

Gene clusters for ribosomal proteins in the mitochondrial genome of a liverwort, *Marchantia polymorpha*

Miho Takemura, Kenji Oda, Katsuyuki Yamato, Eiji Ohta, Yasukazu Nakamura, Naoko Nozato, Kinya Akashi and Kanji Ohyama*

Laboratory of Plant Molecular Biology, Department of Agricultural Chemistry, Faculty of Agriculture, Kyoto University, Kyoto 606-01, Japan

Received January 27, 1992; Revised and Accepted May 6, 1992

GenBank accession no. M68929

ABSTRACT

We detected 16 genes for ribosomal proteins in the complete sequence 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 *Escherichia 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*). Furthermore, several ribosomal proteins encoded by the liverwort mitochondrial genome differed substantially in size from their counterparts in *E. coli* and liverwort chloroplast.

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) (1–3). On the other hand, the completely sequenced human mitochondrial genome has no genes for r-proteins (4), while the yeast mitochondrial genome encodes only one species of r-protein (5). 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 (6–8). To date, only six r-proteins (S3, S12, S13, S14, S19, and L16) have been described in the mtDNA of several species of angiosperms (9–18). We have determined the complete sequence of the liverwort mitochondrial DNA (19) and have identified genes encoding sixteen different r-proteins (S1, S2, S3, S4, S7, S8, S10, S11, S12, S13, S14, S19, L2, L5, L6, and L16). In this paper, we describe the gene organization and the characteristics of the deduced amino acid sequences of these mitochondrially encoded r-proteins.

MATERIALS AND METHODS

Cloning and sequencing of the liverwort mitochondrial DNA were performed in this laboratory as described previously (19, 20). 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.

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 24.1% to 62.1% (*E. coli*), 22.4% to 64.2% (liverwort chloroplast), and 50.5% to 76.8% (angiosperm mitochondria) (Table 1.). The low values in liverwort mitochondria vs chloroplast amino acid sequence comparisons indicate that inter-organelle gene transfer does not occur between the liverwort chloroplast and mitochondrial genomes as observed in *Oenothera rps4* gene (21).

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)(22), *spc* (L14-L24-L5-S14-S8-L6-L18-S5-L30-L15-*secY-X*) (23) and α (S13-S11-S4-*rpoA*-L17) (24). An additional cluster (*rps12-rps7*) had the same order as the homologous genes in the *E. coli str* operon (S12-S7-*fus*) (25). 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,

* To whom correspondence should be addressed

whereas a very similar clustered organization of r-protein genes exists in chloroplast genomes (1–3). 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 (25).

In the mitochondrial genomes of angiosperms, genes for *rps12*, *rps13* and *rps14* are closely linked to non-ribosomal protein genes (9–16). For example, *nad3* and *rps12* genes are co-transcribed in the wheat, maize and rice mitochondrial genomes (9, 10). Exceptionally maize mitochondrial *rps3* and *rpl16* genes are not only closely linked but even overlap, as did the liverwort *rps3* and *rpl16* genes (Fig. 3). Translation of *rpl16* in both liverwort and maize (18) mitochondrial genomes may initiate at a GUG (valine) codon. Although the maize *rpl16* gene encodes three in-frame ATG codons further upstream that could also serve as initiation codons, in the liverwort gene there is a termination codon (TAA) at 24bp upstream of GTG (valine) initiation codon (Fig. 3). 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 (27) as well as gene transfer into nuclear DNA (28). 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 (19). 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.

Table 1. 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	<i>E. coli</i>	Angiosperm mt (ref)	Liverwort cp (1)	Liverwort cp/ <i>E. coli</i> (1)
S 1	24.1		-	-
2	27.1		22.8	44.3
3	25.4	51.4 (18)	24.1	40.6
4	25.0		22.4	40.1
7	35.8		29.9	43.8
8	35.1		26.7	45.5
10	29.2		-	-
11	48.0		48.8	51.5
12	62.1	76.0-76.8 (9,10)	64.2	70.2
13	38.3	55.0-61.2 (11-14)	-	-
14	43.3	69.7 (15,16)	38.4	45.0
19	37.6	49.5 (17)	38.7	63.0
L 2	44.8		43.2	48.4
5	28.6		-	-
6	36.6		-	-
16	50.4	71.9 (18)	45.9	53.8

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 maize mitochondria appeared to be much larger than that its liverwort mitochondrial homologue (Table 2.) (18). However, liverwort and maize S3 amino acid sequences deduced from the corresponding mtDNA sequences showed a high degree of similarity between liverwort and maize in the N-terminal and C-terminal regions (Fig. 1c). 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 (7). 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 (7). 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. 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). 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.

