Analysis of NADH dehydrogenase proteins, ATPase subunit 9, cytochrome *b*, and ribosomal protein L14 encoded in the mitochondrial DNA of *Paramecium*

A.E.Pritchard, C.L.Sable, S.E.Venuti and D.J.Cummings*
Department of Microbiology and Immunology, University of Colorado Health Sciences Center, Denver, CO 80262, USA

ABSTRACT

The mitochondrial (mt) encoded *ndh1*, *ndh3*, *ndh4*, *ndh5*, *rpl14*, *cyt b* and *atp9* gene products were identified by sequence comparisons with known proteins. Amino acid sequence comparisons between predicted *Paramecium* mt gene products and proteins in current databases were quantitated approximately by the means of similarity scores for pairs of aligned sequences. The comparisons show that the *Paramecium* gene products are very divergent from all others with the exception of those from a closely related ciliate, *Tetrahymena*. The similarity scores of comparisons between a *Paramecium* mt DNA encoded protein, cytochrome b for example, and the homologous protein from a group of organisms as diverse as other protozoans, vertebrates, fungi, plants, and prokaryotes were all about the same. The *Paramecium* gene products appear to be equally divergent from proteins representing a number of different kingdoms and organelles.

INTRODUCTION

The identities of a few *Paramecium* mt DNA encoded proteins have already been established by means of comparisons with sequences of known polypeptides (1, 2, 3). These include cytochrome *c* oxidase subunits one (COI) and two (COII), a part of NADH dehydrogenase subunit 1 (protein ND1 encoded by the gene locus *ndh1*), ribosomal proteins L2, S12, S14, and photosystem II protein G. Two other NADH dehygrogenase subunits found encoded in *Paramecium* mt DNA are somewhat unusual (2). ORF400 was identified by its similarity to homologous ORFs in chloroplasts (cp), in the protozoan kinetoplastid DNA of *Leishmania tarentolae*, and now in the bovine nuclear genome (4). These ORFs were shown to be an additional NADH dehydrogenase subunit (4) not encoded in animal or fungal mt genomes. The *Paramecium* gene corresponding to *ndh2* has also been identified but appears to be lacking the 5' end found in most other mt and cp genomes (2). The amino acid sequence of the region that is encoded is

very divergent compared to known ND2 sequences, even for a *Paramecium* gene product, but this region, of all the *Paramecium* mt genome, is the most similar to known ND2 genes.

In general, proteins encoded in the *Paramecium* mt genome have been found to be quite divergent compared to analogous gene products from other sources. Only about 26% of the amino acids in the *Paramecium* COI protein are identical to the aligned residues from other known sequences (1), about 23% for ribosomal protein L2 (2), and less than 20% for ND2 (2). On the other hand the *Paramecium* ORF400 shows 34% identity in alignments with the corresponding gene encoded in cp genomes and the *Paramecium* ribosomal protein S14 shows about 32% identity with the cp DNA encoded proteins (2).

In this communication we identify, by amino acid sequence comparisons, additional genes encoded in the *Paramecium* mt genome that have previously been found encoded in mt, cp, or nuclear genomes of different organisms. As with previously published *Paramecium* mt gene products, the sequences are divergent compared to those from a wide range of taxonomic groups. These results are consistent with the idea that *Paramecium* mt DNA branched from other groups early in evolution and evolved very rapidly.

MATERIALS AND METHODS

The similarity scores tabulated in Figs. 2 and 9 were obtained from the version of FASTP (7) supported on the BIONET computer resource while the sequence alignments shown in the other figures were obtained from the IBI/Pustell version of the

