:KL:::F:K:SD::SDMWDKLRY::::::::::FA:::::::S:
Oenothera    mt                  :-EKR-N::::::::::T::::::KL:::F:N:PA::SDM:DKHRY::::::::::FA::::::::i:
Arabidopsis  ps                  ::*KQ-NS::::::::A::F::R:KL:::F:K:PD::SDM:DKHCY::::::::::FA::::::::i/
Brassica     mt                  ::EKQ-NS::::::::A::F::R:KL:::F:K:PD::SDM:DKHRY::::::::::FA::::::::ST
Yeast        nc  MGNFRFPIKTKLPPGFINA:L::NFK:QQFKEN:ILVKSL:F:ARNM:::T:L:L:AQL::NA::NYMRS:QIK:::VDS
E. coli                          :AK:SMKA:EV::VA:AD::FA::AEL::::IS:V:ASDED:WNAVL::QT:::D::PS:Q::::RQ:
Liverwort    cp                  :AKKSL:Q:EK::QN:EK::KIL:NSL:KKITETSSLDEKW-:FQK::QS:::::AP::LHR::FL:

                 GRPRSVYKLFRISRIVFRELASKGSLIGINKSCW     99
                 :::::::E:::::::::::S:::R:P:M::K::S:   100   69.7%(69/99)
                 :::::::E:::::::::::G:::R:i:M::K::S:    99   70.4%(69/98)
                 ::::::SEF::::::::::G:::R:::M:::::S:   100   68.7%(68/99)
                 ::::::SEF:::Y:::::G::YR:::M::K::S:    100   64.6%(64/99)
                 :HA:F:LSD::LC:YQ:::N:L::N:P:VK:GI:    114   37.4%(37/99)
                 :::HGFLRK:GL:::KV::A:MR:QIP:LK:G       99   43.3%(42/97)
                 :::KAN:RD:GL::HLL::M:HACL:P:VT::S:    100   38.4%(38/99)
```

(l) *rps19*

```
Liverwort  mt  MTR-SIWKGPFVDT--------CLFKQKKIRWRIWSRRSCILPQFVGCYAQIYNGKGFVGLKITEEMVGHKFGEFASTRK
Petunia    mt  :P:R:i:::S:::AFLLRMKKKR---DLLFNRK::::::S:i:E::D:FVR:::::Ti:RC::::GK:::::::::F:::
Oenothera  ps  :P:R:::::S:::AFLLRMKKKR---DLLLNRK::L:::FS::LE:i:N:SVR:::::TP:RC:::::K::::::*:::F:::
E. coli        :P:-:LK:::::I:LHLLKKVEKAVESGD:KPL:T:::::T:F:NMI:LTIAVH::RQH:PVFV:D::::::L:::::P::T
Liverwort  cp  :::-::K::::::ADHLLKKIENLNL:KE:KIIIT:::A:T:V:TMI:HTIAVH::QEHLPIY::DR:::::L:::::P::T

               TSSLGKRALPSKTKIKPIKKVR   93
               RRPSRTNIG:GRKRG:K        94   54.1%(46/85)
               RRCSRTNIGLGRKRG:K        94   49.4%(42/85)
               YRGHAADKKAK:K            92   41.7%(35/84)
               FRGHA:NDKK:RR            92   42.9%(36/84)
```

Fig. 1. (cont.)

(m) *rpl2*

```
Liverwort   mt                     MRNSCWKGKALKQLTFHLKRNSAGRNSSGRITVFHRGGGSKRLQRKIDFKRSTSS-MGIVER
E. coli          MAVVKCKPTSPGRRHVVKVV:PELHKGKPFAPLLEKNSK:G:::NN::::TR:I:::H:QAY:IV::::NKDGIPAV:::
Liverwort   cp   MAIRLYRAYTPGTRNRSVPKFDEIVKCQPQKKLTYN:HIKK:::NR:I::SQ:::::H:::Y:::::Q:NKKYIT:KIKT

                 IEYDPNRSSWIALVRWIEGVLRPGKRLAFSKANSRREKNMPFFGLLFSPSSLPRQAQRIKYEKTRALRPCEQILESSWVL
                 L:::::::AN::::LYKD:ER:-----------------------------------------------------------
                 :::::::NTY:C:INYED:EK:-----------------------------------------------------------

                 GTRDLRAKEVSLGPLGSFLGLPSIAVAGAKPAFFAFRMKGPSSLTGRERLSPLRGENTFSQSEGQRWKTQSGAPRRKSLV
                 --------------------------------------------------------------------------------
                 --------------------------------------------------------------------------------

                 LSWSQGPKARNGLMISAHDIGKKDRRPEMAGPHTIPEHAPRALHAVGPSGSGRVLRTSEPFTYILASENLEVGNTVMNFH
                 --------------------------------------------------------------------------------
                 --------------------------------------------------------------------------------

                 GSKPSTLLNYHQPSQKANDPSGLRVEETAWDSQAWLHPRGDYASSENKYILDSYYQMVGNCIPLAKIPIGTWVHNIERNP
                 -------------------------------YI:A:K:LK:GDQIQSGV:AAIKP-::TL:MRN::V:ST:::V:MK:
                 -------------------------------YI:Y:R:IKLDDTIISSEEAPILI-::TL::TNM:L::AI::::IT:

                 GQGAKLTRAAGTFAQIIQKVENTPQCIVRLPSGVDKLIDSRCRATIGIVSNLNHGKRKFNKAGQSRWLGRRPIVRGVAMN
                 :K:GQ:A:S:::YV::VAR-DGA-YVTL::R::EMRKVEAD::::L:E:G:AE:ML:VLG:::AA::R:V::T:::T:::
                 :K:GQ:V:::::V:K::A:--:GQ-LVTL:::::EIR::SQK:L::::QIG:VDVNNLRIG:::SK::::K::K::::V::

                 PVDHPHGGGEGRTKGGRPS-VSPWGKPTKGGFKTVVRKRRN            501
                 :::::::::::NP:KH:--:T:::VQ:::K-::RSN::TDKFIVRRRSK    273   44.8%(112/250)
                 :I:::::::::API::KKPLT:::H:AL:KRSRKNN:YSDTLILRRRKNS   277   43.2%(108/250)
```

(n) *rpl5*

```
Liverwort   mt   MFSPNRNRLEFHYNQVIRPDLLLKINYENIMEVPRLCKIIV---VPKAPSNF---IKNVKLAMEIVCGQKFIQTRSRGST
Arabidopsis mt        MFP:N:::ED:L:Q:::::L::A:V::::G::E:R:---:::::Y::--I:::G::::::P::::::::Q-::::
Oenothera   mt        MFP:N:::ED:L:Q:::::L::A:V::::G::E:RI---:::TS:TYDFI:::G::::::LR:::::::E-::::
Brassica    mt        MFP:N:::ED:S:Q:P::NRITPTLWKFLGS:E:R:---::NG:Y::--I:::G::::::PR:::::::Q-::::
E. coli               MAK:HDY:KDEVVKK:MTEF::NSV:Q:::VE::TLNMG:GE:IADK-KLLD:AAADLAAIS:::PLI:KA:K:V

                 GKSFRFNKFVLNQESKRDTGYVTYLAR-STLRGHIMYNFLEKLVTIISFYDYPVKIQKNSIQL-----SMATSLLRLFPE
                 :::::S:P:---LG:NK:K:::SD:::Q::::::G:S:::VRIL:VM:LL:F::E:R:::::F-----::E:EFCEFS::
                 ::::::S:P:---L::NK:K:::SD:::Q:I::::G:S:::VRIL:VM:LL:SW:E:WE::::F-----::E:EFCEFS::
                 ::::::S:P:---LG:NK:K:::SD:::Q::::::G:S::SVRIS:VM:LL:F::E:R:::::F-----::E:EFCEFA:Q
                 AG---:-:-------I:Q-::PIGCKV--::::QR:WE:F:R:I::AVPRIRDFRGLSAKSFDGRGDY::GVREQII:::

                 IQDHFEIFEHIRGFDVTIVTSANTQDETVILWSGFLQKEV          188
                 LE::::::::::::N:::I:::::::::LL::::::::DEGETQ      185   68.7%(123/179)
                 LE::::::::::R:N::::::::::::::LL::::::::DEGESFKWKTP   192   64.8%(116/179)
                 L::::::::::::::N::::::::::::::LP::::::::DEGETQ      185   60.9%(109/179)
                 :-:YDKVDRV-::L:I::T:T:KSDE:GRA:LAA:DFPFRK          179   28.6%(52/182)
```

(o) *rpl6*

```
Liverwort   mt                                                                                 ME
E. coli          MSRVAKAPVVVPAGVDVKINGQVITIKGKNGELTRTLNDAVEVKHADNTLTFGPRDGYADGWAQAGTARALLNSMVIGVT

                 AKFFCFLEIIGVGYKASTNAQGSILYLKLGFSHEIRLQVTPSVRVFCLKPNLICCTGMDHQKVTQFAAIVKSCKPPEVYK
                 ED:TKK::LV::::R:AVK--:NVIN:S:::::PVDH:LPAGITAE:PTQTE:VLK:A:K:VIG:V::DLRAYRR::P::

                 GKGIQYRNEIIHKKQGKKK    101
                 :::VR:AD:VVRT:EA:::    177   36.6%(37/101)
```

Fig. 1. (cont.)

(p) *rpl16*

```
Liverwort   mt                                              VLYPKRTKPRKYQKGRCK-GCKADGTQLCFGKYGIKSCE
Maize       mt   MEKHLVMYLTRKSIMLLRKYLLVTEFQVSKCGSHIVKIRRD:::::::YS::S:C::SR:REP::::G::R::T::SR
Oenothera   mt                                          :::::::YS::R::::SR::EP:::::S::R::::::R
Petunia     ps                    MLLRKYLLVTES*VSKCGFLIVKKKRD:::::::YS::R::::SR:::P:::::G::R:::::R
Brassica    mt         MYLTRKSIMLLRKYLLVTESQVSKCGFHIVKKKGD:::::::---::S::::SR:::P:::K:G::R::T:::R
E. coli                                                     M:Q:::::::MH:::NR-:L-:Q::DVS::SF:L:AVG
Liverwort   cp                                              M:S::::::::QHC:NL:-:ISTR:NVI::::FPLQAL:

                 AGRISYQAIEAARRAISRK-F--------RRNSKIWVRVFADIPITSKPAEVRMGKGKGNTKGWIARVLKGQILFEMDCV
                 :::L::R::::::::TIGQ-:RRAMSGQF:::C::::::L::L:::G::::::::R::::PT::::::ST:::P::::G:
                 :::L::R::::::::IIGQ-:HRAMSGQS:::G::::::I::L:::G::T:::::R::::PT::::::ST:::::::G:
                 :::L::R::::::::::IGH-:HRAMSGQF:::G:::::W::::::G::T:::::R::::PT::::::SR:::::::G:
                 :::L::R::::::::TIGHS:RRAMSGQF:::C::::::L::L:::G::::::::R::::PT::::::ST:::P::I::G:
                 R::LTARQ:::::::MT:A-V--------K:QG:::I:::P:K:::E::LA:::::::::VEY:V:LIQP:KV:Y:::G:
                 PSW:TSRQ:::G::::T:Y-A---------::GG:L:I:I:P:K:::IR:::T:::S:::SPEY:V:V:KP:K::Y:ISG:

                 SLSNAQQAATLAAHKLGLSIKFFKWS            135
                 :::::R:::R:::::PCS:T::VQ::            185   71.9%(97/135)
                 :::::R:::::::::ICI:T::VQ::            144   78.5%(106/135)
                 :::::R::LH                            155   78.2%(93/119)
                 ::A::R:::R:::::PCS:T::VQ::            177   74.2%(98/132)
                 PEEL:RE:FK:::A::PIKTT:VTKTVM          136   50.4%(68/135)
                 :ENI:RA:MKI::Y:MPIRTQ:ITT:SLNKKQEI    143   45.9%(62/135)
```

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 deletions, introduced to maximize the matchings of the sequences. Amino acid residues corrected for RNA editing are underlined (Gonzaletz *et al.*, 1993; Hunt and Newton, 1991; Sutton *et al.*, 1993; Bock *et al.*, 1994; Zhuo and Bonen, 1993; Gualberto *et al.*, 1989; Wissinger *et al.*, 1990; Schuster *et al.*, 1990a; Brandt *et al.*, 1993; Schuster and Brennicke, 1991; Conklin and Hanson, 1991; Schuster, 1993). Arrowheads specify the sites of intron insertion.

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-rpl16-rpl5-rps14-rps8-rpl6-rps13-rps11-rps1*) similar to that seen in *E. coli* r-protein operons S10 (S10-L3-L4-L23-L2-S19-L22-S3-L16-L29-S17) (Zurawski and Zurawski, 1985), *spc* (L14-L24-L5-S14-S8-L6-L18-S5-L30-L15-*secY*-X) (Cerretti *et al.*, 1983) and α (S13-S11-S4-*rpoA*-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-*spc*-α 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 endosymbiont hypothesis, which postulated that the organelles of eukaryotes originated from prokaryotic (specifically eubacterial) ancestors in evolution (Gray *et al.*, 1989).

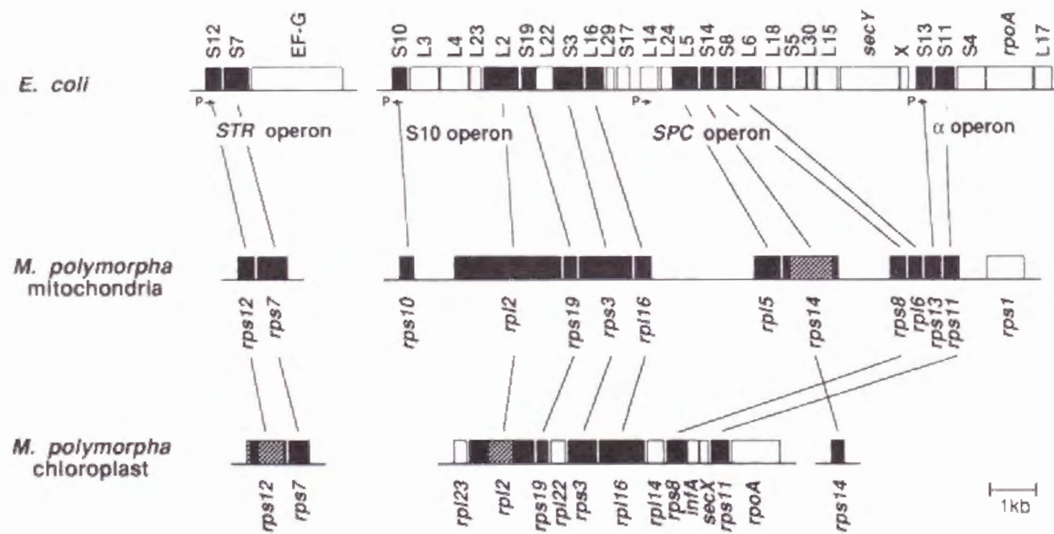

Fig. 2. Organization of r-protein genes in *E. coli*, liverwort mitochondrial and liverwort chloroplast genomes. Solid boxes indicate the common genes detected between the liverwort mitochondrial genome and either the liverwort chloroplast or *E. coli* genomes. Hatched boxes are genes having introns. Open boxes in *E. coli* and the liverwort chloroplast genome indicate genes that are absent in liverwort mitochondrial genome (except that the *rps1* gene appears only in the liverwort mitochondrial genome in this comparison).

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; Wissinger *et al.*, 1990; Wahleithner and Wolstenholme, 1988; Schuster *et al.*, 1990a). For example, *nad3* 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-*spc* 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 got 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,

1984) 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 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 the 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.

Protein	Liverwort mt	*E. coli*	Angiosperm mt	Liverwort cp
S 1	270	557	170	-
2	237	241		235
3	430	233	550-562	217
4	196	206		202
7	230	178	148	155
8	152	130		132
10	102	103	108	-
11	125	129		130
12	126	124	125	123
13	120	118	114-129	-
14	99	99	99-100	100
19	93	92	94	92
L 2	501	273		277
5	188	179	185-192	-
6	101	177		-
16	135	136	144-185	143

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*, *rpl16*, *rps14*, *rps8*, *rps11*, *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

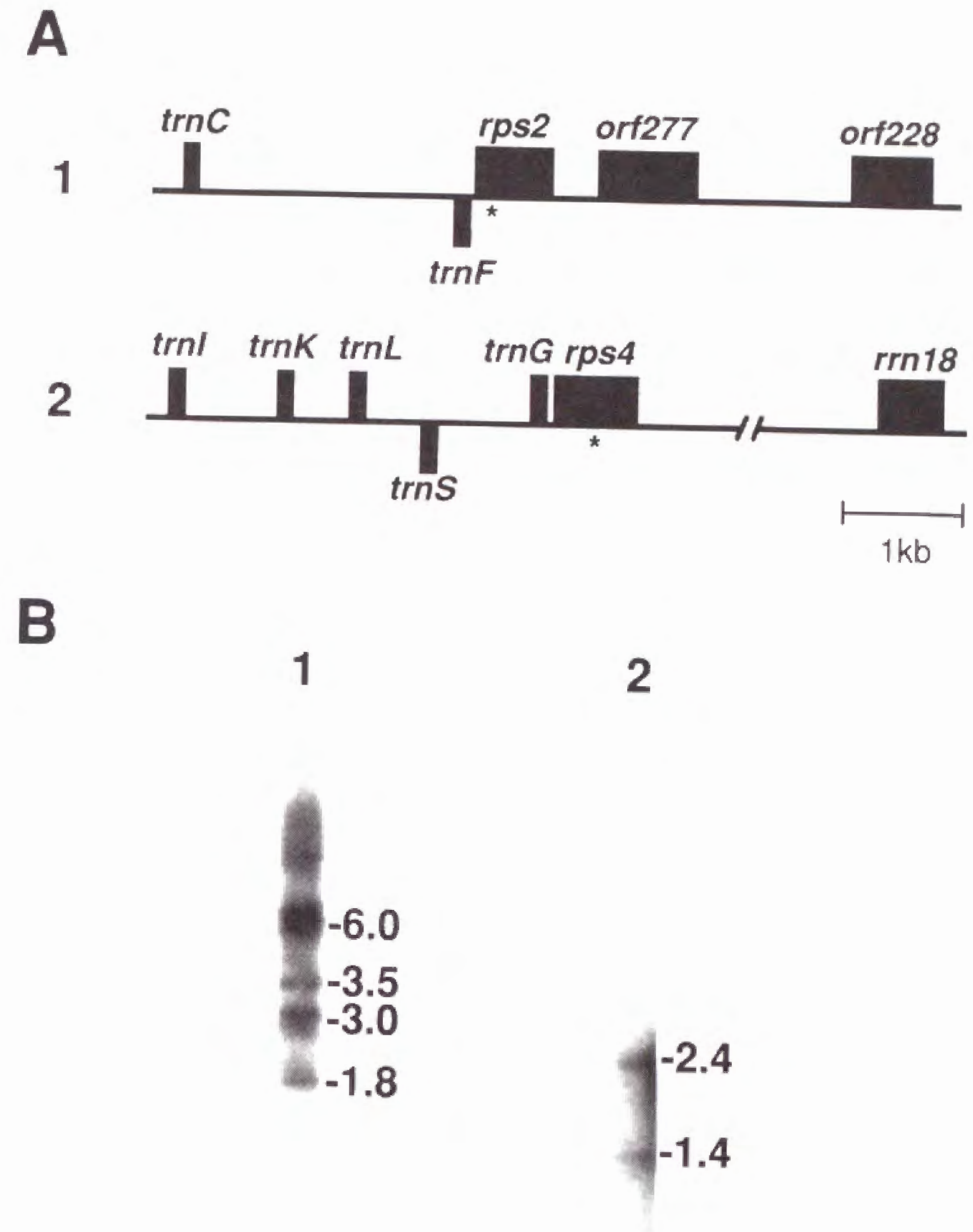

Fig. 5A, B. Transcription analysis of the liverwort *rps2* and *rps4*. **A** Gene organization of the liverwort *rps2* and *rps4* genes. Lane1, *rps2* and lane 2, *rps4*. The genes beneath the horizontal lines are oriented in the opposite direction to the genes above the lines. Each probes is shown by an asterisk under the gene organization. **B** Northern hybridization was performed by the probes specific to lane 1, *rps2* and lane 2, *rps4* genes. The size of each transcripts is indicated by a number given in kilobases (kb).

coding region of *rps2* (714 bp) and was too large to cover those of *orf277* and *orf228* located approximately 0.3 kb and 2.4 kb downstream from *rps2*, respectively. However, Northern analysis using the probes specific to both *orfs* revealed that there were no signals (data not shown), suggesting that these *orfs* may not be transcribed or may be transcribed at very low levels, and that *rps2* is transcribed independently of them.

When probed with a oligonucleotide complementary to *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 *rrn18* located about 4.6 kb downstream from *rps4*. In addition, such signals as 2.4 kb and 1.4 kb were not detected using a probe specific for *trnG* located 30 bp upstream of *rps4* (data not shown). These findings indicate that *rps4* is transcribed individually. The heterogeneity of transcript size might be resulted from RNA processing and/or multiple transcription intiation sites as reported in maize mitochondria (Mulligan *et al.*, 1988a; Mulligan *et al*, 1988b).

Co-transcription of the rps12 and rps7 in liverwort mitochondria

To study the expression of the *rps12* and *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 *rps12* and *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 *rps12* probe, but they were probably rRNAs 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 *rps7* (Fig. 6B, lane 2). The 7.0 kb transcripts were much larger than the coding regions of *rps12* and *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 *atp6* and *ψcob* which were located 1.1 kb and 2.4 kb downstream from *rps7*, respectively, were

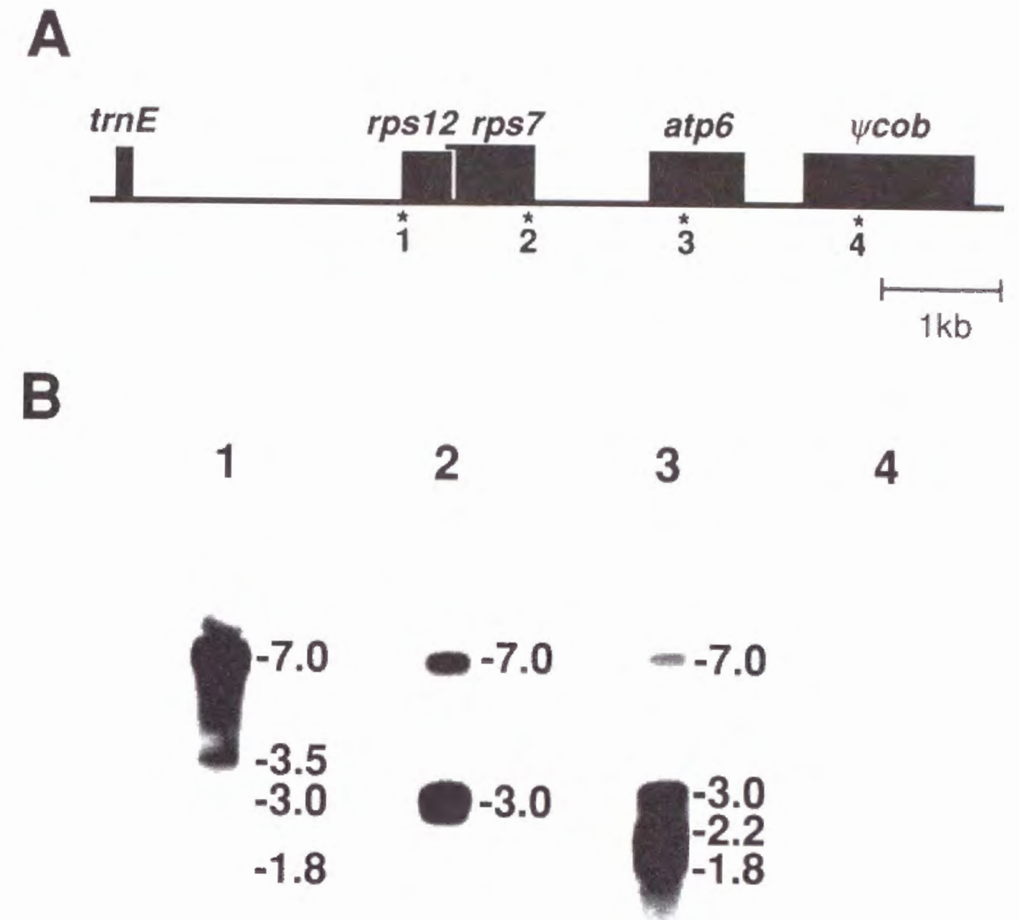

Fig. 6A, B. Transcription analysis of the region containing *rps12* and *rps7*. **A** Gene organization of the liverwort *rps12* and *rps7* genes. Each probes is shown by an asterisk under the gene organization. **B** Northern hybridization was performed by the probes specific to lane 1, *rps12*; lane 2, *rps7*; lane 3, *atp6*; lane 4, *ψcob* genes. The size of each transcripts is indicated by a number given in kilobases (kb).

carried out. Four bands of 7.0 kb, 3.0 kb, 2.2 kb and 1.8 kb were detected in the case of *atp6* (Fig. 6B, lane 3), while no signal was observed in the case of *ψcob* (Fig. 6B, lane 4). These results suggested that *rps12*, *rps7* and *atp6* but not *ψcob*

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 3.0kb 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 function 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-rps8-rpl6-rps13-rps11-rps1*) were analyzed. Using oligonucelotide 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 *ψmad7* located at 1.3 kb upstream of *rps10*, the 9.6 kb transcript was also found, demonstrating that *ψmad7* and *rps10* were probably co-transcribed (see also Chapter III).

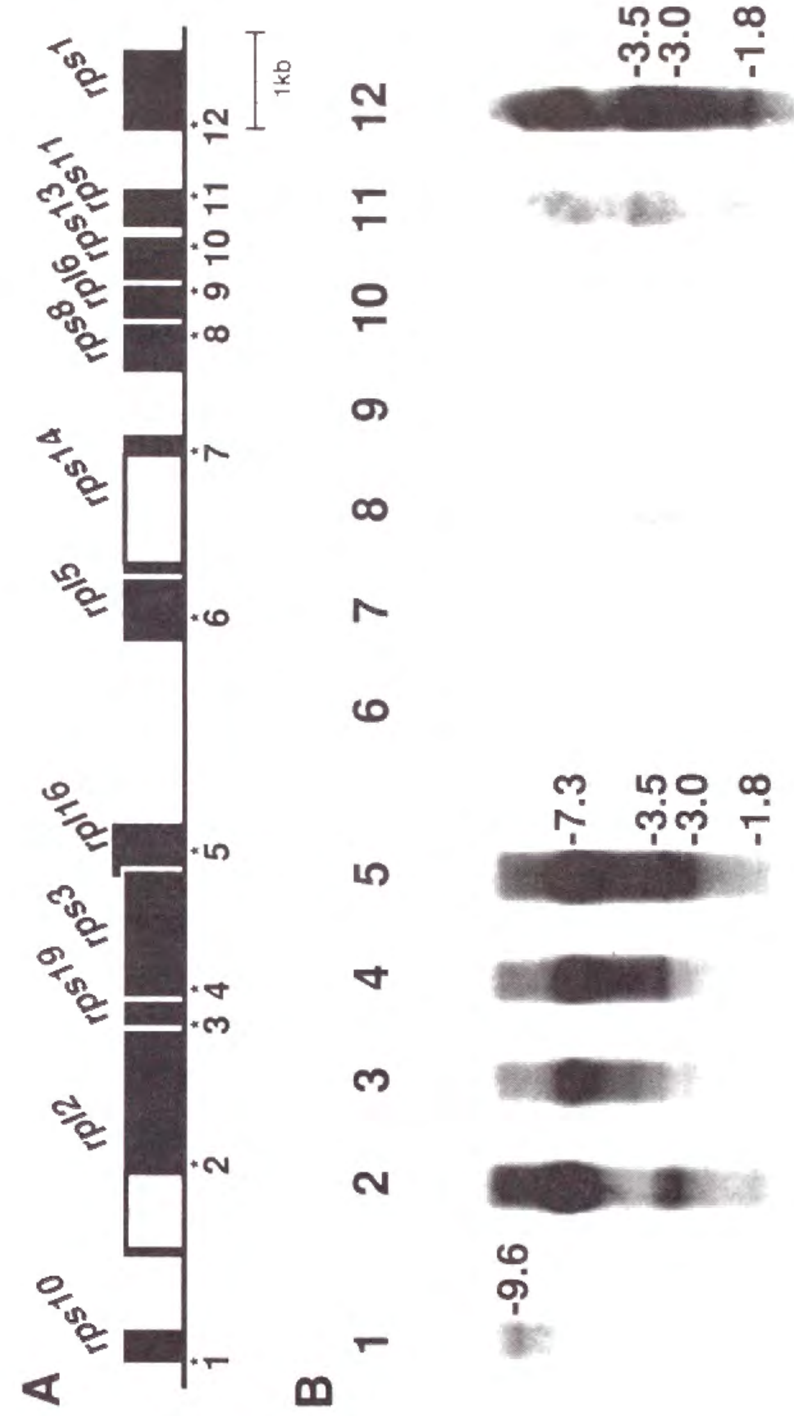

Fig. 7A, B. Transcription analysis of the liverwort ribosomal protein genes organized into a large cluster (*rps10-rpl2-rps19-rps3-rpl16-rpl5-rps14-rps8-rpl6-rps13-rps11-rps1*). **A** Gene organization of the large ribosomal protein gene cluster in liverwort. Filled and open boxes indicate exons and introns, respectively. Each probes is shown by an asterisk under the gene organization. **B** Northern hybridization was performed by the probes specific to lane1, *rps10*; lane2, *rpl2*; lane3, *rps19*; lane4, *rps3*; lane5, *rpl16*; lane6, *rpl5*; lane7, *rps14*; lane8, *rps8*; lane9, *rpl6*; lane10, *rps13*; lane11, *rps11*; lane12, *rps1* genes. The size of each transcripts is indicated by a number given in kilobases (kb).

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; Wahleithner *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 Grienenberger, 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*, *ndhI* 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.

Table 1. Genes for NADH dehydrogenase complex

	Genes for subunits									
	nad1	*nad2*	*nad3*	*nad4*	*nad4L*	*nad5*	*nad6*	*nad7*	*nad8*	*nad9*
Mitochondria										
Liverwort	+	+	+	+	+	+	+	pseudo	-	+
Angiosperms	+	+	+	+	+	+	+	+	?	+
Human	+	+	+	+	+	+	+	-	-	-
Yeast	-	-	-	-	-	-	-	-	-	-
Podospora	+	+	+	+	+	+	+	-	-	-
Chloroplast										
Liverwort	+	+	+	+	+	+	+	+	+	+

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 *Bgl*II-*Pst*I restriction fragment was isolated from a plasmid pLB104 (Oda *et al.*, 1992c) and labeled by [α^{32}-P]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'-GATCATAACGATATCGTGGAAATGCTGCGC-3' (Fig. 6A, 1)
nad3 : 5'-CATTCGTAAGCTGACAATTTCTCTGGATAAGCC-3' (Fig. 6A, 2)
nad4L : 5'-AATAGCGGATTCCGCAGCAGCCACCGTTAA-3' (Fig. 6A, 3)
nad6 : 5'-TATGTTAGATAGGTTGCACTCAATTTCGTTGGTAAT-3' (Fig. 6A, 4)
nad9 : 5'-ACCAAACATATCCCAAGTTTCTCGTTC-3' (Fig. 6A, 5)

nad5 exon1 : 5'-GCCCCAAGCAGCGTCAAAGAGTTCCGAAAA-3' (Fig. 7A, 1)
spacer between *nad5* and *nad4* : 5'-GAGGGGATGTGCGTTAAATAGACCTTCCGG-3' (Fig. 7A, 2)
: 5'-TCTGTTGACCCGGTGTGTTTTTGGCGAATT-3' (Fig. 7A, 3)
nad4 intron : 5'-GGTTCCAATCTAACTAACCGCGGTCGGACC-3' (Fig. 7A, 4)
nad4 exon2 : 5'-TAGGTGCCTCCACATGAGCTTCAGGTAACC-3' (Fig. 7A, 5)
nad2 intron : 5'-GCCGGATCCGCCTGGATCACCTGGAATGAT-3' (Fig. 7A, 7)
nad2 exon2 : 5'-CGGCCAGTGCTCCTAAGATCATAGAAGCAA-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 *frxB*), *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, *nad5*, *nad4*, and *nad2*, 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* (Schuster *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 psuedo-*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.).

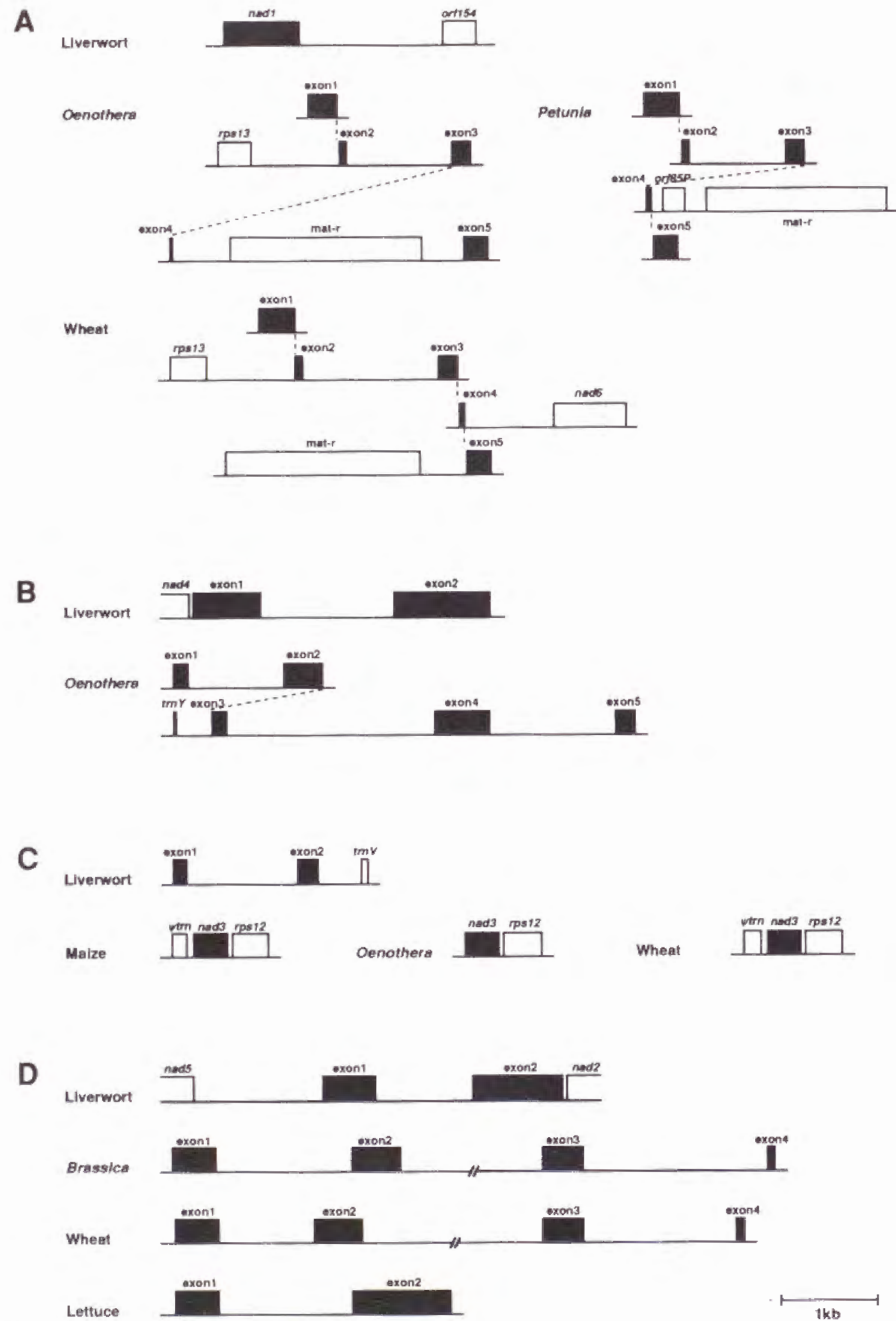

Fig. 1 (Cont.)

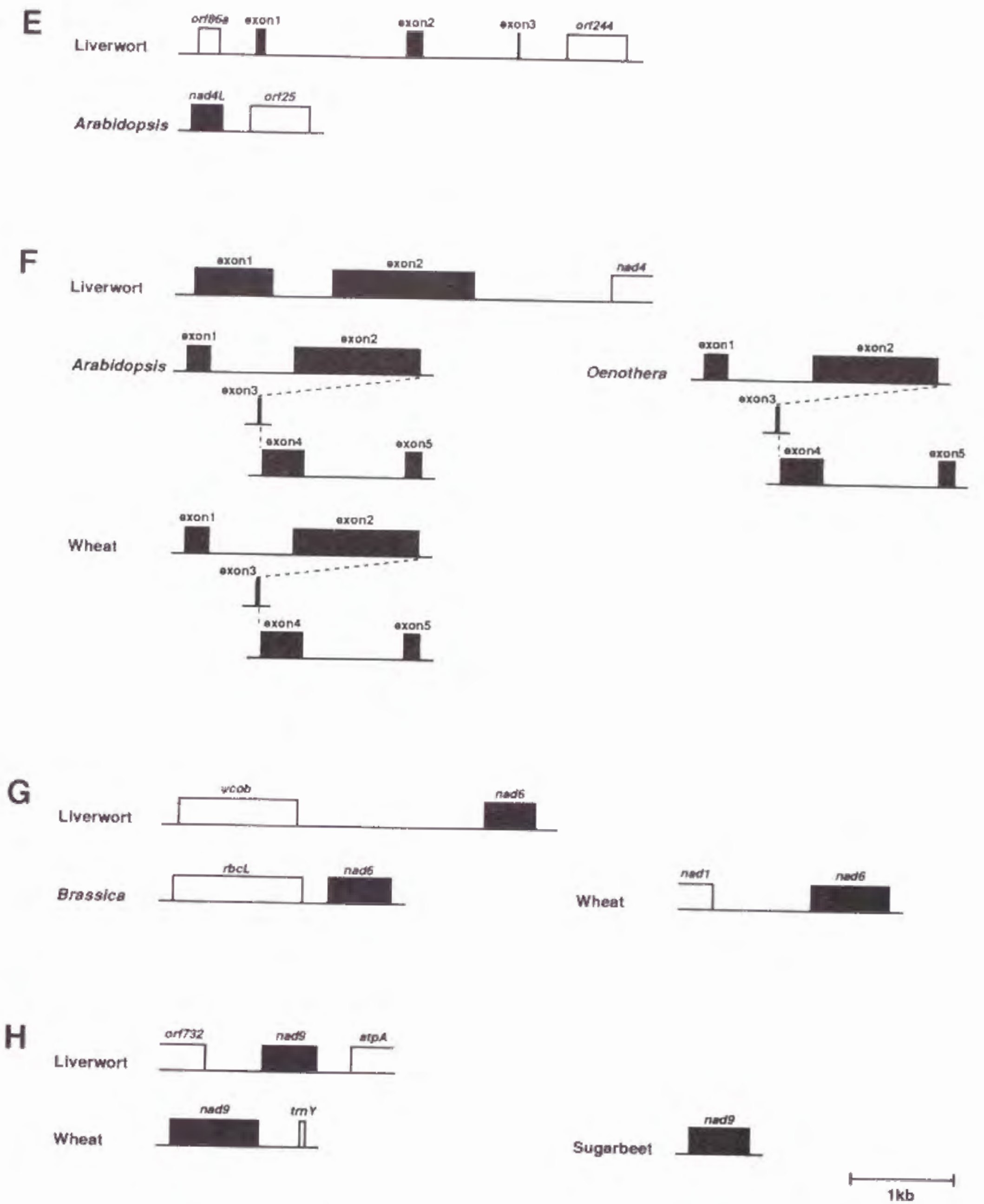

Fig. 1. Gene structures of *nad* genes in liverwort and higher plants mitochondria. The exons of the *nad* genes are indicated by solid boxes. A; *nad1*, B; *nad2*, C; *nad3*, D; *nad4*, E; *nad4L*, F; *nad5*, G; *nad6*, H; *nad9*.

Table 2. Numbers of introns in *nad* genes

	Genes for subunits									
	nad1	*nad2*	*nad3*	*nad4*	*nad4L*	*nad5*	*nad6*	*nad7*	*nad8*	*nad9*
Mitochondria										
Liverwort	0	1	1	1	2	1	0	(2)	-	0
Angiosperms	4	4	0	1-3	0	4	0	0	?	0
Human	0	0	0	0	0	0	0	-	-	-
Yeast	-	-	-	-	-	-	-	-	-	-
Podospora	4	0	1	1	1	4	0	-	-	-
Chloroplast										
Liverwort	1	1	0	0	0	0	0	0	0	0

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 (Bland *et al.*, 1986), maize (Bland *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 (Brears 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, sugarbeet, 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 *ndh9* (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

(a) *nad1*