Table 2. Sizes of ribosomal proteins from liverwort mitochondria, *E. coli*, angiosperm mitochondria, and liverwort chloroplast genomes.

Protein	Liverwort mt	<i>E. coli</i>	Angiosperm mt	Liverwort cp
S 1	270	557		-
2	237	241		235
3	430	233	559	217
4	196	206		202
7	230	178		155
8	152	130		132
10	102	103		-
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		-
6	101	177		-
16	135	136	185	143

Numbers indicate amino acid residues.

(a) *rps1*

Liverwort mt
E. coli

M--SFSQLFPKYNSFNPLRGSIAIQCSVILQONKVLVDTLGKLT----PIICFQHELKRVPIITKQARFHFGEIYEDVE-VFGEPKMLLPKPLEIKCKRKLWV
:TE::A::EESLKEIETRP::IVRGV:VAIDKDV:::A:::SESAI:AEQ:KNAQGELE:QVGDEVDVALDA::DG:::T--:SR--:--:A::HEA:

IELTKIWRSDONLVKGFILNSVKGGYAVAIAGYIAFLPKSL--RS-RKVYFSQWR-I-FSILNMKPKISNIVVKEIGDGKIDYFSPTKSHQKQTKYLGA
:T:E:AY-E:AET:T:V:NGK:::FT:ELD:IR:::G::VDV:PV:DTLHLEGKELE:KVIKLDQ:RN:V::SRRAVIESENSAERDQLLENLQEGME

KLGKHWNRMMKNTNVKKYIFSEKVPPTKTKQGFKHLGPKPLAYTEKKRETTKQSTKNNVFQKDDGGKSLVFVDVLTQSS-----
VKGIVK:LTDYGAFFVLDGGVDGLLHI:DMAWKRV::PSEIVNVGD:ITVKYL:FDRERTRVS:GLKQL:EDPWVIAIKRYPEGTKLGRVTNLTDYGCVF

EIEEGVEGLVHVSEMDWTNKNIHPSKVVNVGDVVEVMVLDIDEERRRISLGLKQCKANPWQFAETHNKGDRVEGKIKSITDFGIFIGLDGGIDGLVHLS

DISWNVAGEEAVREYKKGDEIAAVVLQVDAERERISLGVKQLAEDPFNNWVALNKKGAIVTGKVTAVDAKATVELADGVEGYLRASEASRDVEDATLV

LSVGDEVEAKFTGVDRKNRAISLSVRKDEADEKDAIATVNKQEDANFSNNAMAEAFKAAKGE 270 557 24.1%(65/270)

(b) *rps2*

Liverwort mt
E. coli
Liverwort cp

MYNSNLLVIQKLLSTNAYLGHRIPTSDFGYLYGRNEMAIIDLEKTLICLRRTCNLIGSIIISAKGH-LLLVNTNPEYNI IQQMAKTNQSYINHK
MATVSMRDMKAGVHFGHQT:YWNPKMKPFIF:A::KVH::N:::VPMFNEALAELENK:A:R::K-I:F:G:KRAASEAVKDA:LSCD:FFV::R
MKQKSWNIH:EEMMEAGVHFGHQA:KWNPKMAP:IFTE:KGIH::N:TQ:ARF:SEA:-DLVANA:S::KQF:I:G:KYQAADL:ESS:L:ARCH:V:Q:

WIGGFLLTNWKMVKKHFQDFAHPNLKDAFTSSPFDFYFRFKMKQKCFE---GIMTHNI-PDCLVIINANQNSMAILEANQLQIPIVALVDSNIPNRL
:L:M:::TVRQSI:RLK:LETQSDGTFDKLTKEALM:TEELE:LENSLG::KDMGGL::A:FV:D:DHEHI::K:::N:G::VF:I::T:SDPDG
:L:M:::STIETRLQK:K:LENKKKTGTINRLPKKEAANLKRQLDHLQKYL:G::KYMTSL::IVI::DQOKEFT::Q:CIT:G::TIC:::TDCDPM

HKLITYPVPVNDSDIKFVYLFENLITKTVILSKRSQRPKVKVRL 237
VDF---VI:G:::A:RA:T:YLGAVAA::RE-G:::DLASQAEESFVEAE 241 27.1%(64/236)
TDI---I:A:::ARASIRWIL:KL:LAICE-G:YNSI:N 235 22.8%(54/237)

(c) *rps3*

Liverwort mt
*Maize mt
E. coli
Liverwort cp

MAQKVNPIVRLNLRSSDSSWFSDDYYGKLLYQ-DLNFDRDYFGSIRPPTGNTFGFRLGRCI IHHFPKRTFIHVFFLDRLSQSRHQGLGAIPSVKLIIRRI
:R:G:::D:::P:::V:L:S:::RL:::L:::F:LPR:PLRLKRRDKSRPGKD:GRWWA
:G:::H:NGI::GIVKPNW:T::ANTKEFADNLD:SKV:Q:LTKELAKASV-----SRIVIE:PAKSIRV-----
:G:::I::LGF:::GITQNRH:Y::ANKK:S:VFEE--KKI::CI--ELYVQKHIKNSSNYGL:ARVEI::KTDLIQV-----

NDNTVKQRNEVGIWPKK---RYEYHDLPSIQK-IDQLLRVSDWMADIHSTFQS---IWPKDENDRRAS
FGKVGPIGCLHSSEGTEERNEVRGRGAGKRVESIDREKQ:IR:::MQ:G:::T:RK:NFSKS:::RAFKHPKYAGVVND:AFLI:::SFIK

EERYAFSRFAPSILVAVRAEKKAIFGSEGDFGFGTGRAFLDYFVMQYFFNLKNQIQFDPMVN-RSPVAQGVAKTSMIGKAIPAKTEQGTQSGESICQPR
TKLFX:FFLPKKSRSDG:TSHLLKRTLPAV-----:PS:N:S:::T::KMH:::V:VLNHF::P:::EP:TM:G:KGGSLDKRIR:RIAFFVES

STLYFDAI-----
:SDKKLARAKRRLIHFIRQANDLRFAGTTKTTISLFPFGATFFFSRDGTVGVNPNPFYAYAREQLLGLQLIKCRNLMGKDKVMELIEKIDLGRIGKL

IFLRYARFRKATSLSSRYYLKMQSLFNSQTKNTLIQPVKIASVYQASLIAQEISWKLEQK-SFRQICRSIFKQIKK-P-YVKGIR
IKGIEMMIE:I::KR:IPYGYN-YLNEV-Q::R:FL::R:N:::ES:::K:::D:::FQ:RNNPI:::S:FSK:V:D:PLIM:KG:E:::
-----TIHTARPGIVIGKKG--EDVE:L-RKVVADIAGVPAQ:NIAEVRKPELD:K:V:DS:TSQ::RRV-M::RAMKRAV--NAMR-L-GA:::K
-----EIYTGFPALLVE:RGQIEQ::LNVQNILSSEDRLRLMLIE::KP:GEPKIL:KK:AL:::SRV-A::RTMKKAI-ELA:K-G-NI:::K

IGCSGRLNGAEIAKTECKKYGETSLHVFSDQIDYAKTQASTPYGILGVKVVWYFLTQKKGTSCAISKTYKIS 430
:C:::G:::R:::G:::K:CN:::NQK:::PAEV::RD::S:::RI::SQ-N:::R:::E::E: 559 51.4%(207/403)
VEV:::G:::R:::WYRE:RVP::TLRAD::NTSE:H:T:VI:::IFKGEILGGMAAVEQPEKPAAPKQKQKGRK 233 25.4%(60/236)
:QIA:::RV:WARE:RVP:QTIRAR:N:CYYA:Q:I::V::I:::IFQDEE 217 24.1%(53/220)