---

* To whom correspondence should be addressed

Cytochrome b

```
                 10        20        30        40        50        60        70 +      80    +  90       100
Paramecium: ILNIFNYFKNLRVSFHEVFSLFGFFTFMTIIVQLVSGTMLAFSSVPEPMLIPTVRDEEDIEDLYTDDFFWLHERGVDLIFIFSYFHLLRKLYLNVFDLET
Yeast                                          MM
[284]    38                     LCLVI*I*T*IFM*Y**NI*LAFSSVEHIMR*VHNG*I--LRY**AN*ASFF*MVMFM*MAKG**YGSYRSPR>
                                A AAAAAA A AAAAA  A A    AA A A   A  AA A   A AA   A AA A AA   A

R.sphaeroides                                                     V
[264]    51                      LA*-*LVL*I*T*IV**IDYT*HVD*-AFASV*HIMR*VNGGAMRYI*AN*AS*F*LAV*I*IF*G**YGSYKAPR>
                                A A AAAAAAA A AAA   A   A    A  A AA   AA A  A AA  A AAAA AA   A

Human                                       MM
[255]    33                    **SLLGACL*L*ITT*LF**Y*PDASTAFSSIAHITR*V-N-*GWIIRY**AN*ASMF**CLFL*IG*G**YGS*-*YS>
                               AA A   AAAAA A AAAA   A  A    AA A A   AA A A AA   AAAA A AA   A A

T.brucei                                        FI                                              L
[236]    20           *I-YGVG***-***IALQ**CG-*CLAW*F**CCSNWYFVLFLW*-F*L-G-FV--IRSV*ICFTS*LYLLL*I*IFKSIT*II**THI>
                      AA A  AAA AAA  A AA  A  A AA     AA AAA AA  A     AA    AAAAA A AAAA A A AAA

Maize mt                                                         Y
[205]    38                   **CLAGICLVI*I*T*VF**MHYT*HVD*-AFNSV*HIMR*VEGGWLLRM*AM*ASMFL*VVHL*IF*G**HASYSSPR>
                              AA A  A AAAAAA A AA   A   A    A  A AA   AAAAA A A AA   AAAAA AA   A
```

```
                 110       120       130       140       150       160       170 +      180   +190       200
Paramecium: EASWKSGVFSFLVFQVVVVFFGLVLCCTHLSEITLTIAANIFHTFFNFKGKAYWFLFTDKQLNTDTLIRLAYAHYVSAFYLSFLGLLHGIDIHYOWKNEPF
Yeast                      VV                          P
[284]   111 VTL*NV**II*-ILTIATA*-*GY**GQM*HWGA*VIT*L*SAI*VGNDIVS*-*WGGFSVSNP*IQ*FFAL**LVP*IIAAMVIM*LMAL*IHGSSN*L>
            A AA  A A AA  A A   AA AAA  AA AAA   AA   A A    A   A  AA AAA   AAA A A A AAA A AA       AA

R.sphaeroides                                  G
[264]   126 *IT*IV*MVIY*LMMGTA*M*Y**PWGQM*FWGA*VITGL*GAIPGIPSIQA*-*LGGPAVDNA**N*FFSL**LLP*VIAA*VAI*IWAF*TTGN*N*->
            A A  AA AAA   A A AA   A A AA   AAA  AA AA    A A    AA  AA AA AA AAA A A A AA   A    A A

Human                                          P                                          F
[255]   111 *-T*NI*IILL*ATMATA*M*Y**PWGQM*FWGA*VIT*LLSAIYIGTDLVQ*-IWGGYSVDSP**T*F-FT*FILP*IIAA*AT**LLFL*ETGS*N*L>
            A A  AA AA  A A A AA   A A AA   AAA  AA AAA   A A  A  AA A A  AAA A A A AA   A  AA A AA   A AA

T.brucei                    A                     W
[236]   104 -LV*FI*-*ILF**IIII*I*Y**P**MM*YWG**VFS**IA*VPILGILC**-IWGSEFI*DF**LK*HVL*VLLP*I*LIILI**LFCL**FMSSDA*>