```
Liverwort  mt                         MRIYLIGLV-AKILGIIIPLLLGVAPLVLAERKVMASMQRRKGPNVVGILGLLQPLADG
Wheat      mt                              MYIA:P:E::CL:L:::::::::::::::::::FV:::::D:::SF:::::::::
Oenothera  mt                        MNRKSKMYIAVP:E:::::L::::::::::::::::::::FV:::::D:::SF:::::::::
Petunia    mt                              MYIA:P:E:::::L::::::::::::::::::::FV:::::D:::SF:::::::::
Podospora  mt                          MY:S:IISLIEVVLVLV:A:::I:YVTI::::T:::::::L:::F::YY::::AF::A
Human      mt                                 MPM:NL:LL:V:I:IAM:::M:T:::ILGY::L::::::::PY:::::F::A
Liverwort  cp  MISNINLEDKFFSFFFTLGFSKEFFNFLWIIFS::ILMLGVTI::LV::WL:::IS:AI:Q:I::EYA:P::II:A::::

               LKLMIKEPILPSSANIFIFLMAPVLTFTLALCAWAVIPFDYGMVFSDLNIGVLYLFAISSLGVYGIITAGWAS-NSKYAF
               :::IL::::S:::::F:L:R::::A::M:S:V::::V:::::::L::::::L:::::::::::::::::I:::S:-::::::
               :::IL::::S:::::FSL:R::::A::M:S:V::::V:::::::L::::::L:::::::::::::::::I:::S:-::::::
               :::IL::::S:::::F:L:R::::A::M:S:V::::V:::::::L::::::L:::::::::::::::::I:::S:-::::::
               :::LL::YVA:TQ:::IL:FLG::I:LIPS:LGY::::YGPSLAIN:FSL:IY:IL:V:::AT:::LL:::G:A::::::
               M::FT:::LK:ATST:TLYIT::T:AL:I::LL:TPL:MPNPL:--N::L:L:FIL:T:::A::S:LWS:::::-::N::L
               I::FL::D:V:AQGDVWL:NIG:I:VLIPVFLSYL::::E:NVILANFS:::FFWI:V:::VVPL:LLM::YG:-:N::S:

               LGALRSAAQMVSYEVSIGLLLISVILCAGSCNFSEIVLAQTRI----WFVFPLFPVFLMFFISCLAETNRAPFDLPEAEA
               ::::::::::::::::::::::I::T:LI:V::::L::::M::KQ:----::GI:::::LV:::::::::::::::::::::::
               ::::::::::::::::::::::I::T:LI:V::::L::::M::KQ:----::GI:::::LV:::::::::::::::::::::::
               ::::::::::::::::::::::I::T:LI:V::::L::::M::KQ:----::GI:::::LV:::::::::::::::::::::::
               ::S:::T::LI:::LVLSSAILL::MLT::L:L:VNIES:RA:----:NI:::L:::II:::GSV:::::::::::A:::S
               I::::AV::TI::::TLAII:L:TL:MS::F:L:TLITT:EHL----:LLL:SW:LAM:W:::T::::::T::::A:G:S
               ::G::A:::SI:::IPLA:SVL:IA:LSN:LSTVD::E:::SKYGFLS:NLWRQPIG:IV:::AS:::CE:L::::::::E

               ELVAGYNVEYSSMGFALFFLGEYANMILMSSLCTLLFLGGW---------------LPILDIPIFQVIPGSIWFS-----
               ::::::::::::::::::::::::::::::::G:::::::::::---------------:::::L::SKK::C::::::-----
               ::::::::::::::::::::::::::::::::G::::::::::::---------------:::::L:::KK::::::::::-----
               ::::::::::::::::::::::::::::::::G:::S:::::::---------------:::::L:::KK::::::::::-----
               :::S:FMT:HAAVV:VF:::A::GSIV::CI:TSI:::::YLSINSLDVFNFFYSI:FNIGPIDLNFFNIFYY:YKEIFV
               :::S:F:I::AAGP:::::MA::T:I:M:NT:T:TI:::TT---------------YDA:SPELY-----TTY:VT----
               ::::::QT:::G:K::F:Y:AS:L:LLVS:LFV:I:Y::::---------------HFSIPFFSLFKNFEWNLM:NGISE

               ---------------IKVLFFLFVYIWVRAAFPRYRYDQLMRLGWKVFLPLSLA---WVVFVSGVLVAFEWLP
               ---------------::::L:::L::::::::::::::::::::G::R::::::::::---R::P::::S:T:R:::
               ---------------:::I::::L::::::::::::::::::::G:::::::::::::---:::S::::::T:Q:::
               ---------------:::I::::L::::::::::::::::::::G:::::::::::::---:::S::::::T:Q:::
               NNSIIEGLIYGLTIGL:SSILI:LF:::::S:::I:F::::AFC:T:L:::LF:---LI:LLPCI:YSYNI::VNVL
               ----------------:T:LLTSLFL:I:T:Y::F::::::H:L::N::::T::LLM:Y:SMPITISSIPPQT
               VISIIIGIVITL---V:SYL:::IS:MT:WTL::I:I:::LN::::FL::IA:G---NLLLTTSFQLFLL

               328
               325   82.7%(268/324)
               331   84.3%(273/324)
               325   84.6%(274/324)
               368   47.5%(154/324)
               318   46.4%(149/321)
               368   39.4%(128/325)
```

(b) *nad2*

```
Liverwort  mt                 MFEHDFLALFPEIFLINATIILLIYGVVFSTSKKYDYPPLVRNVGWLGLLSVLITILLVAVGSPL-AV
Oenothera  mt                 ::NL-:::V:::::I:::::F::::H:::::::::D::::::S:::::::::::::::L::L:A:A::LTI
Sugarbeet  mt                           MLS:TIKGKARRRKERA:GDRDFLTFSSKTKKTENVN:SFEKG:RFFDRG:MIPGPS
Podospora  mt  MIFISIIGLLLSNAVTLRQDMSVNFNR:ALIDLIYCILHDTMSLSIINKGIGLHGGLLHITNIT::FHIFIFFL:I:-IL
Human      mt       INPLAQPVIY-STI:AGT-::T:L------------------------------------------------
Liverwort  cp      MKLELD::FLYGSTIL::CI::FSLL:I::IDLT:PK-:DTIWLYFI-SLTS:-:I:II:-L:FQYKTD:I--I

               ANLVYNNLII-DNFTYFCQIFLLLSTASTMVMCLDYFKQESLNAFESIVLILLSTCSMLFMISAYDLIAMYLAIELQSLC
               :H:FW::FFRR:::::::::L::::::G:IS::F:F:E::RFD:::F:::::::::::::::::::::::::::::::::::
               PRSARWPIG:AAFGLCLLFLIKNSGS:RESAGNNRKEGVHVAA:SAPFLVNRAAGSATTTKERIHPK:TNAS:MAACGMA
               QLTSFYPRKA-WIPEHSSLKDIIYQKFLNYRTKIFNKMG:HMKII:YPLIL:FVISGAV:LM:TN::VSIF:S:::::YG
               -----------------------------------------------------------SSHWFFTWVGL:MNM:-
               S-FL-GSFQT-:S:NRIF:S:IVFCSILCIPLSIE:I:CAKMAIP:FLIF::TA:VGGM:LCG:N::VTIFVSL:CL:::
```

Fig. 2. (Cont.)

```
FYVIAA-SKRDSEFSTEAGLKYFILGAFSSGILLFGCSMIYGFTGVTNFEELAKIFTGYEITLFGAQSSGI---------
:::L::-:::K:::::::::::L:::::::::::::::::::::S::A:H:DQ::::L:::::::--::R::::---------
GSDLFG-YIIQV:-:GVT:TAGLMENN:HGSVQRALF:LRILRSL-RV-NS::R:QNFW-GPSIPSS:PAK---------
L:LLSTI-Y:N::L::AG::I::L::GL::CFI:L:T:LL:INS:T:SLDG:YILNSISDVKDGA:DMPALTSWYKSYYL
AFIPV-LT:KMNPR::::AI:::LTQ:TA:M:::MAILFNNMLS:-QW-TMTNT-TN--------QY::----------
S:LLCGYT:::IR-:N::AI::LLI:GT::S::AY:F:WL::LS:-GE-TNIQ::TN:L-LNAETYN:::T---------

FMGILFIAVGPLFKITAVPFHMWAPDVYEGSPTIVTAFFSIAPKISILANMLRV-F-I-YSFYDPTWQQLFFFCSIASMI
:::::::::::::::::::::::::::I::::::P::::L::::::::F::I::::-:-:-:GS:GA:L::I::::::::::
------TPLP:GLN:FPDSY-:::::I::::::P:::::::::ER::S::I::::-:-:-:GS:GA:L::I:::::::LRL
NFAL:VFSI:::::VS:A:::F:S::::DAI:::::T:VA:IA::::FIFL:ELVYHTNNYLSEFS:TY:LLIS:LF:L:
----:M:MMAMAM:LGMA:::F:V:E:TQ:T:---------LTSGLLL:TWQKLAPIS:M:QISPSLNVS:LLTL::L:IM
:IAFIC:L::LA::LSL::::Q:T::I:::::::P:V::L:VTS::AG::LAT:I-LN:LF::SPNE:KIFLEILA:L:::

LGALAAMAQNKVKRLLAYSSIGHVGYLLIGFSCGTIEGIQSLLIGIFIYVLMTVNVFA-IVL-ALRQNRFKYI----ADI
::::::::::T:::::::::::::::::::IC::::::::::::::::::::L:::A:T:I:A::-:::-::::Y:V:::----:::
RSTG-:::NEGKASSSIGQLDYGGL:FVLVLMWNR-::::::::::L:::AS:DDRC::-::S-::::T:V:::----:::
I:TVVGLT:FRI:::::::T:S:::FI:LAL:GCS::ST:AFIFYLIQ:SISNL:::I-:II-TIGFSLYG::TTNKEYL
A:SWGGLN:TQLRKI::::::T:M:WMMAVLPYNPNMT:LN:T:Y:------ILTTT:FLL:N-:NSST-TTL----LLS
::N:V:IT:TSM::M::::::SQI::I:::LIT:DLK:YT:MT:YV:F:IF:NLGT::C:I:YS::TGT-DN:----R:L

--LGALAKTNPILAITLSTMFSYAGIPPLAGFCSKFYLF-FAALGC-GAYLLALI-GVVTSVISCFYYIRFV----KIMY
--:::::::::::::F:::::::::::::::::::::::-:::::::-:::F::SV-:::::::G::::::L:----:R:F
--::::::::::S:::F::::::::::::::::::::::-:::::::-:::F::PV-:::::::G:WAAG:LP----RVSQ
--KDL:D:N:SPIQVISQKGYF:--:N::LSLSLAITI:S:VGIPPLVGFFAKQM-VLSAALDNGYIFLTLI----A:LT
RTWNK:TWLT:LI-PS-T-LL:LG:L:::T::LP:WAII-EEFTKN-NSLIIPT:MATI:-LLNLYF:L:LIYSTSITLL
--YAG:YIKD:L:SFS:TCLL:LG:L:::T::FG:L:::-WCGWQS-:F:::VF:-ALI:::::LY::LKII----:LIL

FDTPKTWVLY-KPMDREKSLLLAITVF-FITFFFLYPSPLFLVTHQMALCLCL
::::R::I::-E::::N::::::M:SS-:::L::::::::::S:::::::S:Y:
:GDRRQFSVH-RTRSLPNQ:RHGWECM-LRKIGSSLIHQPSVYSISLYESTITTRDEPWFGEFELALGVIGLPVTAHDRI
SVIGAVYY:N-IIKKIFFY:PDHSINP-S:GE:LFKKGLI:EAGDFKGRITLISSPFSITISIITLVILLFIFMNKEWLS
PMSNN-VKMKWQ-FEHT:PTPFLP:LIALT:LLLPI-::-:MLMIL
TKKNNEINP:IQAYIITSPTFFSKNPIE:VMI:CVLG:T-::GIIINPIFSFFQDSLSLSVFFIK

                         489
                         488   83.2%(405/487)
LRCSPPVVGTTRAGPGLNSER    515   32.1%(151/471)
MGTILVQVLFSN             556   23.0%(112/487)
                         347   23.5%(77/328)
                         501   34.0%(163/479)
```

(c) *nad3*

```
Liverwort  mt    MEFAPIFVYLVISLLLSLILIGVSFLFASSSSLAYPEKLSAYECGF-DPFDDARSRFDIRFYLVSILFIIFDLEVTF
Wheat      mt   ML:::::CI:::::::V::::L::P:::::N::-T::::::::::::::-:::G::::::::::::::::::::::::::
Oenothera  mt   ML:::::CIS::::::::::LV:P:::S:N::-T:::::::::::::-:::G::::::::::::::::::::::::::::
Petunia    mt   MS:::::CI:::::P:V:::PL:LP:::S:N::-T::::::::::::::-::SG:::::::::::::::::::P:P::::
Podospora  mt  MSSMTLP:LPVSI:A::FLP:-----N:IFAPHN-P:Q::Y:IF::::HSFLGQN:TQ:GVK:FIFALVYLLL:::ILL
Human      mt     :N::L:LMINTLLA::LM:I----TFWLPQLN-G:M::STP::::::-::MSP::VP:SMK:F::A:T:LL::::IAL
Liverwort  cp  MFLLQKYDY:FVFLLIISFFS:::FSLSKWIAPIN-KG:::FTS::S:I-E:MGE:CIQ:Q::Y:MFALV:V:::V:TV:

               LFPWAVSLNKIGLFGF-WSMMVFLFILTIGFVYEWKKGALDWE                   118
               F::::::::::D::::-:::::A::L::::::L::::R:::::                   118
               F:::::F:::D:::::-:::::A::L::::::L::::R:::::                   118
               S:::::PP:::DP::S-:::::A::L:::::SL::::R::S:::                  118
               T::F:::EYVNNIY:L-IILLG:IT:I:::::::LG:S::KIDSRQVITMTRFNYSSTIEYLGKI  137
               :L:LPWA:QTTN:PLMVM:SLLLII::ALSLA:::LQKG:::TE                  115
               :Y:::M:FYNF:ISS:-IEALI:IL::I::L::A:R::::E:S                   120

               87.2%(102/117)
               84.6%(99/117)
               76.1%(89/117)
               33.9%(38/112)
               37.2%(42/113)
               35.9%(42/117)
```

Fig. 2. (Cont.)

(d) *nad4*

```
Liverwort  mt                               MLQVLAPFYSNLSGLILLPLLGSLIILVIPNSRVRLIRGITIWTSLITFLYSLF
Wheat      mt                               ::EHFCEC:FD::::::C:V:::I:L:F::::SI:::::GLCV::::::::::V
Brassica   mt                               ::EHFCEC:F:::::::C:V:::I:L:F:::::I::::L:GLCA:::::::::V
Lettuce    mt                               ::EHFCEC:FD::::::C:V:::I:L:F:::::I::::L:GLCA:::::::::V
Podospora  mt  MSGLLYALLIIPMIGIFFILSFDSYNFNITSNNSNSG:FSEAGAGKNSG:E:LKVLVI:NELNFYKK:AFI:TIMNLIV::I
Human      mt                                              MLK:IVPTIMLLP-LTWLSKKHMIW:NTT:HSLIISIIPL
Liverwort  cp                                       MNHFPW:TIIV:F:ISAG:V:PFL:STGNKI::WY:LGVC:LE::LITY

               FW-IRFENDTAKFQFVETIRWLPYSNINFYIGIDGISLFFVILTTFLTPICILVGFYSVKSYKKEYMIAFFICESFLIAVFC
               L:-:Q:DPS:::::::SL:::::E::HL:M::::L:::::::::::I::::::::WSGMR:FG:::I:::L:::FLM:::::
               L:-:Q:DSS:::::::SL:::::E:::::L:::::::::::::::::I:::::::WSGMR::G:::I:::L:::FLM:::::
               L:-:Q:DPS:::::::SL:::::E::H::L:::::::::::::::::I:::::::WSGMR::G:::I:::L:::FLM:::::
               IY-:L:DFS:NQ::::QDA--:EL:VY:I:L:V::L:IY::L:::IIM::ALISNWN:ITHNI:A:L:IILLL:TL:L:::L
               LF-FN-QINNNL:SCSP:FSSD:LTT-PL-LMLTTWL:PLT:MAS-QRHLSSE--PL:R:---:L:LSMLISLQIS::MT:T
               IFCYHYQFNDHLI:LK:DYN:ISFI:FHWRL::::F:IGLIL::G:I:TLAT:AAWPVTRNPRLF:FLMLAMYSG-Q:GL:A

               SLDLLIFYVFFESVLIPMFIIIGVWGSRQRKIKAAYQFFLYTLMGSLFMLLAILFI-FFQTGT-T-DLQILLTTEFSERR-Q
               M::::L::::::::::::::::::::::::::::::::I:::L::V:::::::L:-LL::::-:-::::::::::::::::-:
               M::::L:::::::::::::::::::::::::::::L:::::L::V:::::::L:-L:::::-:-::::::::::::::::-:
               M::::L:::::::::::::::::::::::::::::L:::::L::V:::::::L:-L:::::-:-::::::::::::::::-:
               V::V:L::I::::I:P:L::L::LF::SN-:VR:SFYI:::::L::::L::S::TM-SSLI::-::-Y:DV:SKSN:EYTT-:
               ATE:IM::I:::TT:::TLA::TR::NQPERLN:GTY:LF:::V:::PL:I:--IYTHNTL:SLNIL:LT:TAQ:L:NSWAN
               :Q:I:L:FFMW:LE:L:VYLLLAM::GK-:RLY::TK:I:::AA::::I:IGG:IMA:YNSNEF:F:F:F:INKKYPLEL-E

               ILLWIAFFASFSVKVPMVPVHIWLPEAHVEAPTAGSVILAGILLKLGTYGFLRFSIPMFPEATLYFTPFIYTLSVIAIIYTS
               :::::::::::A:::::::::::::::::::::::::::::::::::::::::::::::::C:::::::::A:::::::
               :F::::::::A::::::::::::::::::::::::::::F::::::::::::::::::::::C::::::::A:::::::
               :F::::::::A:::::::::::::::::::::::::::::::::::::::::::::::::::C::::::::A:::::::
               :F:FLGV:IA:A::T:T:FLNS::LK::::S:LG::IV::::V:::SL::IF:LIL:LL:KIS:NY:SI:FSIGI:T:::A:
               N:M:L:YTMA:M::M:LYGL:L:::K::::::I:::MV::AV:::::G::MM:LTLILN:LTKHMAY::LVL-:LWGM:M::
               :IIYLS:LIAYA::L:II:F:T:::DT:G::HYSTCML:::::::M:A::LI:INMELL:H:HSF:A:WLVIVGA:Q:V:AA

               LTTIRQIDLKKIIAYSSVAHMNFVTIGMFSLNIQGIEGSILLMLSHGLVSSALFLCVGA-LYDRHKTRIVKYYGGLVSTMP-
               :::L:::::::::::::::::::L::::::::::::G::::::::::::::::::::::V-::::::::L:R::::::::::-
               :::L:::::::::::::::::::L::::::::::::G::::::::::::::::::::::V-::::::::L:R::::::::::-
               :::L:::::::::::::::::::L::::::::::::G::::::::::::::::::::::V-::V:::::L:R::::::::::-
               FS:L:T::I:EL::::::S:AAVYL::V::NI::::::G:::G:G::F:::G::I:A:GV::::SG:::IS::R:TAQV::-
               SICL::T:::SL:::::IS::AL:VTAILIQTPWSFT:AVI::IA:::T::L::CLANS-N:E:THS::MILSQ::QTLL:L
               ::SLS:RN::RR::::::S::G::L::IG:ITNL:LN:A::Q:I:::::IGAS::FLA:I-S:::TR:LVLDQM::IGNS::K

               I-FSTIFLFFTLANMSLPGTSSFIGEFLILVGAFQR-N-S-L--VATL--AAL-GMILGAAYSLWLYNRVVFG-N--FKPNF
               N-:::::F::::::::::::::::::P:::::::-:-:-:--::::--:::-:::::::::::::::::::S:-:--L::D:
               N-:::::F::::::::::::::::::::::::::-:-:-:--::::--:::-:::::::::::::::::::S:-:--L::D:
               N-:::::F::::::::::::::::::::::::::-:-:-:--::::--:::-:::::::::::::::::::S:-:--L::D:
               L-::IL:FILC:G:CGA:L:LN:V:::MS:Y:T:E:-L-P-:--LGLF--SSS-S::FS:::TIYM:::IA::GS--:SKF:
               MA:-WWL:A-S:::LA::P:INLL::LSV::TT:SWS:ITL:LTGLNMLVT::YSLYMFTTTQWGSLTHHINNMKPS:TREN
               :-:-:L:TSCSM:SLA:::M:G::A:LM:FL:VIDNP:Y:S:FKIIIIIIQGI-:I::TPI:L:SMLRQMFY:YK--:SNTL

               ILKFSDLNRREVLIFLPFIVGVIWMGVYPEVFLECMHTSVSNLVQHGKFD    495
               LY::::::G:::F:::::L::::V::::::K:::D::::::::::::::::::    495   84.6%(419/495)
               LH::::::G:::F::I::L::::V::::::K:::D:::::::::::::::::H    495   85.3%(422/495)
               LH::::::GT::F::I::L:::V::::::K:::D:::::::::::::::::H    495   84.6%(419/495)
               EENIG:VTK::FFLLFTL:IFT:MF:I::SFI:DGL:YN:TS:LFGV          519   41.1%(202/492)
               T:M:MH:SPILL:SLN:D:ITGFSS                                459   30.6%(135/441)
               EPY:M:AGP::IF:LICLFFPI:SI:I::NFV:SIWNSK:NP:LSNNF:        499   33.3%(162/486)
```

(e) *nad4L*

```
Liverwort    mt   MDLVKYLTFSMILFLLGIWGIFLNRKNILIMLMSIELMLLAVNLNFLVFS--VYLDDMMGQLFALFVLTVAAAESAIGLA
Arabidopsis  mt   :::I::F:::::I:I:::::::L:::R:::::::::::::::::S:::::::--:S:::::::V:::L::::::::::::::
Podospora    mt         MNITL::::I::L:FV::::::IL::I::::I::::ITFLI::S:--LNM::II::TY:IYIIV::G:::::::G
Human        mt     MPLI:MNI-:LA:TISLL:MLVY:SHLMSS:LCL:G:M:SLFIMATLMT--LNTHSLLANIVPIAM:VF::C:A:V:::
Liverwort    cp     M:EHI::L:AF::CI:VF:LITS:-:MVRA::CL::IFN:::I:LVA::NFLDSSQIK:EI:SI:IIAI::::AT::::
```

Fig. 2. (Cont.)

```
ILVITFRIRGTIAVEFINCMKG   100
:F:::::V::::::::::SIQ:   100   85.0%(85/100)
:::APY:L::S:SI:YK         89   52.8%(47/89)
L::SISNTY:LDY:HNL:LLQC    98   22.4%(22/98)
:VLAIY:N:KSTRIDQF:LL:W   100   36.7%(36/98)
```

(f) *nad5*

```
Liverwort    mt              MYLLIVILPLI-GSFAAGFFGRFLGSR-GVAVVTTTCVSLSSIFSCIAFYEVALCASACYIKIAPW
Wheat        mt              ::::::F:::L-::SV::::::::::E-:T:IM::::::F:::L:L::::::::G:::::LR::::
Oenothera    mt              ::::::F:::L-::GV::::::::::E-:S:IM::::::F:::L:L::::::::G:::::L:::Q:
Arabidopsis  mt              ::::::F:::L-::SV::::::::::E-:S:IM::::::F:::L:L:::::::PG:::::LR::Q:
Sugarbeet    mt              ::::::F:::L-::SV::::::::::E-:T:::::::::F:::L:L:::::::PG::::::R::::
Podospora    mt              :::S:I::::L-:CIVS::L::KV:VK-:AQLI:CSN:VITT:L:IL::I::GFNNIPVT:NLFR:
Human        mt       MTMHTTMTTLT:TSL:P:ILTTLVNPNKKNSY--PHYVKSI:AS:FI-I:-L:-PTTMF-MC:DQEVIISN-WH:
Liverwort    cp  MELIFQNVWFVPLFPF:ASIL:GIG-LF:FPNSIKK:--R:-LSSFISIMFLNIAMLL-SFH:FWQQITG:PIHRYLWS:

IFSELFDAAWGFLPDSLTVILLLVVTIVSS-LVHIYSISYMSEDPHSPRFFCYLSIFTFFMLMLVTGDNFIQLFLGWEGV
:S::M:::S:::::::::VM:I:::FI::-:::L:::::K::::::RP:M::::::::::::::::::::L:::::::::
:S::M:::S:::::::::VM:I:::FI::-:::L:::::::::::::::M:::::::::P::::::::L:::::::::
:S::M:::S:::::::P::VM:I:::SI::-:::L:::::::::::::::M:::::L::::P:::::::SL::::::::::
:S::M:::S:::V:::P::VM:I:::FI::-:::L:::::::::::::::M:::::L::::P:S:::::SL::::::::::
:D::W:NII:::Q::::::SM:IP:L:I::-:::::::::::::S:::NQ:::S:::L:::M:II:::AN:YLLM:V:::::
ATTQTTQLSLS:KL:YF-SMMFIP:ALFVTWSIMEF:LW::NS::NINQ::K::L::LIT::I:::AN:LP::::I:::::
VLYKN:VLEI:Y:L:P::S:M:VL::T:AV-M:M:::D:::FY:EGYIK::::::L::AS::G::LSP:L::VYIF::L:

GLASYLLINFWFTRIQANKAAIKAMLINRVGDFGLALGIMGCFTIFQ--TVDFSTIFACASAFSEPHHYFLPCNMGFHAI
::::::::H:::::L::D:::::::::V:::::::::::::F::::L::--:::::::::::::::PRN---EWI::::RLN::
::::::::H:::::L::D:::::::::V:::::::::::::::L::::L::--::::::::::::::PRN:::SWIS:::RLN::
::::::::H:::::L::D:::T:::PV:::::::::::::S:R::L::--:::::::::::::R:::PRN:::SWIS:::RLN::
::::::::H:::::L::D:::::::::V:::::::::::::S::::L::--:::::::::::::::PRN:::SWI::::RLN::
:VC::::VS::::::A::QSS:S:F:A:::::CF:TI:MFAILWSLG--NL:Y::V:SL:PYINS----------KVVL:
:IM:F:::SW:YA:AD::T:::Q:I:Y::I::I:FI:A-LAW:-:LHSNSW:PQQM--------ALLNANP--SLT--P-
:MC:::::G:::::PS:AN:CQ::FVT::I:::::L:::L:FYW:TG--SF::QQL-S-KR-:F:LLS:NQI-:LV:-:-

TVICIL-VFIGAVGKSAQIGLHTWLPDAMEGPTPVSALIHAATMVTAGVFMIARCSPLFEYSPNALIVITFVGAMTSFFA
:L::::-L::::::::::::::::::::::::::::::::::::::::C::::::::::::::T::::::::A::::::L:
:L::::-LL:::::::::::::::::::::::::::::::::::::::::::::::::::::T::::::::A::::::L:
SL::::-LL:::::::::::S::::S:::::::::::::S::::::::::::::::::::::::P:T:::::::SA::T:::L:
:LF:::-:::::::::::::S:::S::::::::::::::::::::::::::::::::::::::P:T::::::::A::::::L:
IG::L:---:::MA::S:V:::V:::M::::::::::::::::::::::::YLLM:S:::I::NSTV:LLCLWL::I:TV:S
LL--G:-L-LA:A:::::L:::P:::S::::::::::::L:SS:::V::I:LLI:FH::A:N::LIQTLTLCL::I:TL::
:L-:A:FL:L:P:A::::FP::I:::::::::::::I:::::::::::A::I:LV::MF:::QML:FVMSI:SWT::I:ALLG

ATTGILQNDLKRVIAYSTCSQLGYMIFACGISNYSVSVFHLMNHACFKALLFLSAGSVIHAM-S--D-E----QDMRKMG
:::::::::::::::::::::::::::::::::::::::::::F:::::::::::::::::-:--:-:----::::::
:::::::::::::::::::::::::::::::::::::::::::F:::::::::::::::::-:--:-:----::::::
:::::::::::::::::::::::::::::::::::::::::::F:::::::::::::::::-:--:-:----::::::
:::::::::::::::::::::::::::::::::::::::::::F:::::::::::::::::-:--:-:----::::::
SLI:LF:Q:I:K::::::M::::M:VI:I:L:S:N:AL:::I:::FY::::::G:::::::V-A--:-N----::F::Y:
:VCALT:::IIIIV:F::S:::::L:MVTI::NQPHLAFL:ICT::F:::M::MCS::I::NLNN--E-Q-D-IRK:---:
::IALA:K:::KGL::::M::::::ML:L::GS:KAGL:::IT::YS::::::GS:::::S:EPIVGYHPNKS:N:IF::

GLASLLPFTYAMMLIGSLSLIGFPFLTGFYSKDVILELAYTKYT-ISG-NFAFWLGSVSVFFTSYYSFRLLFLT-FLAPT
::::SF:L:::::::M:::::::::::::::::::::::::::::-:::-:::::::::::::L::::::::::::::-::V::
::::SF::::::::M:::::::::::::::::::::::::::::-:::-:::::::::::::L::::::::::::::-::V::
::::SF:L::::::::::::::::::::::::::::::::::::-:::-:::::::::::::L::::::::::::::-::V::
::::SF::::::::M:::::::::::::::::::::::::::::-:::-:::::::::::::L::::::::::::::-::V::
::INF::L::SV:::A::::VA:::M:::::::F:::S::GQFS-F::-VAVYIIATIGAI::TL::VKV:Y::-::SNP
::LKTM:L:STSLT::::A:A:M:::::::::::H:I:T:NMS::NAWALSITLIAT:LTSAYS-TRMIL:TLTGQPRF::
::RQYM:I:AITF:F:T:::C:I:PFAC:W:::E::VNSWLHFP-:L:-SI::FTAGLTA:YM-FRIYF:T:EGD:RGHF
```

Fig. 2. (Cont.)