(d) *rps4*

Liverwort mt
E. coli
Liverwort cp

MFASRFKVCROILENVWTKKLTQKFLISELQKQKKN--KKQSDFSI--QLQTIKLLSFLYGNLPIK--MQRAKTHTYIDKKN--LLFNIEK
MARYLGPKLKLSRREGTD:F::SGVRAID-T:C:IEQAPG:HGARKPRLSDYGVQ:REKQKVR:YGVLER:FRNYKEAARLKGNTGEN::ALL:G
MSRYRGPVVKIIRRLGA:P-GLTNKTLK:K-SGYI-NQS-TSNKK--VS:YRIR:EEKQLRFHYGLTER:LL:YVRIAR:AKGSTGQV::QLL:M

RLDVILVRLNFCSTMFQARQLISHKNICVNYKKNIPGFQVNSGDLISIQENSDFKSNIRKNFQTNIRRMKPNHLEVNYKTLKAVVLYEPOQIQFPY
::NVVY:MG:GA:RAE:::V:::A:M::GRV:::ASY:::PN:VV:::R:KAKKQ--RV--AALELAEQ:E::TW:::DAGKMEGTFKRK:ERSDLSA
:::N:IF::GMAP:IPG:::VN:RH:LI:NNT:D::SYNCKPK:V:T:KDR:KSQ--I:I::-LNSFQKQI:::TFDLMQI:GL:NQIIDREWIYL

KIDLDDL 196
D:NEH:IVELYSK 206 25.0%(49/196)
:NEL:VVEYYSRQV 202 22.4%(44/196)

(e) *rps7*

Liverwort mt MNLFGKSNFNCVFSSSLDFHSSRLSEKVGTKKNRICRETESFYALCLSHRRYLCLYALEGLLPSRPRGRRASTYNCSNLYGYIRGLNGKQKLIKLVHI
E. coli MPRRRVIGQRK:LPDPKFGSE:LA:F:N:MSRKSIAEKQVAKPPIYRNR:VNM::NR
 Liverwort cp CMIDGKTRSRRAIVYKTFHRLAPHGDVIKLLVN---AIENVKPICEVKKVRISGTTTLVPSIIATNRQETLAIRWMLSEAAKRRMGKKSISLDQCLYAEI
 L:V:::STAES::SALET::QRSKSE:EMAFEV:L::R:TV::SR:VG:S:YQ::VEV-PV:RNA::M::IV:A:R::GDKSMALR:ANE:SD-A
 ILKN::SLAYR:L::AMKNIKQKTKKNP:F:LR-Q:VRK:T:NVT::AR::D:S:YQ::LE:KSTQGA:::L:GASR::SGQNAFAK:SYE:ID-A
 LEASQKMGIAARKRDDLHKLAEANRSFSHYRWW 230
 A:N---K:T:V:::E:V:RM:::KA:A:::LSLRSFSHQAGASSKQPALGYLN 178 35.8%(57/159)
 ARD---N:::IR:KEET::M:::A:A:F: 155 29.9%(47/157)

(f) *rps8*

Liverwort mt MHTLSNLLSSIKNQAQKRVLYFSSFKIKSKRKRKRVCSACKMMPRVFVSRCLWDFCRILYNEGYIHGFQAEADG-S-LRIVLKYHSSGIGV---IKKM
E. coli MSMQDPIADMLTRI:NGQA:N:AAVTMPS:K:KVAIANV:KE::F:ED:KV:G:TKPE:ELT:::FQ--:KA:VES:Q--
 Liverwort cp MGNDTIANMITSIRNANLGI:TVQVPATNITRNIKILFQEGFIDNFIDNKQTKDI:ILN:::Q--:KKKSY:TTL
 KTISKPGFRIYSSKNRLSKKREGLGITLSTSKGNLICDREAQKTNFGGGEILCQVF 152
 -RV:R::L::KR:DQ:P:VMA:::AVV:::VMT--:A:RQAGL--:::I:Y:A 130 35.1%(46/131)
 RR:::L:::NHKEIP:VLG:M::V:::R:IMT--:::RQKKI--:::L:Y:W 132 26.7%(35/131)