            A A  AAAA AAAA A Y  A  AA AA  A   AAA  AA A  A   AA AAAAA A  A A A A A AAA   AAA    A A

Maize mt                                           I                                        Q
[205]   117 *FV*CL**VI**LMI*TA*I*Y*PPWGQM*FWGA*VITSLASAIPVVGDTVT*-*WGGFSVDNA**N*FFSL*HLLPLI*AGAS***LAAL**G-S*N*L>
            A A  A  AA AAA  A A  A AA A  A AA   AAA  AA    A A    A A A AA AA  A A A A A A AAA    AAA   A AA
```

```
                 210       220       230       240       250       260       270       280       290       300
Paramecium: YDGLSSEMLWWDEALSMELTNFFVLLVFITLAFFLLFEEPEALSYEIFNWGDIGLSTDVRFYGVAPHWYFRPFMAWLIACPFHKTGIFGLLFFFVTLYYQ
Yeast             F                               P
[284]   211 GITGNLDRIPNHSYFIKD*VTV*-*FML*-**L*VFY-S*NT*GQDNYIP*N-P*V*PA---SID*E**LL**Y*I*RSI*-D*--LL*VITM*AAILVL>
              A A   A A    A AAAAA AAAAAAA A A  AA AA  A A   AA AAA AA A  A A    A AAAAA A A

R.sphaeroides              FW KDLFA              Y        P                                              GIV
[264]   225 -T*VEVRRTSKAD*EKDT*PPY**I*ALVL*G**AVVAM*NY*GHDNYIQAN-P***PAHI--*-*E***L**Y*-ILR-A*-AADVW-VVILVDG*TFD>
             AA     AA A A  AAAAA AA A AA  A  A A   A AAA   A A    A A AAA AA A A A A  A   A  A    A AA   A AA

Human                                              P                                F
[255]   210 GITSH*DKITFHPYYTIK-DALGL**FLLS*MTLT**-S*DL*GDDNYTLAN-P*N*PPH---IK*E***LAYTI-*RSV*NKLG*VLA**LSILI*AMI>
            AA  A    A        AAA  AA AA  A A     AA A  A AA A A   A A AAA  A   A    A  AAA AAA  A A

T.brucei
[236]   203 C*RFAFYCERLSFCNWFY*RDM*LAFSILLCMNYVI*INW-YFVFHEES*VIVD-TLKTS-DKIL*E*F*LYLFGF*K*I*DKFM*L*LMVILLFS*F>
            A A     A     A   A A  A A  AAAA   A A A AA  A A    A A AAA  A A A    AAA AAA AA AA  A

Maize mt                    KD     I SI          P
[205]   217 --*VH***---*K-IAS-YPY*Y**VGRVAS***WI*FA*NV*GHDNYI-PANPMP*PPHI--*-*E***L*IH*I*RSI*DKAG*VAAIAPV*IS*>
            AA AAA  A  A    AAAAA A AAA AA  A A  A AA   A A A AAA A A A    A  AA A  AA A   AA A   AA A
```

```
                 310       320       330       340       350       360       370       380       390
Paramecium: PNLHGVSDQNSYGKKTLTISSTVLAKKNTATPFSISIDSNLYHQITYFFFIMCCLYTPSFLPYGRFFNQIGGNWGFLFSYFYVFCYLAFTD
Yeast
[284]   303 LV*P-FT*RSVVRGN*FKVL*KFFF-FIFVFN*VLLGQIGAC*-VEVPYVL*GQIA*FIYFA*--*LIIVPVISTIENVL**I>
            A   A    AA A A    AA A A    A A A  A A A    A AA A A A AA A AA A    AAA

R.sphaeroides                                W
[264]   328 AKFF**IAMFG-AIAVMALAPWLDTS*VRSGAYRPKFRMWFFLVLDFVVLTWVGAMPTEY-**DWISLIASTY*-*AY-FLVILPL*GA*E>
            A AA       A A   A A    A   A  A     AA A A    A AA A   A A A  AA AA  A  AA

Human
[255]   305 *I**-M*K*Q*MMFRP*SQ*LYW*LAADLLILTW*GGQPVS*-PF*IIGQVASV**FTTI*>
            A AA AA A A   A A A  A A    A A    A A A A    A  AA  A
```