```
NSFKR--DLSRCHD-API-LMAIP-LI-LLAFGSIFVGYLAKDMMIGLGTNFWA-NSLFILPKNEILAESEFATPTIIKL
:::G:--:RL::::-:::-P:::::-::-:::L::L::::::::::::::::::::-:::V:::::::::::::::A:::::
:::G:--:IL::::-:::-P:::::-::-:::L::L::::::::::::::::::::-:::V:::::::::::::::A:::::
:::G:--:IL::::-:::-P:::::-::-:::L::L::::::::::::::::::::-:::LV::::::::::::::A:::::
:::G:--:I:::::-:::-P:::::-::-:::L::::::FWPCIDHPMDHG-LNGHC:KGLCPGDRRPPHKVHLAPM
:GTRAYYNYKPALE-GDL-FLTL:-::-I::IF:::F:FIT::IF::::S::FGD:::::H:TH::MIDT:::V:VLF::
LTNINE--NNPTLLNPIKG:A:GSLFAGF:ITNN:SPASPFQTTI-P:-Y-LKL-TA:AVTFLGLL-TA--LDLNYLTNK
FDDVKK--::SISIWGSLEFNKEQ-FK-:DKKSTLYPKEANNI:LFP:-IILTI-PTV::GFIGILFD:NKMNVDSLSYW

IPILFSTLG-SFVAYSVNFVVNPLIFALKTSTFGNRL---YCFPNKRW-FFDKVFNDFLARSFLRFGYEVSFKALD-K-G
::::::::-A:L::N::L:ADQFQR:FQ::::C::---:S::::::-:::Q:L::::V::::::::::S:::E::::-:-:
::::::::-A:L::N::L:ADQFQR:FQ::::C::---:S::::::-:::Q:L::::V::::::::::::E::::-:-:
::::::::-A::::N::P:ADQFQR:FQ::::C::---:S::::::-:::Q:L::::V::::::::::::E::::-:-:
GHYNYPRRH-:KLERET:PPGRLPSYKSYDTGA:FSK---KPRS
L:FI:TISF-:II:LTISEFLSE:VIYF:L:RL:YNI---FG:::Q:F-LIELPY:KYITNLI:DL:GQMT-:I::-:-:
LKMKSPLCTFY:SNM-LG:-YPSIT-HRTIPYL:LLTSQNLPLLLLDLTWLE:LLPKTISQHQISTSIIT:TQKGMI:-L
LTLSINSFNY:NSEKFLE:LF:AIP-SVSIAF::ILIA--FYLYGPNFS:LK:EKKKLQLK:EIDI-VLK::SNFIYNWS

AI-EILGPYGISYTIRKMAQQISKIQSGFVYHYAF-VMLLGLTIFISVIGL-W-DFI-SFWVDNRLYFIYIVSFLFINI
::-:::::::::::F:RL:ER::QL::::::::::-A::::::L:VTFFCM-:-:SL-:S:::S::S::LL::SFL::KS
::-:::::::::::F:RL:ER::QL::::::::::-A::::::L:VTFFCM-:-:SL-:S:I:::SS::W:::SFYNKKS
::-:::::::::::F:RL:ER::QL::::::::::-A::::::L:VTFFCM-:-:SL-:S::::::S::L:::SFYNTKS

S:-:LF::F:LEKGLVNFSKN::SLSTSH:TT::L-YI:VAFILYLLYNY:SF-N:L-P:LIAGLTILTT:
YFLSFFF:LILTLLLITV-NPNK:NSY-PH:VKSIVASTFIISL:PTTMFMCL:QE
YYRAYIDGFYS:FF:KGL-RFLI::V:-:IDRWIIDGIIN:IG::SFFG:ESLKY:EGGRI(SSY:F::IFCM:::FLYS

       669
SQE    670   83.6%(557/666)
SQQ    673   83.9%(561/669)
SQE    673   81.8%(547/669)
       562   76.0%(427/562)
       652   47.1%(305/648)
       640   29.0%(176/606)
YII    692   34.7%(227/654)
```

(g) *nad6*

```
Liverwort    mt                         MILFYVFVVLALVSGAM-VIRAKNPVHSVLFLILVFCNTSGLLVLLGLDFFAMIFLVVYVG
Wheat        mt      MRLLAPAFKFHFKGGRRT:::S-:LSSF:::::L:-:V::::::::::F:::::D:::::I::::::::::::I:
Brassica     mt                         :::S-:LSSP:::::L:-:A:::::::::::P:P::RD:::::L::::::::::::P::HI:
Podospora    mt     MNNNYPLFWINEILTNGFVEYILDI:SIM:FLT:IY-::LT:::IV:::::::L:GGI:SY:NII::N:IGLSYII::I:
Human        mt                         :MYALFLLSVG::M:FVGFSSKPS:IYGG:V::VSGVVGCVIILNF:GGYMGLMVFLI:L:
Liverwort    cp                   MKLPESFYETIFL:LESG:IL:SLG::LLT:I:Y:A:::GF::VCI:L:YL::NA::V:AAQILI:::

AIAVLFLF-VVMMLHIRIEEIHENVLRYLPVGGIIGLIFLLEIFLMVDNDYIPILPTKLSATYLTYTVYAGKIHSWTNLE
:::::::-::::FN:Q:A::::E::::::::S:::::::WW:M:FIL::ET::L:::HRNT:S:R:::::::VR::::::
::::S:::-::::F::Q:A::::E:::::::S:::::::WW:M:FIL::ES::L:::QRNT:S:R:::::::VR::::::
:VSI:::::-IL:LIN::TS:LQS:TSNSI:LTIF::I::SNFL:P:LPY:I--VMLSNFYNN:FSEDF:T--:DVNI:DN
GMM:V:GYTTA:AIEEYP:---AWGSGVEVLVSVLVGLAMEVG:VLWVKE:DGVVVVVNFNSVGSWMI:E:EGSGFIRED
:VN::II:-A::LINKKQY---S:FFV:WTI:DG:T:TLCTS:::LLN:FISNTSWS:IFLMTKPNL:VKDI:LIN:VRH

TLGNLLYTTY-FFL--FLVSSLILLVALIGAIVLTMHKTTKVKRQDVFIQNAIDFQNTIKKVR
::::::::Y:-:VW--:::::::::::M:::::::::R::::::::::RR::L::RSH:MNRTISPFGHSHRRSFSSGAG
::::::::Y:-SVW--:::P::::::M:::::::::R:::::::::::RR:::::RR::MRRTTDPLTIY
N:N::YNNVL-Y:MTSVIWDGSVIDFNH:T::GNI:YTIYNIWLIIASFILLLAMVGS:VITIKQRKI
PI:AGALYD:GRW:--VV:TGWP:F:GVYIV:EIARGN
I-:SE:L:EF-LLP--:ELM:I::::::::::T:A-RREK:IELEKNDFF:F

              199
GPPDNYKETFKMWI   247   75.3%(149/198)
              205   75.3%(145/198)
              221   28.2%(55/195)
              174   15.2%(26/171)
              191   32.8%(60/183)
```

Fig. 2. (Cont.)

(h) *nad9*

```
Liverwort   mt
Wheat       mt   MLCIILFPPERWFSGFGIVTKHPGFYTRFNTRACSRSWIHNSKKCVCSFGSLLVASLSLLPLHSHAFLGWTNPTGDFRQVF
Sugarbeet   mt
Potato      mt
Bovine      nc                                                                                  ES
Neurospora  nc                                   MASKLCRSRALASALRSAKPSPAIRCLATTSRNLINMPERPNPRQFPREPLPG
Liverwort   cp

                 MKNIYISTGFTKKKRRFMDNQLFFKSLIATLPK-WIHKCQTS-KHENILYTNPNSLFQLLYFLKYHTNTRFKVLIDICGV
                 LLRARSG:KIKLSLFS:::::FI:QYSWEI:::K:V::MKR:-E:G:RF:::TDY::P::C:::W::Y::VQ::::::::
                                  ::::FI::YSWE::::K:VK:IEK:-E:G:RFD::TDY:::::C:::L::Y:::Q::::::::
                                  ::::FI::YSWE::::K:VK:MER:-E:G:RFD::TDY:::::C:M:L::Y::VQ::::::::
                 SAADTRP:VRPRNDVAHKQLSA:GEYVAEI:::-YVQQV:V:CFN:LEICIH:DGVIPV:T::RD:S:AQ::S:A:LTA:
                 AL:AAVVNPAD:YQSKADNLHKYGSW:MGC:::-Y:QQFSVW-:D:LTI:IS:AGVIPVFS::::N:AAEYTQVS::TA:
                       MLNIL:NNNNKIQGRLSIW::KHNL:-HRPLGFDY-QGIET:QIRSEDWPS:AVA:YVYGFNYLRSQCAYDVE

                 DYPSRKRRFEVVYNLLSIDYNTRIRILTSVDEITP-ICSVVSIFPSAGWWERETWDMFGVYFSNHPDLRRILTDYGFEGH
                 ::::::::::::::::::TR::S:::VQ::A::V:R-:S::::L::::::::::V::::::S:I::::::::::::::::::
                 ::::::::::::::::::TR::S:::LQ:CA::V:R-:S::::L::::::::::V:::::::S:I:::::::::::::::::
                 ::::::Q:::::::::::R::S:::VQ::A::V:R-:S::::L::::::::::V::::::FSI::::::::::::::::::
                 :I:T:QN:::I::::::LRF:S:::VK:YT::L::-:E:S:PVYKA:N:Y:::I::::::F:A::::::::::::::::::
                 :F:TKDQ::::::::::VRH:S:::VK:YA::VS:-VP:ITPLYDG:N:Y:::VY:L:::F:TG:::::::M:::::D::
                 PGGLLASVYHFTKITDNA:QPEE:C:KIFILRKN:K:P:IFWVWK::DFQ:::SY::::IFYE:::C:K:::MPDSWL:W

                 PLRKDFPLSGYVEVRYDDSEKRVVSEPIEMTQEFRYFD-FASPWEQMSRSDESNQK
                 :::::::::::::::::P::::::::::::::::::::::-::::::::R:DG
                 :::::::::::::::::P::::::::::::::::::::::-::::::::RNGNEG
                 :::::::::::::::::P::::::::::::::::::::::-:::::::R:DGF
                 :F:::::::::::L::::EV::::A::V:LA::::K:::-LN::::AFPAYRQP-PESLKLEAGDTKPEAK
                 :::::::MT::T:I:::EEK::I:T::L::::A::N:EGGS:A:::VGAGIDRKPESFKLPTP--KPETKPEEKK
                 :::::YIVPNFY:LQDAY

                 212
                 287   72.3%(149/206)
                 192   77.0%(147/191)
                 190   77.2%(146/189)
                 228   49.8%(105/211)
                 283   43.9%(93/212)
                 169   22.8%(36/158)
```

Fig. 2. Amino acid sequence comparison of liverwort mitochondrial *nad* gene products with their counterparts from other organisms. Identical amino acid residues with liverwort mitochondrial *nad* 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 identity with its corresponding liverwort mitochondrial *nad* gene products are also shown at ends of sequences.

was also named as *nad9*. The predicted amino acid sequences of the liverwort *nad9* gene have relatively high levels of similarities to those of higher plants (Fig. 2h and Table 3.). The putative liverwort protein displays significant similarity with the 30 kDa subunit of complex I from bovine mitochondria (49.8% homology, Pilkington *et al.*, 1991) and the corresponding 31 kDa protein from *Neurospora crassa* (43.9%, Videira *et al.*, 1990), both of which are encoded by the nuclear genome.

Table 3. Amino acid sequence homology(%) of liverwort mitochondrial NADH dehydrogenase subunit proteins to those of angiosperm mitochondria, *Podospora* mitochondria, human mitochondria, and liverwort chloroplast.

Protein	Angiosperm mt	*Podospora* mt	Human mt	Liverwort cp
ND1	82.7-84.6	47.5	46.4	39.4
2	32.1-83.2	23.0	23.5	34.0
3	76.1-87.2	33.9	37.2	35.9
4	84.6-85.3	41.1	30.6	33.3
4L	85.0	52.8	22.4	36.7
5	76.0-83.9	47.1	29.0	34.7
6	75.3	28.2	15.2	32.8
(7)	(88.2)	-	-	(42.8)
9	72.3-77.2	-	-	22.8

In addition, products of the three *nad* genes, *nad2*, *nad4*, and *nad5* genes, showed significant amino acid sequence similarities each other as shown in Fig. 3. Especially in the middle of the proteins (Phe-227 to Lys-400 in NAD2, Phe-224 to Phe-397 in NAD4, and Lys-236 to Lys-404 in NAD5), 19 amino acid residues were conserved in these three proteins. Functional significance of these conserved 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 5'- 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

```
NAD2  227 FLFKITAVPFHMWAPDVYEGSPTIVTAFF-SIAPKISILANMLRVFIYSF 275
NAD4  224 FSVKVPMVPVHIWLPEAHVEAPTAGSVILAGILLKLGTYG-FLRFSIPMF 272
NAD5  232 AVGKSAQIGLHTWLPDAMEG-PTPVSA-LIHAATMVTAGVFMIARCSPLF 279

NAD2  276 -YDPT--WQQLFFFCSIASMILGALAAMAQNKVKRLLAYSSIGHVGYLLI 322
NAD4  273 -PEATLYFTPFIYTLSVIAIIYTSLTTIRQIDLKKIIAYSSVAHMNFVTI 321
NAD5  280 EYSPNA--LIVITFVGAMTSFFAATTGILQNDLKRVIAYSTCSQLGYMIF 327

NAD2  323 GFSCGTIEGIQSLLIGIFIY-VLMTVNVFAIVLAL-RQNRFKYIADLGAL 370
NAD4  322 GMFSLNIQGIEGSILLMLSH-GLVSSALFLCVGALYDRHKTRIVKYYGGL 370
NAD5  328 ACGISNYSVSVFHLMNHACFKALLFLSAGSVIHAMSDE---QDMRKMGGL 374

NAD2  371 AKTNPILAITLSITMFSYAGIPPLAGFCSK 400
NAD4  371 VSTMPIFSTIFLFFTLANMSLPGTSSFIGE 400
NAD5  375 ASLLPFTYAMMLIGSLSLIGFPFLTGFYSK 404
```

Fig. 3. Amino acid sequence comparison of parts of 2, 4, and 5 subunits of NADH dehydrogenase (NAD2, NAD4, and NAD5, respectively) deduced from liverwort mitochondrial DNA. Black boxes indicate identical amino acid residues . Amino acids common to all the three proteins are depicted by asterisks. Residual numbers are shown both sides of the sequences.

sequence is 330 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 (*trnfM*-CAU), and these three genes are on the same DNA strand, although the possibility of its expression and its function are unknown (Fig. 4B).

A

```
                                                M  I  L  F  Y  V  F  V  V
nad6   ATATATTTTTTGTTCACAGCTAATAAAATTAAAAGGGAAAATTTCACCATGATACTTTTTATGTTTTGTGGTC
orf100 TGACACAGCTAAAAAAAAAAATAATAAAAT--AAATTGAAAATTTCATCATGATACTTTTTATGTTTTATTATT
                                                M  I  L  F  Y  V  F  I  I

       L  A  L  V  S  G  A  M  V  I  R  A  K  N  P  V  H  S  V  L  F  L  I  L  V
nad6   CTCGCTTTAGTGTCAGGCGCTATGGTGATACGTGCCAAAAATCCAGTTCATTCTGTTTTATTTTTAATCCTAGTT
orf100 CTCGCTTTAGTTTCGGGTGCTCTCGGTATACGTGCCAAAAATCCCGTTCATTCTGTTGTCTTTTTCATCCTAGTT
       L  A  L  V  S  G  A  L  G  I  R  A  K  N  P  V  H  S  V  V  F  F  I  L  V

       F  C  N  T  S  G  L  L  V  L  L  G  L  D  F  F  A  M  I  F  L  V  V  Y  V
nad6   TTTTGCAATACTTCCGGGTTACTTGTTTTGTTAGGTCTTGACTTCTTTGCTATGATTTTTTTAGTGGTTTATGTA
orf100 TTTTTCAATACTTTCGGTTTACTTGTTTTGTTAGGTCTTGACTTCTTTGCTATGTTTTTTTTAGTTGTTTATGTA
       F  F  N  T  F  G  L  L  V  L  L  G  L  D  F  F  A  M  F  F  L  V  V  Y  V

       G  A  I  A  V  L  F  L  F  V  V  M  M  L  H  I  R  I  E  E  I  H  E  N  V
nad6   GGAGCTATTGCCGTTTTATTTTTGTTTGTCGTTATGATGTTACATATAAGGATAGAAGAAATTCACGAGAATGTA
orf100 GGA------GCCGTTTTCATTTTGTTTGTCGTTATGATGTTACATATAAGGATAGAAGAAATTCACGAGAATGTA
       G        A  V  F  I  L  F  V  V  M  M  L  H  I  R  I  E  E  I  H  E  N  V