(g) *rps10*

Liverwort mt MTAKICIVIKSFENQRSGLLNTRKIGLPPKQTLTYLVRSPHIDKKSREQFEMRIHKQLLV-IETETHKLRKLNWLKLDLGLVQ
E. coli MQNQIRIRIRLKAFDHRLIDQATAEIV:TAKRTGAQVRGP:P::TRKERF:::I:::VN:DA:D:Y:I:T:L-R:::D:VEP:E:TVDA:MR:D:AAGVD::
 VKIIFYYQTRLKVKCKS 102
 ISLG 103 29.2%(26/89)

(h) *rps11*

Liverwort mt MQKK----HGITNMQKKHCITYIQSTFGNTIITLDYNGNKTWSSSGSVGFKGSRSTNYAAQATA-ENAAVAIQLGFKFVVRKIKGLGYGKÉSSLR
E. coli :A:APIRARKRV-RK:VSDGVAH:HAS:N::V:I::RQ::ALG:ATA:GS::R::K::PF:::V:AGRC:DAVKEY:I:NL::MV::P:P:R::TI:
 Liverwort cp :P:SVKKINLRKGRRLPKGVH:AS:N::V:V::IR:QVVS::A:AC:::TKK::PF:::A:::I:IL:OQ:M:QA::M:S:P:P:RDTA::
 GLKLGGLIITKIRDVTPPHNGCRPPKRRV 125
 A:NAA:FR::N:T:::I::: 129 48.0%(60/125)
 AIRRS:I:LSFV:::M:::R::: 130 48.8%(61/125)

(i) *rps12*

Liverwort mt MPTMNQLVRKGRSKRRTKTRALNKCPOKQGVCLRVSTRSPKPNALSALRKAIVRLTNRNEIIAYIPGEGHNLQEHVSMVMVRGGRVQDLPGVYKHCIRG
 Maize mt :::K:::I:H:::E:::D:::SDQ:::T:::S:::HD:F:H:::S:::I:L:::K:S:::S:R::
 Rice mt :::K:::I:H:::E:::D:::SDQ:::T:::S:::HD:F:H:::S:::I:L:::K:S:::S:R::
E. coli :A:V:::P:AR:VAKSNVP::EA:::R:::T:Y:TT:::VCR:::GF:V:S:G:::G:G:::IL:OQ:M:QA::M:S:P:P:RDTA::
 Liverwort cp :::IQ:::I:NK:QPIENRTKSP::KG::RR:::T:Y:TT:::R:::SGF::T:::I:::L:::K:::R:::I::
 VKDLQIGIPGRRRGRSKYGTCKPKDYI 126
 :::L:::D:::K:::AER::SK 125 76.8%(96/125)
 :::L:::D:::K:::AER::SK 125 76.0%(95/125)
 AL:CS:VKD:KQA:::V:R::A 124 62.1%(77/124)
 TL:AV:VKD:QQ:::V::S: 123 64.2%(79/123)

(j) *rps13*

Liverwort mt MSYILGTNLNSNKQVKIALTRIFGIGPKKAIQVCDQLGSDTIKVNKLTKYQFDQILKIIISQNYLVDSELKRVIQRDIKRLISIGCYRGRFHNAGLPLRG
 Wheat mt :::S:ARSLPDE::R:S:KMD:::LRYR::I:GN::M:E:::I::EQM:A:DHV:HW:::GERA::E:::SR:::I:QD:S::
 Maize mt :::S:ARSLPDE::R:S:KMD:::LRYR::I:GN::IHE:::I::EQM:A:DHV:HW:::GERA::E:::SR:::I:QD:S::
 Tobacco mt :L::S:AR:VGDE::R:S:K:D:::RYR::I:GN::IKE:::I::EQM:G:DHV:HW:::GERA::E:::S:::I:QD:S::
 **Oenothera* mt :::S:AR:VADE::R:S:KMD:::RSR::--GN::RKE:::I::EQMRG:DHV:HW:::GERA::E:F:::S:::I:QD::
E. coli VAR:A:I:IPDH:HAV:::S:Y:V:KTRSKAIIAAA:IAEDV:ISE:SEG:I:TLRDEVAK-FV:EGD:R:E:SMS:::MDL:::L:RR:::V::
 QRTHTNAKTRKLYVSIKRS 120
 :::R:A::QIWK 116 57.8%(67/116)
 :::R:A::QIWKGNERRLPKEQATD 129 55.0%(66/120)
 :::R:::IWK 116 61.2%(71/116)
 ::S:::R:S::RIRK 114 56.9%(66/116)
 ::K:::R:::GPRKP:KK 118 38.3%(46/120)

(k) *rps14*

Liverwort mt MSNQ--IIRDHKRRLLVAKYELKRMHYKAIQDQNLNPKIRYEFYFKLSKLPNSSKTRVRNRCIFGTRPSRVYKLFRIIRIVFREL
 Broad bean mt ::EKR-N:::A:::R:KL:::F:K:SD::SDMWOKLRY:::FA:::S:::E:::S:
 **Oenothera* mt :-EKR-N:::AT:::R:KL:::F:N:PA::SDM:DKHRY:::FA:::E:::EF:::G:
 Yeast nc MGNFRFPIKTKLPPGFINA:L:NFK:QOFKEN:ILVKS:L:F:ARMN:::T:L:L:AQL::NA::NYMRS:QIK::VDS:HA:F:LSD::LC:YQ::N
E. coli :AK:SMKA:EV::VA:AD::FA::AEL:::IS:V:ASDED:WNAVL::QT:::D:PS:Q:::RQ:::HGFLRK:GL::KV::A
 Liverwort cp :AKKSL:Q:EK::QN:EK::KIL:NSL:KKITETSSLEDEK--FQK::QS:::AP::LHR::FL::KAN:RD:GL::HLL::M
 ASKGSILIGINKSCW 99
 ::R:P:M::K::S: 100 69.7%(69/99)
 ::R:::M::K::S: 99 69.7%(69/99)
 :L:N:P:VK:GI: 114 37.4%(37/99)
 :MR:QIP:LK:G 99 43.3%(42/97)
 :HACL:P:VT::S: 100 38.4%(38/99)



Figure 1. Amino acid alignments of r-proteins deduced from the liverwort mitochondrial DNA sequence, with counterparts from *E. coli*, the liverwort chloroplast, 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. Asterisks indicate amino acid residues corrected for RNA editing (14, 16, 17, 18). Arrowheads specify the sites of intron insertion.

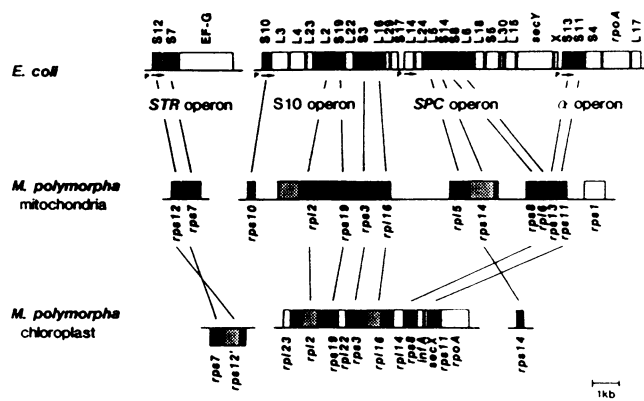


Figure 2. Organization of r-protein genes in *E. coli*, liverwort mitochondrial and liverwort chloroplast genomes. Solid boxes indicate the common genes detected between the liverwort mitochondrial genome and either the liverwort chloroplast or *E. coli* genomes. Hatched boxes are genes having introns. 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).

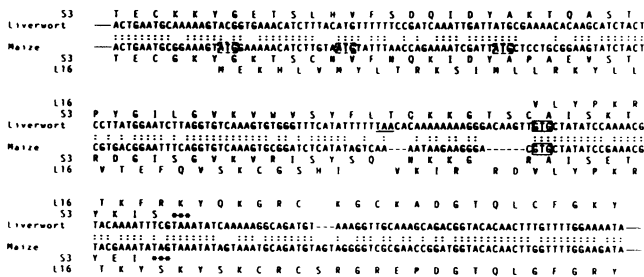


Figure 3. Overlapping regions of *rps3* and *rpl16* genes in liverwort and maize mitochondrial genomes. Amino acids are denoted by their 1-letter symbols. Initiation codon (GUG) are boxed in both liverwort and maize *rpl16* genes. Three in-frame ATG codons in maize are boxed by dashed line (18). Termination codon (TAA) 24bp upstream from the GTG initiation codon in liverwort *rpl16* gene is underlined.