       L  R  Y  L  P  V  G  G  I  I  G  L  I  F  L  L  E  I  F  L  M  V  D  N  D
nad6   TTGCGCTATTTACCTGTAGGTGGTATTATTGGACTTATTTTTTTGTTGGAAATCTTTTTAATGGTAGATAATGAT
orf100 TTGCGCTATTTACCTATAGGTGGTTTATTGGGCTTATTTTTTTTTT----CTATTTTTAATCCCTTCTTTCGTA
       L  R  Y  L  P  I  G  G  F  I  G  L  I  F  F  F     Y  F  *
```

B

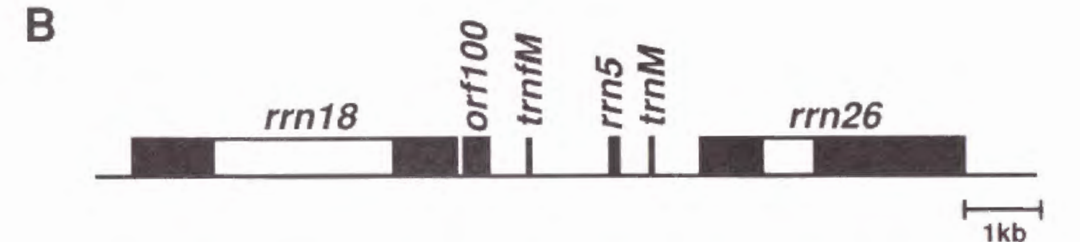

Fig. 4A, B. A The alignment for nucleotide sequences of the *nad6* gene (upper) and its duplicated fragment (lower). The boundaries of the duplicated region are shown with vertical arrows. Dashes indicate the deleted nucleotides and the identical nucleotides are shown by colons between the two sequences. The deduced amino acid sequences of putative products of the *nad6* gene and the duplicated sequence, *orf100*, are also given above and below the nucleotide sequences, respectively. **B** The gene organization in the vicinity of *orf100*. The coding region for each gene is illustrated with the filled box(es) and the introns are shown as open boxes.

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). It

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 threonine 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 (GAGTGACC and GGTCACTC) 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

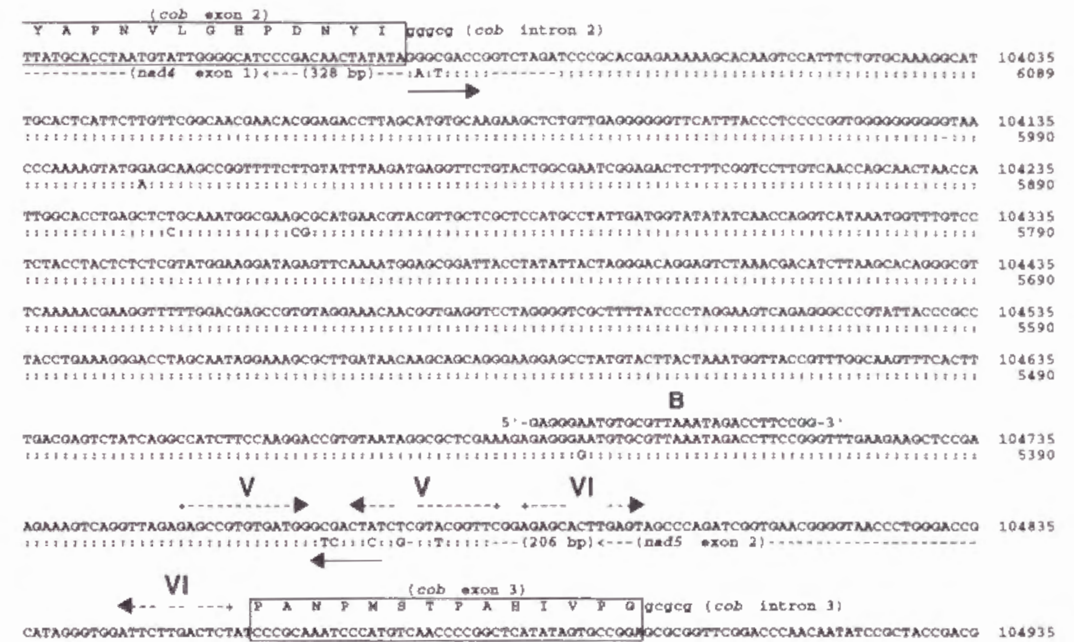

Fig. 5. Nucleotide sequence comparison of repeated regions between the second intron of *cob* and *nad5-nad4* spacer region, of which the respective direction of transcription is shown by solid arrows. Identical nucleotides are shown by colons. Coding regions for *cob* gene (exon 2 and exon 3) are boxed. Inverted repeats are depicted by dashed arrows. Inverted repeats corresponding to the fifth and sixth stem structures typical to group II introns are depicted by V and VI, respectively. Nucleotide sequences which are typical for the 5' ends of group II introns are also shown in small letters (gggcg and gcgcg). Numbering of the nucleotides is according to Oda *et al.* (1992c). B indicates the oligonucleotide probe used for the Northern blot analysis as mentioned in the text.

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.*,

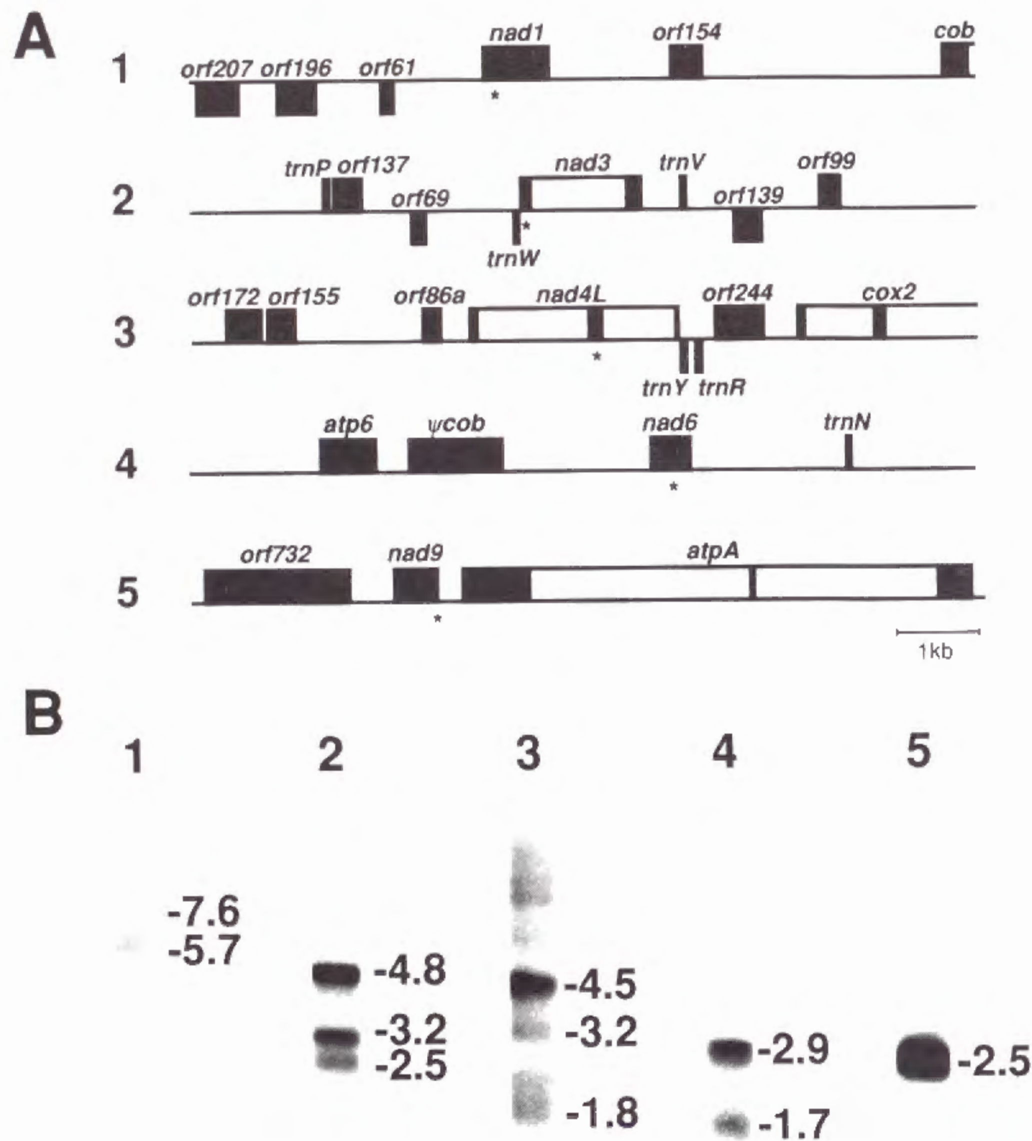

Fig. 6A, B. Transcription analysis of the liverwort *nad1*, *nad3*, *nad4L*, *nad6* and *nad9*. **A** Gene organization of the liverwort *nad* genes. The location of each gene is illustrated with a box, filled for exon and open for intron, respectively. lane 1, *nad1*; lane 2, *nad3*; lane 3, *nad4L*; 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).

1986; Covello and Gray, 1991; Brown *et al*., 1991). In these cases, the consensus sequences which have been reported at the transcription initiation sites (for example, 5'-AAATN$_{1\text{-}6}$TAAG(TA)GA-3', Lonsdale, 1989), are moderately similar to the consensus promoter sequence of yeast mitochondria (5'-ATATAAGTA-3') as defined by mutagenesis (Biswas *et al*., 1987). However, such sequences were not found in the regions upstream of the *nad* genes from liverwort mitochondria. Schuster *et al*., (1986) discovered potential hairpin structures, which were analogous to the bacterial terminators at 3' terminal regions, in the mRNAs from the higher plant mitochondria. Indeed, some potential palindromic structures upstream of and downstream from the liverwort *nad* genes were found, but they were only poorly homologous with comparable structures from higher plants.

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 *Bgl*II-*Pst*I 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 1 (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

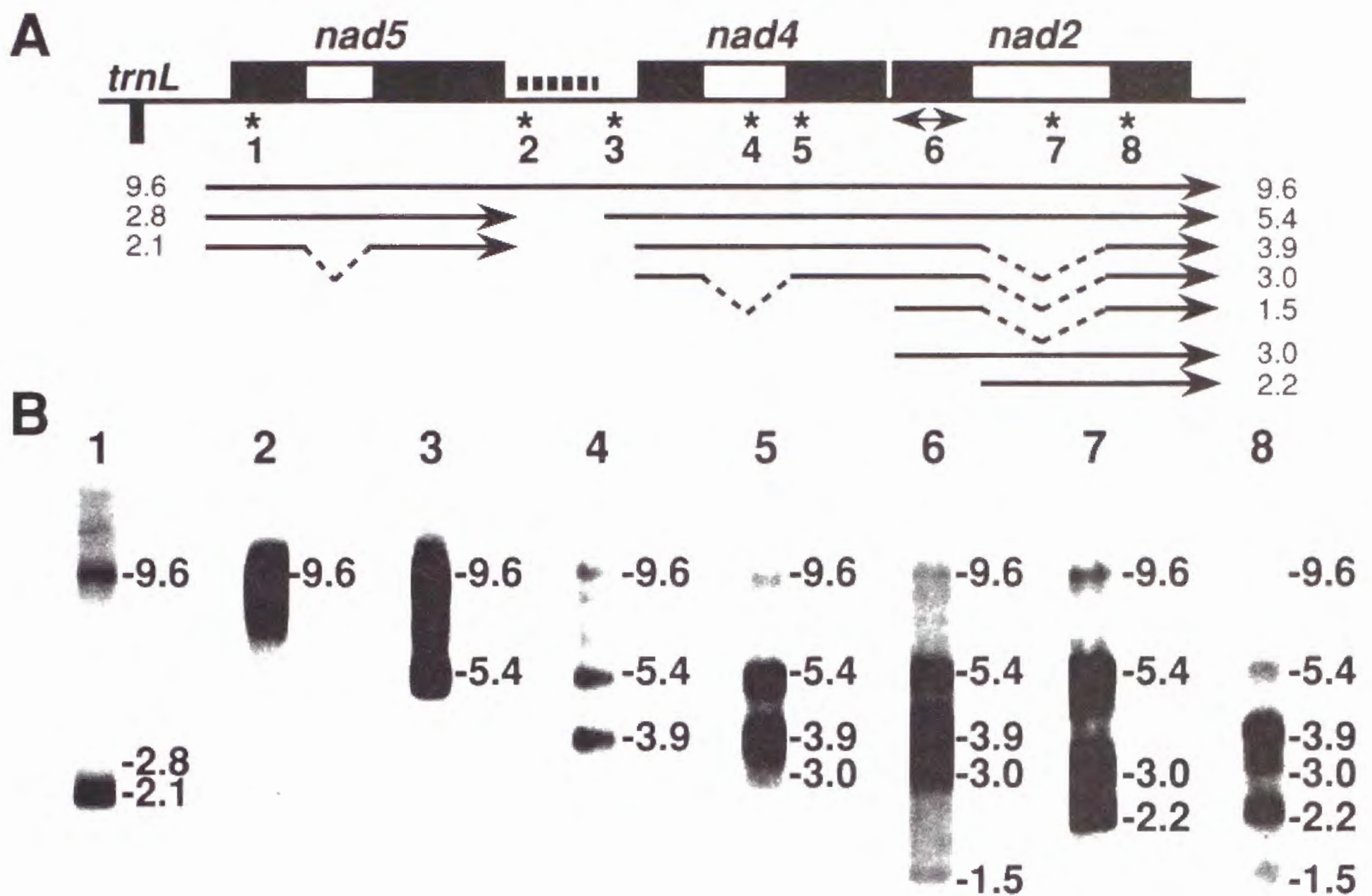

Fig. 7A, B. Transcription analysis of the liverwort *nad2*, *nad4* and *nad5*. **A** Organization of the liverwort*nad2*, *nad4* and *nad5*. Coding regions and introns are indicated by filled and open boxes, respectively. Locations of oligonucleotides and a 664-bp *Bgl*II-*Pst*I DNA fragment used as 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*. **B** Northern hybridization of total mitochondrial RNA was carried out. Molecular sizes are indicated in kilobases (kb).

1 (Fig. 7B, lane 1), a 3.0 kb band for *nad4* was detected by oligonucleotide 5 (Fig. 7B, lane 5), and 3.9, 3.0, or 1.5 kb bands for *nad2* were revealed by probes 6 and 7 (Fig. 7B, lanes 6 and 7). In addition, two bands of 3.0 kb and 2.2 kb corresponding to premature RNA molecules including both exons and an intron of *nad2* was detected (Fig. 7B, lanes 6, 7, and 8). From the Northern blot analysis, the RNA transcripts containing intron sequences are relatively predominant compared with mature RNA encoding subunits of the NADH dehydrogenase in liverwort mitochondria.

Chapter III
ψnad7 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* (*ψnad7*) (Oda *et al.*, 1992a; Oda *et al.*, 1992c). In this chapter, the author described the detail structure of *ψnad7*, 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 pH7.0, 2mM $MgCl_2$, 10mM EDTA, 10mM β-mercaptoethanol, 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 pH7.0, 2mM $MgCl_2$, 10mM β-mercaptoethanol, 10mM EDTA, and 1M sucrose. The nuclear fraction between 60% and 90% Percoll was resuspended in resuspention buffer [1M hexylene glycol, 10mM PIPES-KOH pH7.0, 2mM $MgCl_2$, 10mM EDTA, 10mM β-mercaptoethanol, and 20% glycerol] and centrifuged at 1,000 x g for 10 min. The nuclei pellet was washed with resuspention 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 isothyiocyanate procedure (Chomczynski and Sacchi, 1987). Poly(A)$^+$mRNA was purified by use of oligo(dT)-latex (Oligolatex™-dT30, Daiichi 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 *Xho*I. 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 $NaPO_4$ pH7.2, 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.

Denaturated 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℃. On the other hand, hybridization for poly(A)$^+$mRNA was carried out at 42℃ in a solution [0.5M $NaPO_4$ 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 [γ^{32}-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-^{32}P]dCTP (3,000Ci/mmol, Amersham) using a Random primed DNA labeling kit (Boehringer Mannheim). The oligonucleotide sequences are:

exon 1	: 5'-ACCGCATATTGGATTACTTCATAGAGGCAC-3' (Fig. 4A, 1)
intron 1	: 5'-AAATTCCGGTGTGTCGGACCTGTCATCTGA-3' (Fig. 4A, 2)
exon 2	: 5'-GGATTCAGCGGTGTAATGTTAAGAGGCTCC-3' (Fig. 4A, 3)
intron 2	: 5'-GTCCAACAAGCTCGTGTGAAGATCGAATGACT-3' (Fig. 4A, 4)
exon 3	: 5'-CATAGGTACTCAAGATATTGTGTTTGGAGAGGTAGA-3' (Fig. 4A, 5)
exon 1- exon 2	: 5'-ACTCGGATTTTCAATCATTTACTTGCTTTA-3' (Fig. 4A, 6)
exon 2- exon 3	: 5'-ATGTCCAAACATCACATACTAGCAGATGTT-3' (Fig. 4A, 7)

Results and Discussion

Structure of ψnad7 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-rpl5-rps14-rps8-rpl6-rps13-rps11-rps1*) (Oda *et al.*, 1992a; Oda *et al.*, 1992c). These reading frames were designated as a pseudogene, *ψnad7*, based on the following observations.

This hypothetical *ψnad7* 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 *ψnad7*, 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 maturase-like reading frame in introns

In the hypothetical first intron of *ψnad7*, reading frames which showed significant sequence similarities with the RNA maturase encoded in the first intron of mitochondrial *cox1* 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 cox1* (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

```
Liverwort    mt                                                 M-AKTKQIKNFTFHFGPQHPAAHGVLRLVLEMNGEVV
Wheat        mt                                                 :TTRNG:::::LN::::::::::::::::::::::::
Trypanosoma  mt                                                 :LFLVVFLHLYR:T::::::::::CCL:YFC::FI
Bovine       nc           ARQWQPDVEWAEQYGGAVMYPTKETAHWKPPPWNDVDPPKD:LV-S:L:LN:::::::::::::::M:LS::M:
Neurospora   nc  AEPSYEGQGTRLVPTGDDFAPNNDLYGLEALKADGAPRVPPQDHILAR:V-RHY:VN:::G:::::::::I::LK::EI
Liverwort    cp                                                 :MIL::N-:PMIVSM::H::SM::::::IVTLD::D:

                 TGA
ERAEPHIGLLHRGTEK*IEYKTYLQALPYFDRLDYVSMMAQEHAYSLVVERLCNCEVPLRAQYIRVFFCEITRIFNHLLA
:::::::::::::::L:::::::::::::::::::::::::::::A::::L::::::::::::::L::::::::L::::::
VYIDCI::Y:::::::LC:::SVE:C::::::::::::VCCN::LL::CF:YMLR:CLS::CAFM:LLIV:F::S::G::C
RKCD::::::::::::L:::::::::::::::::::::::::CN:Q::::A::K:L:IRP:P:::W:::L:G::::LL::IM:
V::D::V::::::::::LC::R::::::::::::::::::::IN:QCFA:A::K:L:V:I:E::KW::TM:A:T:::L:::MS
LDC::VL:Y::::M::IA:NR:IV:Y:::VT:W::LAT:FT:AITVNAP:K:T:IQ::K::S:::II:L:LS::AS:::W

                          TGA                                              TGA
LTTHAIDVGALTPFLWAFEEREKL*EFYERVSGARMHASYIRPGGVGQDLPLGLSEDIFLFTQQFASRIDELEEK*TNNR
:::::M:::::::::::::::::L:::::::::::::P::::::A::::::::CR::DS::::::::::::::::ML:G::
ISCMVL:L:C:S:L::S::::D::MT:FDLCC:C:::LAFMVLL:ILD:FVF:FVDFLL:LIISCLFVM:CYDLLFVG::
V::::L:I::M:::F:M::::::MF:::::::::::::A:V:::::H:::::::MD::YE:SKN:SL:::::::ML::::
VLS::M::::::::::G:::::::M::::::::::::L::A:V:::::H::I:::::LD::YMWAT::GD::::T::ML:D::
:GPFMA:I::Q:::FYI:R:::MIYDLF:SAT:M::MHN:F:I:::AV:::Y:WIDKCLD:CDY:LPK:N:Y:RLI:::P

TAG                                    TAA
I*KQRLVDIGTVTAQQAVDWGFSGVMLRGSGVCWNLRK*ALYDVYDRLDFEV--GTRRDCYDRYYIRIEEMRQSIRIIIM
:W::::::::::::::::K::::::::::::::::::D::RA:P:::::Q::::D:PV:::G:::::::C::::::::::V:::-V
LFYL::RGLSPFDLYDL:FNSL:::LS:SL:MV:DC:LFSC:EL:FMFCYDYCFCFIG:AF::LFL:LFD::M:LL:C-K
:WRN:T::::I:::ED:LNY::::::::::::IQ:D:::TQP:::::QVE:D:PI:S:G::::::LC:V:::::::::::-S
:WID::RG::V:S:AD:LNLS:T:::::::::P:DI::SQP::A::QVE:D:PV:ING::::::LC:M::F:::L::::-H
:FLK:VEG:::::REE:IN::L::P:::A:::Q:D:::VDH:EC::E::WKIQWQKEG:SLA::LV::G::KE:VK:::-Q

                                                                  TAA
QCLNQMPSGMIKA-DDRKLGPTARSRMKQSMESLIHHFKLYTESVSVRASSTYTAVEAPKG*FGVFLVSNGTNRPYRCKI
::::::::::::::-:::::C:PS:::::L:::::::::E::::GF::P:::::::::::::::E::::::::::S::::::::
::FFVGFFVFGFVCLFDY:YCDITIETIIMLFYSLWCCC:PGI:FAC---------::H:::EYCLL:CFCVGLCSRLRLR
::::K::P:E:-KV::A:VS:PK:AE::T:::::::::::::::GYQ:PPGA::::I::::::E:::Y:::D:SS::::::::
::::K::A:PV-RVE:Y:IS:PP:AS::EN::A:::::L:::KGYA:PPGD::S:I:::::EM::YV::D:SE::::VH:
:A:KAI:G:PFENLEA:R:NQGKN:EWNLFEYQF:SK-:P-SPTFKLPKQEH:VR::::::EL:I::IGDDSVF:W:L::

TAPGFAHLQGLDFMSKHHILADVVTIIGTQDIVFGEVDR    392
R:::::::::::::::::M:::::::::::::::::::::   394   88.2%(345/391)
C:-D:L:ICL::VCLRGPL:H:L:AVL:NI:V:::S:::    386   33.5%(128/382)
K::::::::A:::K:::G:M:::::A::::::::::::::   430   70.7%(275/389)
R:::::::G:F:HL:RG:M:::A:AV::IM:L::::::::   436   61.2%(238/389)
RS:::IN::I:PQLV:GMK:::IM::L:SI::IM:::::    392   42.8%(166/388)
```

Fig. 1. Amino acid sequence comparison of ND7 related gene products. Hypothetical product of *ψnad7* 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 *ψnad7* 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 *ψnad7* 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 *ψnad7* gene product are also shown at the ends of sequences.

Fig. 2. Secondary structures of the two group II-like introns in *ψnad7*. (**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).