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) (29). 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.

ACKNOWLEDGEMENTS

The authors thank Prof. M.W.Gray, Dalhousie University, Halifax, Canada for critical reading of the manuscript. This research was in part supported by a Grant-in-Aid for Science Research in Priority Area from the Ministry of Education, Science and Culture, Japan, and by the Research Grant from Yamada Science Foundation.

REFERENCES

- 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.
- 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., Toidoh, N., Shimada, H. and Sugiura, M. (1986) *EMBO J.* **5**, 2043–2049.
- 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.
- 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.
- Butow, R.A., Perlman, P.S., Grossman, L.I. (1985) *Science* **228**, 1496–1501.
- Myers, A.M., Crivellone, M.D. and Tzagoloff, A. (1987) *J. Biol. Chem.* **262**, 3388–3397.
- Kitakawa, M., Grohmann, L., Graack, H.R. and Isono, K. (1990) *Nucl. Acids. Res.* **18**, 1521–1529.
- Kang, W., Matsushita, Y. and Isono, K. (1991) *Mol. Gen. Genet.* **225**, 474–482.
- Gualberto, J.M., Wintz, H., Weil, J.H. and Grienenberger, J.M. (1988) *Mol. Gen. Genet.* **215**, 118–127.
- Suzuki, T., Kazama, S., Hirai, A., Akihama, T. and Kadowaki, K. (1991) *Curr. Genet.* **20**, 331–337.
- Bland, M.M., Levings, III, C.S. and Matzinger, D.F. (1986) *Mol. Gen. Genet.* **204**, 8–16.
- Schuster, W. and Brennicke, A. (1987) *Mol. Gen. Genet.* **210**, 44–51.
- Bonen, L. (1987) *Nucl. Acids. Res.* **15**, 10393–10404.
- Wissinger, B., Schuster, W. and Brennicke, A. (1990) *Mol. Gen. Genet.* **224**, 389–395.
- Wahleithner, J.A. and Wolstenholme, D.R. (1988) *Nucl. Acids. Res.* **16**, 6897–6913.
- Schuster, W., Unseld, M., Wissinger, B. and Brennicke, A. (1990) *Nucl. Acids. Res.* **18**, 229–233.
- Conklin, P.L. and Hanson, M.R. (1991) *Nucl. Acids. Res.* **19**, 2701–2705.
- Hunt, M.D. and Newton, K.J. (1991) *EMBO J.* **10**, 1045–1052.
- Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T. and Ohyama, K. (1992) *J. Mol. Biol.* **223**, in press.
- Oda, K., Kohchi, T. and Ohyama, K. (1992) *Biosci. Biotech. Biochem.* **56**, 132–135.
- Schuster, W. and Brennicke, A. (1987) *EMBO J.* **6**, 2857–2863.
- Zurawski, G. and Zurawski, S.M. (1985) *Nucl. Acids. Res.* **13**, 4521–4526.
- Cerretti, D.P., Dean, D., Davis, G.R., Bedwell, D.M. and Nomura, M. (1983) *Nucl. Acids. Res.* **11**, 2599–2616.

24. 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.
25. Post, L.E. and Nomura, M. (1980) *J. Biol. Chem.* **255**, 4660–4666.
26. Gray, M. W., Cedergren, R., Abel, Y. and Sankoff, D. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2 267–2271.
27. Palmer, J. D. and Shields, C. R. (1984) *Nature* **307**, 437–440.
28. Stern, D. B. and Lonsdale, D. M. (1982) *Nature* **299**, 698–702.
29. Ohyama, K., Ogura, Y., Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T. and Yamada, Y. (1991) In Osawa, S. and Honjo, T.(eds.), *Evolution of life*. Springer-Verlag, Tokyo, pp.187–198.