```
Liverwort   nad7i1
Yeast       cox1i1   GAPTSAYISLMR-----TALVLWIINRYLKHMTNSVGANFTGTMACHKTPMI-SVGGVKCYMVRLTNNLQVFIRIT
Podospora   cox1i1   GAPFNVRFKSGELDPISYCCCLLNLLTYL--MTKGLRECSMSVNPYLTIA-IKSVESGEVKASYVLRLLTMVGLCV

                             *DLILTIIPWEPFRHSRQDRRPRIEVSLQAYLCTKWDDPTQSGPTDCTRLVPKDLKHQTPGPSEVNSI
                     ISSYHLDMVKQVWLFYVEVIRLWFIVLDSTGSVKKMKDTNN::GNTKSEGSTERGNSG:D-RGMVVPNTQMKMRFL
                     SIGIKIAIALYWVL:KISASLIKNSYSFTTSERGYIVY------NAKGRRRLNVGNSGL::GRNSYGN:SVV:GVT

                     ALISAYKLLKEKDYMNYLYLTKLLLTFQIELIYQLQNHENLLRLIAQKGLYCKLFD-TNLHITTYGKLKSKQGNMT
                     NQVRY:SVNNNLKMGKDTNIELSKD:STSD:LEFEKLVMDNMNEENMNNNLLSIMKNVDMLMLA:NRI:::P::::
                     SGKWIHT-----KVRKSKVSP:SEARKGQGSLGEMLMYNERGQC:NA---:EVICK-LEALY:A:MNI::EP::::

                                   ▼
                     PGLRTPKTWIAWDN/KNHCQTQGQIIPIQNT-SRRTFIPKPNGDQRPLGIPSLRDKVVQEAIQAVIEPAFERKFLP
                     ::TTLETLDGMNMMYL:KLSNELGTGKFKFKPM:MVN:::K:GM:::SVGNP::KI:::VMRMILDTI:DK:MST
                     :RVDSETLDGISKEWFEKISE:LKSEQFRFRPT::VY:::A::KM:::::A:P::KI:::VFR:IL:QVL:PR:HS

                                             ▼                                   ▽
                     SSHGFRPSRSPRTALREIRTNGRV/NWAIEGDIKGYFYNINHHKLASFLDAELR----DP*LLQLYWKLV--GGQD
                     H:::::KNM:CQ::IW:V:NMPGGS::F::V:L:KC:DT:S:DLI----IK::KRYIS:KGFID:VY::LRA:YI:
                     :::::::G:GCHS::AT::YWNGIK-:F:::::::F:D::D::I:----EKL:VKHFQ:QRFID::::M:KA:YVE

                              ▼                                        ▼
                     MD*KRSPKL/VVGLRGVLSPMLPNIYLGPLDQFCEEL-KIR--YKAPTSL/SPDRRFRLFKKSKE------EDHL-
                     EKGTYHKPMLGLPQGSLI::I:C::VMTLV:NWL:---DYINL:NK-GKVKKQHPTY-KKLSRMI------AKAKM
                     F:KDK:S-IIG:PQG:IA::I:S:LV:NE::E:VQNIVDEFNEKLKGGKHT:KNPAYVVIDSRIGKITRLERKLKS

                                                                  ▼                          ▽
                     ---------RARLEWL-----RKSIERRVV-KIY--YLRYANDWI/GVVGSKKVALAIRAEIASFL--KLHL*LNW
                     FSTRL-KLHK-ERAKG----PTFIYNDPNFKRMK--:V:::D:ILI::L:::NDCKM:K-RDLNNFLNS:G:TM:E
                     KGQELDSGRKLERMK:IKVRATMPSMIPNPDLAKIY:V:::D::LI::A::SET:R::KER::AY:KDI:K:E:SM

                     DNTKITHISSQLALFLGTHIKVLRAESI--RNHRILVVGQRTRSATFRLHLLAPIERIVKHLHGKGLC--------
                     EK:L::CATELP:R:::YN:SITPLKRMPTVTKT:-RGKTIRSRN:T:PIIN:::RD:INK:ATN:Y:-KHNK--N
                     EK:L::NA:EDR:Y::::E:QRISSVKGEIKRFKN-IK:HPQ:IP:TSTVMN:::SKL:TK:AD::IVIWKSKALN

                               ▽
                     TPTGKPKPVR*WIFLDHHELIFRYQDIMSGYMNYYSFVDNYGMLK-RVAYIVRFSAAGTLKRKFKMLSVASVFKGS
                     GRM:V:TR:GR:LYEEPRTI:NN:KALGR:IL:::KLAT::KR:RE:IY:VLYY:CVL::AS:YRLKTMSKTI:KF
                     EDNLI:Q:ILK:VN:PIRDI:L::KM:WN::I:::::A::KPR:V-LIYW:L:K:L:K::AT:L:LGT:RK:YLKF

                     GTGRGKEL*
                     :YNLNIIENDKLIANFPRNTFDNIKKIENHGMFMYMSEAKVTDPFEYIDSIKYMLPTAKANFNKPCSICNSTIDVE
                     :VNLRF:ILGTDNKSIEFTKGNLLPTPKNFKGKTNFVDNLKVVEWSLRTVS---------FFNYVCASCGASDNLQ

                     MHHVKQLHRGMLKATKDYITGRMITMNRKQIPLCKQCHIKTHKNKFKNMGPGM*
                     VHHVKHIRTIDVKLSGFDKQLAAI--NRKQVTLCISCHNKVHTGKYDGMSLKYMKDISKPELNQQ*
```

Fig. 3. Amino acid sequence comparison of the hypothetical protein (Liverwort nad7i1) in the liverwort *ψnad7* first intron with the RNA maturase encoded by the first intron in the yeast *coxI* (Yeast cox1i1) and with the reverse transcriptase-like ORF encoded by the first intron in *Podospora coxI* (*Podospora* cox1i1). Frame shifts are indicated by diagonals and filled triangles. Open triangles show translational stop codons. Other symbols are same as described in Fig. 1.

intron (data not shown). These findings suggest that both DNA segments are traces of RNA maturases in the liverwort mitochondrial genome. A maturase-related open reading frame was identified in a group II intron of broad bean mitochondrial *nad1* (Wahleithner *et al.*, 1990). However, this did not show significant amino acid sequence similarity with those in the liverwort *ψnad7* introns.

Expression of ψnad7 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}P-labeled oligonucleotide probes (Fig. 4B). Probes for the putative three exons and two introns hybridized with two 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 *ψnad7* itself (5,668 bp) but also regions upstream and/or downstream. Northern hybridization probed for *rps10* which is located 1.3 kb downstream from *ψnad7* demonstrates the existence of the same 9.6 kb band as shown in Chapter I, supporting that *ψnad7* is co-transcribed with at least *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 band 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 *ψnad7* 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 *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

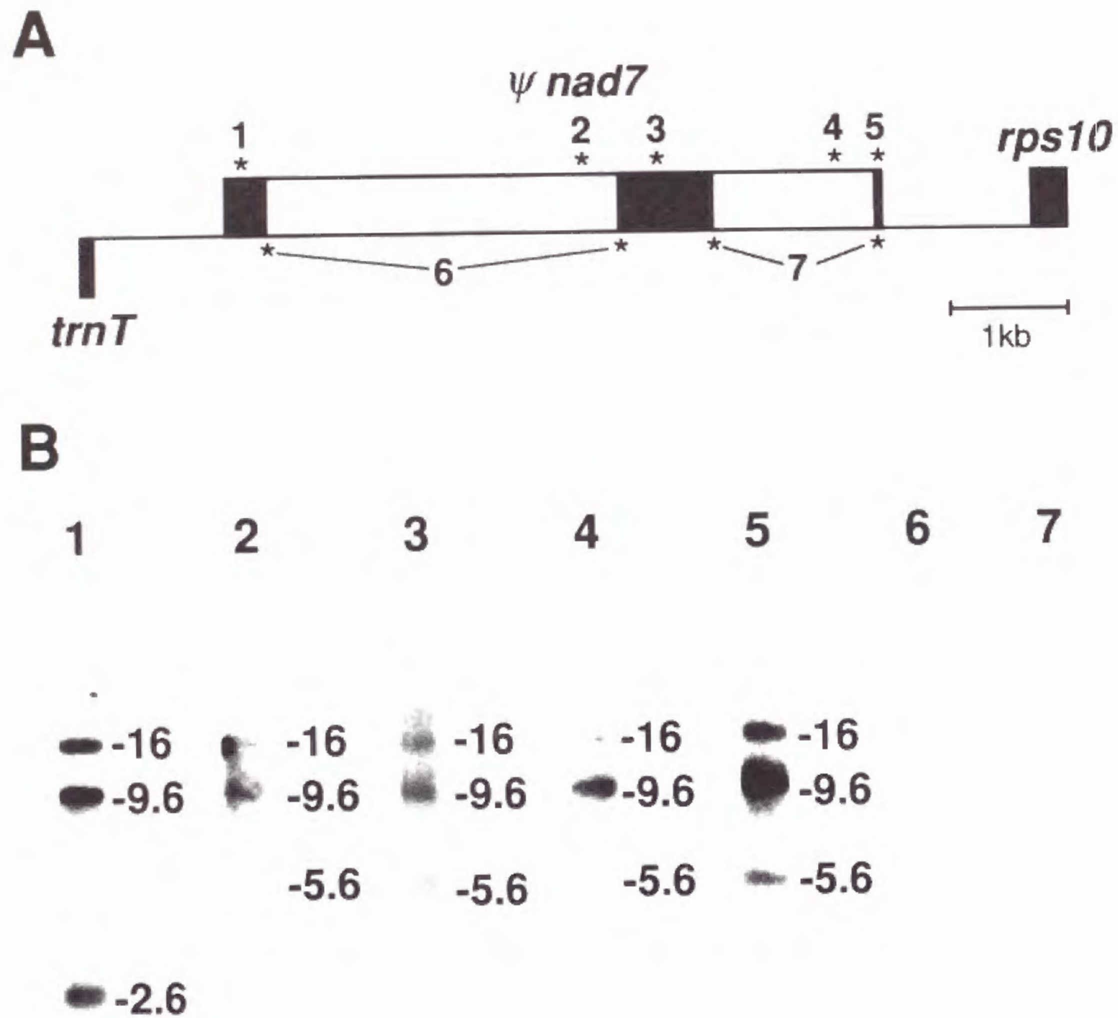

Fig. 4A, B. Northern blot analysis of mitochondrial RNA. **A** Gene organization of the liverwort *ψnad7* gene. Filled and open boxes indicate hypothetical exons and introns, respectively. Asterisks with numbers show positions of oligonucleotide probes used for RNA blot analysis. **B** Total mitochondrial 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).

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 *Xho*I 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, lane1). When the liverwort mitochondrial DNA digested with *Xho*I was probed with the same probe, a single 26 kb band was detected which corresponded to mitochondrial *ψnad7* (Fig. 5B, lane 2). In contrast, the exon 2 specific probe hybridized with a 42 kb *Xho*I chloroplast DNA fragment which corresponded to the *nad7* homologue ORF392 (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 *ψnad7* (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.

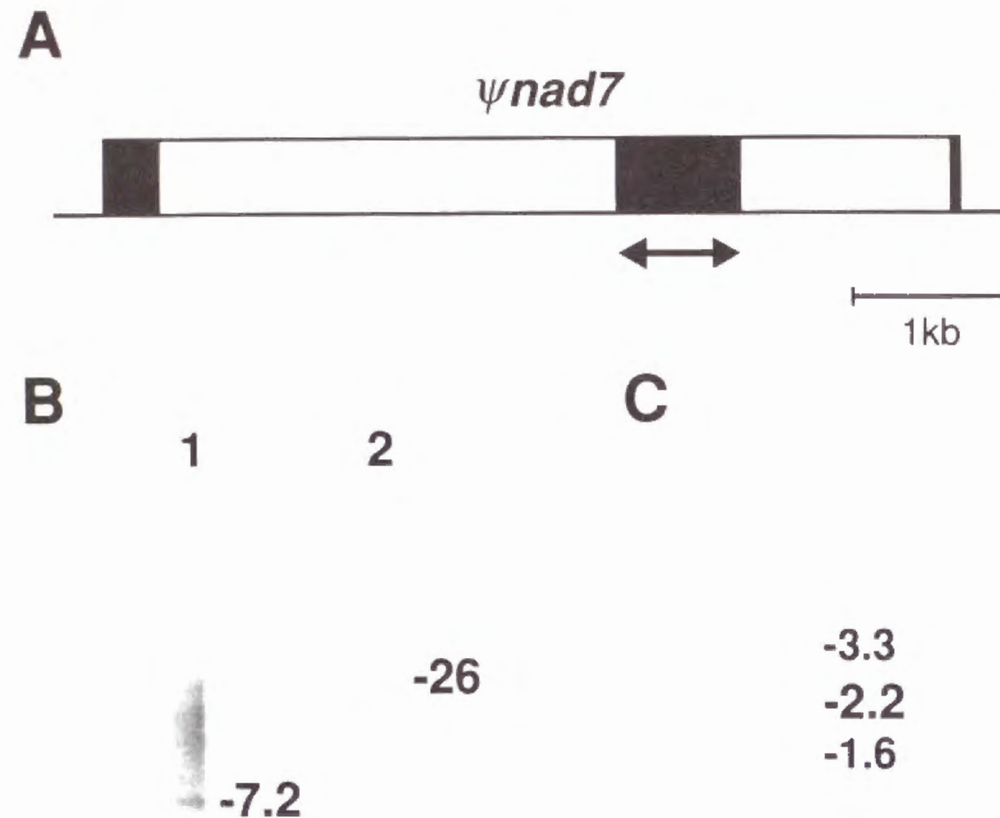

Figure 5A, B and C. A Gene organization of the liverwort mitochondrial *ψnad7* gene. Bar with arrows shows an exon 2 specific DNA probe generated by PCR amplification using cloned mitochondrial DNA as a template. **B** Southern hybridization analysis of nuclear DNA (lane 1) and mitochondrial DNA (lane 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

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 (Gälber *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 *ψnad7* 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.

References

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) *Nature* **290**, 457-465.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley- Interscience, Secaucus, NJ, pp4.3.2

Bedwell, D., Davis, G., Gosink, M., Post, L., Nomura, M., Kestler, H., Zengel, J.M. and Lindahl, L. (1985) *Nucl. Acids. Res.* **13**, 3891-3903.

Binder, S., Marchfelder, A., Brennicke, A. and Wissinger, B. (1992) *J. Biol. Chem.* **267**, 7615-7623.

Biswas, T.K., Ticho, B. and Getz, G.S. (1987) *J. Biol. Chem.* **262**, 13690-13696.

Bland, M.M., Levings III, C.S. and Matzinger, D.F. (1986) *Mol. Gen. Genet.* **204**, 8-16.

Bock, H., Brennicke, A. and Schuster, W. (1994) *Plant Mol. Biol.* **24**, 811-818.

Bonen, L. (1987) *Nucl. Acids. Res.* **15**, 10393-10404.

Bonen, L., Williams, K., Bird, S. and Wood, C. (1994) *Mol. Gen. Genet.* **244**, 81-89.

Bonitz, S.G., Coruzzi, G., Thalenfeld, B.E. and Tzagoloff, A. (1980) *J. Biol. Chem.* **255**, 11927-11941.

Brandt, P., Sunkel, S., Unseld, M., Brennicke, A. and Knoop, V. (1992) *Mol. Gen. Genet.* **236**, 33-38.

Brandt, P., Unseld, M., Eckert-Ossenkopp, U. and Brennicke, A. (1993) *Curr. Genet.* **24**, 330-336.

Brears, T. and Lonsdale, D.M. (1988) *Mol. Gen. Genet.* **214**, 514-522.

Brown, G.G., Auchincloss, A.H., Covello, P.S., Gray, M.W., Menassa, R. and Singh, M. (1991) *Mol. Gen. Genet.* **228**, 345-355.

Butow, R.A., Perlman, P.S. and Grossman, L.I. (1985) *Science* **228**, 1496-1501.

Carignani, G., Groudinsky, O., Frezza, D., Schiavon, E., Bergantino, E. and Slonimsky, P.P. (1983) *Cell* **35**, 733-742.

Cerretti, D.P., Dean, D., Davis, G.R., Bedwell, D.M. and Nomura, M. (1983) *Nucl. Acids. Res.* **11**, 2599-2616.

Chapdelaine, Y. and Bonen, L. (1991) *Cell* **65**, 465-472.

Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156-159.

Chomyn, A., Mariottni, P., Cleeter, M.W.J., Ragan, C.I., Matsuno-Yagi, A., Hatefi,

Y., Doolittle, R.F. and Attardi, G. (1985) *Nature* **314**, 592-597
Chomyn, A., Cleeter, M.W.J., Ragan, C.I., Riley, M., Doolittle, R.F. and Attardi, G. (1986) *Science* **234**, 614-618.
Christianson, T. and Rabinowitz, M, (1983) *J. Biol. Chem.* **258**, 14025-14033.
Clayton, D.A. (1984) *Ann. Rev. Biochem.* **53**, 573-594.
Clayton, D.A. (1991) *Annu. Rev. Cell Biol.* **7**, 453-478.
Conklin, P.L. and and Hanson, M.R. (1991) *Nucl. Acids. Res.* **19**, 2701-2705.
Conklin, P.L., Wilson, R.K. and Hanson, M.R. (1991) *Genes. Dev.* **5**, 1407-1415.
Covello, P.S. and Gray, M.W. (1991) *Curr. Genet.* **20**, 245-251.
Covello, P.S. and Gray, M.W. (1992) *EMBO J.* **11**, 3815-3820.
Cummings, D.J., McNally, K.L., Domenico, J.M. and Matsuura, E.T. (1990) *Curr Genet* **17**, 375-402.
de Souza, A.P., Jubier, M.F., Delcher, E., Lancelin, D. and Lejeune, B. (1991) *Plant Cell* **3**, 1363-1378.
de Zamaroczy, M. and Bernardi, G. (1986) *Gene* **47**, 155-177.
Dupuis, A., Skehel, J.M. and Walker, J.E. (1991) *Biochem.* **30**, 2954-2960.
Ecke, W., Schmitz, U. and Michaelis, G. (1990) *Curr. Genet.* **18**, 133-139.
Fearnley, I.M., Runswick, M.J. and Walker, J.E. (1989) *EMBO J.* **8**, 665-672.
Gäbler, L., Herz, U., Liddell, A., Leaver, C.J., Schröder, W., Brennicke, A. and Grohmann, L. (1994) *Mol. Gen. Genet.* **244**, 33-40.
Gass, D.A., Makaroff, C.A. and Palmer, J.D. (1992) *Curr. Genet.* **21**, 423-430.
Geiss, K.T., Abbas, G.M. and Makaroff, C.A. (1994) *Mol. Gen. Genet.* **243**, 97-105.
Gonzalez, D.H., Bonnard, G. and Grienenberger, J.-M. (1993) *Curr. Genet.* **24**, 248-255.
Gray, M. W., Cedergren, R., Abel, Y. and Sankoff, D. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2267-2271.
Gray, M.W., Hanic-Joyce, P.J. and Covello, P.S. (1992) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **43**, 145-175.
Grohmann, L. Brennicke, A. and Schuster, W. (1992) *Nucl. Acids. Res.* **20**, 5641-5646.
Grohmann, L., Thieck, O., Herz, U., Schröder, W. and Brennicke, A. (1994) *Nucl. Acids. Res.* **22**, 3304-3311.
Gualberto, J.M., Wintz, H., Weil, J.H. and Grienenberger, J.M. (1988) *Mol. Gen. Genet.* **215**, 118-127.
Gualberto, J.M., Lamattina, L., Bonnard, G., Weil, J.H. and Grienenberger, J.M. (1989) *Nature* **341**, 660-662.

Haouazine, N., de Souza, A.P., Jubier, M.-F., Lancelin, D., Delcher, E. and Lejeune, B. (1992) *Plant Mol. Biol.* **20**, 395-404.
Haouazine, N., Takvorian, A., Jubier, M.-F., Michel, F. and Lejeune, B. (1993) *Curr. Genet.* **24**, 533-538.
Herz, U., Schröder, W., Liddell, A., Leaver, C.J., Brennicke, A. and Grohmann, L. (1994) *J. Biol. Chem.* **269**, 2263-2269.
Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.R., Meng, B.Y., Li, Y.Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. and Sugiura, M. (1989) *Mol. Gen. Genet.* **217**, 185-194.
Hunt, M.D. and Newton, K.J. (1991) *EMBO J.* **10**, 1045-1052.
Kang, W., Matsushita, Y. and Isono, K. (1991) *Mol. Gen. Genet.* **225**, 474-482.
Kim, K., Schuster, W., Brennicke, A. and Choi, K. (1991) *Plant Physiol.* **97**, 1602-1603.
Kitakawa, M., Grohmann, L., Graack, H.R. and Isono, K. (1990) *Nucl. Acids. Res.* **18**, 1521-1529.
Knoop, V., Schuster, W., Wissinger, B. and Brennicke, A. (1991) *EMBO J.* **10**, 3483-3493.
Kohchi, T., Shirai, H., Fukuzawa, H., Sano, T., Komano, T., Umesono, K., Inokuchi, H., Ozeki, H. and Ohyama, K. (1988) *J. Mol. Biol.* **203**, 353-372.
Koslowsky, D.J., Bhat, G.J., Perrollaz, A.L., Feagin, J.E. and Stuart, K. (1990) *Cell* **62**, 901-911.
Kubo, T., Mikami, T. and Kinoshita, T. (1993) *Mol. Gen. Genet.* **241**, 479-482.
Labacq, P. and Vedel, F. (1981) *Plant Science Letters* **23**, 1-9.
Lamattina, L., Weil, L.H. and Grienenberger, J.M. (1989) *FEBS Letter* **258**, 79-83.
Lamattina, L. and Grienenberger, J.M. (1991) *Nucl. Acids. Res.* **19**, 3275-3282.
Lamattina, L., Gonzalez, D., Gualberto, J. and Grienenberger, J.-M. (1993) *Eur. J. Biochem.* **217**, 831-838.
Lonsdale, D.M. (1989) In Stumpf, P.K. and Conn, E.E. (eds.), *The biochemistry of plants*, Academic Press, New York, **15**, pp 229-295.
Makaroff, C.A. and Palmer, J.D. (1987) *Nucl. Acids. Res.* **15**, 5141-5156.
Mi, H., Endo, T., Schreiber, U., Ogawa, T. and Asada, K. (1992) *Plant Cell Physiol.* **33**, 1233-1237.
Michel, F. and Dujon, B. (1983) *EMBO J.* **2**, 33-38.
Mulligan, R.M., Lau, G.T. and Walbot, V. (1988a) *Proc. Natl. Acad. Sci. USA* **85**, 7998-8002.

Mulligan, R.M., Maloney, A.P. and Walbot, V. (1988b) *Mol. Gen. Genet.* **211**, 373-380.

Myers, A.M., Crivellone, M.D. and Tzagoloff, A. (1987) *J. Biol. Chem.* **262**, 3388-3397.

Nixon, P.J., Gounaris, K., Coomber, S.A., Hunter, C.N., Dyer, T.A. and Barber, J. (1989) *Biol. Chem.* **264**, 14129-14135

Nozato, N., Oda, K., Yamato, K., Ohta, E., Takemura, M., Akashi, K., Fukuzawa, H. and Ohyama, K. (1993) *Mol. Gen. Genet.* **237**, 343-350.

Nugent, J.M. and Palmer J.D. (1991) *Cell* **66**, 473-481.

Nugent, J.M. and Palmer J.D. (1993) *Curr. Genet.* **23**, 148-153.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T. and Ohyama, K. (1992a) *J. Mol. Biol.* **223**, 1-7.

Oda, K., Kohchi, T. and Ohyama, K. (1992b) *Biosci. Biotech. Biochem.* **56**, 132-135.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T. and Ohyama, K. (1992c) *Plant Mol. Biol. Rept.* **10**, 105-163.

Ogawa, T. (1991a) *Proc. Natl. Acad. Sci. USA* **88**, 4275-4279.

Ogawa, T. (1991b) *Plant Physiol.* **96**, 280-284.

Ogawa, T. (1992) *Plant Physiol.* **99**, 1604-1608.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1986) *Nature* **322**, 572-574.

Ohyama, K., Fukuzawa, H., Kohchi, T., Sano, T., Sano, S., Shirai, H., Umesono, K., Shiki, K., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1988) *J. Mol. Biol.* **203**, 281-298.

Ohyama, K., Ogura, Y., Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T. and Ymada, Y. (1991) In *Evolution of life* eds. by Osawa, S. and Honjo, T., Springer-Verlag, Tokyo, pp.187-198.

Palmer, J.D. and Shields, C.R. (1984) *Nature* **307**, 437-440.

Palmer, J.D. and Herbon, L.A. (1988) *J. Mol. Evol.* **28**, 87-97.

Pilkington, S.J., Skehel, J.M. and Walker, J.E. (1991) *Biochem.* **30**, 1901-1908.

Post, L.E. and Nomura, M. (1980) *J. Biol. Chem.* **255**, 4660-4666.

Preis, D., van der Pas, J.C., Nehls, U., Roehlen, D.-A., Sackman, U., Jahnke, U. and Weis, H. (1990) *Curr. Genet.* **18**, 59-64.

Rasmussen, J. and Hanson, M.R. (1989) *Mol. Gen. Genet.* **215**, 332-336.

Rothenberg, M. and Hanson, M.R. (1987) *Mol. Gen. Genet.* **209**, 21-27.

Schinkel, A.H. and Tabak, H.F. (1989) *Trends Genet.* **5**, 149-154.

Schuster, W. (1993) *Mol. Gen. Genet.* **240**, 445-449.

Schuster, W., Hiesel, R., Isaac, P.G., Leaver, C.J. and Brennicke, A. (1986) *Nucl. Acids. Res.* **14**, 5943-5954.

Schuster, W. and Brennicke, A. (1987a) *Mol. Gen. Genet.* **210**, 44-51.

Schuster, W. and Brennicke, A. (1987b) *EMBO J.* **6**, 2857-2863.

Schuster, W. and Brennicke, A. (1991) *Nucl. Acids. Res.* **19**, 6923-6928.

Schuster, W., Unseld, M., Wissinger, B. and Brennicke, A. (1990a) *Nucl. Acids. Res.* **18**, 229-233.

Schuster, W., Wissinger, B., Unseld, M. and Brennicke, A. (1990b) *EMBO J.* **9**, 263-269.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) *EMBO J.* **5**, 2043-2049.

Siculella, L. and Palmer, J.D. (1988) *Nucl. Acids. Res.* **16**, 3787-3799.

Stern, D.B. and Lonsdale, D.M. (1982) *Nature* **299**, 698-702.

Stern, D.B. and Palmer, J.D. (1986) *Nucl. Acids. Res.* **14**, 5651-5666.

Stern, D.B., Bang, A.G. and Thompson, W.F. (1986) *Curr. Genet.* **10**, 857-869.

Sutton, C.A., Conklin, P.L., Pruitt, K.D., Calfee, A.J., Cobb, A.G. and Hanson, M.R. (1993) *Curr. Genet.* **23**, 472-476.

Suzuki, T., Kazama, S., Hirai, A., Akihama, T. and Kadowaki, K. (1991) *Curr. Genet.* **20**, 331-337.

Umesono, K., Inokuchi, H., Shiki, Y., Takeuchi, M., Chang, Z., Fukuzawa, H., Kohchi, T., Shirai, H., Ohyama, K. and Ozeki, H. (1988) *J. Mol. Biol.* **203**, 299-331.

Videira, A., Tropschug, M. and Werner, S. (1990) *Biochem. Biophys. Res. Comm.* **171**, 1168-1174.

Wahleithner, J.A. and Wolstenholme, D.R. (1988) *Nucl. Acids. Res.* **16**, 6897-6913.

Wahleithner, J.A., Macfarlane, J.L. and Wolstenholme, D.R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 548-552.

Walker, J.E. (1992) *Quart. Rev. Biophys.* **25**, 253-324.

Ward, B.L., Anderson, R.S. and Bendich, A.J. (1981) *Cell* **25**, 793-803.
Weiss, H., Friedrich, T., Hofhaus, G. and Price, D. (1991) *Eur. J. Biochem.* **197**, 563-576.
Wintz, H., Chen, H.C. and Pillay, D.T.N. (1989) *Curr. Genet.* **15**, 155-160.
Wissinger, B., Hiesel, R., Schuster, W. and Brennicke, A. (1988) *Mol. Gen. Genet.* **212**, 56-65.
Wissinger, B., Schuster, W. and Brennicke, A. (1990) *Mol. Gen. Genet.* **224**, 389-395.
Wissinger, B., Schuster, W. and Brennicke, A. (1991) *Cell* **65**, 473-482.
Xue, Y., Davies, D.R. and Thomas, C.M. (1990) *Mol. Gen. Genet.* **221**, 195-198.
Yamato, K., Ogura, Y., Kanegae, T., Yamada, Y. and Ohyama, K. (1992) *Theor. Appl. Genet.* **83**, 279-288.
Yamato, K., Nozato, N., Oda, K., Ohta, E., Takemura, M., Akashi, K. and Ohyama, K. (1993) *Curr. Genet.* **23**, 526-531.
Ye, F., Bernhardt, J. and Abel, W.O. (1993) *Curr. Genet.* **24**, 323-329.
Young, E.G., Hanson, M.R. and Dierks, P.M. (1986) *Nucl. Acids. Res.* **14**, 7995-8006.
Zanlungo, S., Quiñones, V., Moenne, A., Holuigue, L. and Jordana, X. (1994) *Plant Mol. Biol.* **25**, 743-749.
Zhuo, D. and Bonen, L. (1993) *Mol. Gen. Genet.* **236**, 395-401.
Zurawski, G. and Zurawski, S.M. (1985) *Nucl. Acids. Res.* **13**, 4521-4526.

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-rpl5-rps14-rps8-rpl6-rps13-rps11-rps1*, very similar in organization to *E. coli* ribosomal protein operons (*str* and S10-*spc*-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 *ψnad7*. 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 wheareas *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, *ψnad7*, 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 *ψnad7*, is a relic of a gene transfer event from mitochondrial genome into nuclear genome during mitochondrial evolution in *M. polymorpha*.

List of Publications

(a) K. Ohyama, Y. Ogura, K. Oda, K. Yamato, E. Ohta, Y. Nakamura, M. Takemura, N. Nozato, K. Akashi, T. Kanegae, and Y. Yamada (1991) Evolution of organellar genomes in *Evolution of life* (eds. by S. Osawa and T. Honjo, Springer-Verlag, Tokyo) pp.187-198.

(b) Y. Ogura, T. Yoshida, Y. Nakamura, M. Takemura, K. Oda, and K. Ohyama (1991) Gene encoding a putative zinc finger protein in *Synechocystis* PCC6803. *Agric. Biol. Chem.* **55**, 2259-2264.

(c) K. Oda, K. Yamato, E. Ohta, Y. Nakamura, M. Takemura, N. Nozato, K. Akashi, T. Kanegae, Y. Ogura, T. Kohchi, and K. Ohyama (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. *J. Mol. Biol.* **223**, 1-7.

(d) K. Oda, K. Yamato, E. Ohta, Y. Nakamura, M. Takemura, N. Nozato, K. Akashi, and K. Ohyama (1992) Complete nucleotide sequence of the mitochondrial DNA from a liverwort, *Marchantia polymorpha*. *Plant Mol. Biol. Rep.* **10**, 105-163.

(e) Y. Ogura, M. Takemura, K. Oda, K. Yamato, E. Ohta, H. Fukuzawa, and K. Ohyama (1992) Cloning and nucleotide sequence of a *frxC*-ORF469 gene cluster of *Synechocystis* PCC6803: conservation with liverwort chloroplast *frxC*-ORF465 and *nif* operon. *Biosci. Biotech. Biochem.* **56**, 788-793.

(f) M. Takemura, K. Oda, K. Yamato, E. Ohta, Y. Nakamura, N. Nozato, K. Akashi, and K. Ohyama (1992) Gene clusters for ribosomal proteins in the mitochondrial genome of a liverwort, *Marchantia polymorpha*. *Nucl. Acids. Res.* **20**, 3199-3205.

(g) K. Oda, K. Yamato, E. Ohta, Y. Nakamura, M. Takemura, N. Nozato, K. Akashi, and K. Ohyama (1992) Transfer RNA genes in the mitochondrial genome from a liverwort, *Marchantia polymorpha*: the absence of chloroplast-like tRNAs. *Nucl. Acids. Res.* **20**, 3773-3777.

(h) N. Nozato, K. Oda, K. Yamato, E. Ohta, M. Takemura, K. Akashi, H. Fukuzawa, and K. Ohyama (1993) Cotranscriptional expression of mitochondrial genes for subunits of NADH dehydrogenase, *nad5*, *nad4*, *nad2*, in *Marchantia polymorpha*. *Mol. Gen. Genet.* **237**, 343-350.

(i) K. Yamato, N. Nozato, K. Oda, E. Ohta, M. Takemura, K. Akashi, and K. Ohyama (1993) Occurrence and transcription of genes for *nad1*, *nad3*, *nad4L*, and *nad6*, coding for NADH dehydrogenase subunits 1, 3, 4L, and 6, in liverwort mitochondria. *Curr. Genet.* **23**, 526-531.

(j) E. Ohta, K. Oda, K. Yamato, Y. Nakamura, M. Takemura, N. Nozato, K. Akashi, K. Ohyama, and F. Michel (1993) Group I introns in the liverwort mitochondrial genome: the gene coding for subunit 1 of cytochrome oxidase shares five intron positions with its fungal counterparts. *Nucl. Acids. Res.* **21**, 1297-1305.

(k) K. Ohyama, K. Oda, E. Ohta, and M. Takemura (1993) Gene organization and evolution of introns of a liverwort, *Marchantia polymorpha*, mitochondrial genome in *Plant mitochondria:with emphasis on RNA editing and cytoplasmic male sterility* (eds. by A. Brennicke and U. Kuck, VCH Weinheim) pp.115-130.

(l) K. Ohyama, K. Oda, K. Yamato, E. Ohta, M. Takemura, and K. Akashi (1995) The mitochondrial genome of a liverwort, *Marchantia polymorpha* in *The molecular biology of plant mitochondria* (eds. by C.S. Levings III and I.K. Vasil, Kluwer Academic Publishers, Netherlands) in press.

(m) M. Takemura, N. Nozato, K. Oda, H. Fukuzawa, and K. Ohyama (1995) Active transcription of the pseudogene for the subunit 7 of the NADH dehydrogenase in *Marchantia polymorpha* mitochondria. *Mol. Gen. Genet.* in press.

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 Kohchi, 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 Bioresources, 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. Katsuyuki 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. Yoshifumi 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