Figure 1. Amino acid sequence alignments of cytochrome b encoded in *Paramecium* mt (top line, complete sequence), yeast mt (8), *Rhodopseudomonas sphaeroides* (9), human mt (10), *Trypanosoma brucei* mt (11), and maize mt (12) DNAs. Below each organism name is the FASTP similarity score in brackets; the next number, outside of the brackets, is the number of the first amino acid aligned on that line. Letter(s) above a sequence line represents amino acid 'inserts' at that point in the sequence which are necessary for alignment purposes; a dash represents the absence of an amino acid, and an asterisk represents identity with the top *Paramecium* reference sequence. The carets below a sequence represent scores > +1 in the FASTP scoring for conservative amino acid replacements (6). The 4 invariant histidines (+) are at positions 72, 86, 173, and 187 in the *Paramecium* sequence.

program (6). The slight differences in scoring were insignificant for short proteins but for longer ones the BIONET scores were slightly different from the IBI/Pustell scores. The small differences probably reflects the uncertainty inherent in the algorithm and either score was considered valid. For more

Cytochrome b

ORGANISM	SIZE	% IDENTITY RELATIVE TO:				SCORES RELATIVE TO:			
		Para	T.bru	R.sph	Yst	Para	T.bru	R.sph	Yst
Paramecium	391	100	22	25	21	2255	210	219	245
Yeast	385	21	24	49	100	245	474	862	2089
A.nidulans	387	21	28	55	62	232	471	855	1411
N.crassa	385	20	25	48	42	186	401	786	1300
Human	380	24	26	47	50	215	457	853	1155
Mouse	381	22	26	47	50	207	452	855	1170
D.yakuba	378	21	27	48	54	192	453	863	1190
T.bru	363	22	100	25	24	210	2352	415	474
Maize mt	388	21	26	55	51	176	421	916	1155
R.sph	437	25	27	100	49	219	412	2441	862
LivCpb6	215	--	25	35	32	---	240	448	381

Figure 2. Cytochrome b similarity comparisons. The % of the amino acids that are identical at corresponding positions for each paired alignment, and the FASTP program similarity score obtained for each alignment are given. The proteins that are not shown in the alignments in Fig. 1 are encoded in *Aspergillus nidulans* mt (13), *Neurospora crassa* mt (14), mouse mt (15), *Drosophila yakuba* (16), and tobacco cp (cytochrome b6, ref. 17) DNAs. Abbreviations are T.bru for *Trypanosoma brucei* and R.sph for *Rhodopseudomonas sphaeroides*. No valid alignment was obtained between the *Paramecium* and liverwort protein sequences.

accurate comparisons, however, all scores considered should be from the same program.

Since the *Paramecium* mt gene products are quite divergent, identities were based on a number of considerations including: (1) similarity scores, (2) an appropriate size for the gene between possible initiation codons ATG, ATA, ATT, ATC, GTG and stop codons TAG and TAA (ref. 5), (3) regions in the sequence alignments that were conserved in almost all of the possible pairwise comparisons, and (4) specific amino acids, such as histidines, that had previously been observed to be invariant.

RESULTS AND DISCUSSION
Cytochrome b

The *Paramecium cyt b* gene product is identified by the aligned sequence comparisons shown in Fig. 1 and is further characterized by the relative comparisons summarized in Fig. 2. This gene product is typical of most of the *Paramecium* genes in terms of its extreme divergence from most of the other known cytochrome b proteins. This particular gene product is analyzed in more detail because many cytochrome b sequences from a wide variety of organisms and organelles are available. Since the *Paramecium* gene product is so divergent, it was difficult to identify. We rely on FASTP analysis which scores conserved amino acid replacements (shown as carets below the sequence line in Fig. 1) as well as identities at corresponding positions

Figure 3. Amino acid sequence alignments with ND1 encoded in bovine mt (22), *Marchantia polymorpha* cp (Liverwort Cp, ref. 23), and *Aspergillus nidulans* mt (24) DNAs. The + marks above two regions of the *Paramecium* sequence identifies the homologous segments noted in ref. 21. Other symbols are given in the Fig. 1 legend.

```
ND1
ORGANISM        Para  Bovine  LivCp  Human  Mouse  Droso  N.cras
SIZE            261    318    369    318    315    324    371
%ID vs Para     100     32     33     32     31     34     35
%ID vs Bovine    32    100     38     78     79     48     42
SCORE vs Para  1452    457    454    428    442    428    428
SCORE vs Bovine 457   1677    533   1403   1449    861    685


ND3
ORGANISM        Para  LivCp  TobCp  Podo  Human  Mouse  Starfish
SIZE            120    120    120    130    115    114    116
%ID vs Para     100     25     22     24     21     22     28
%ID vs LivCp     25    100     72     35     31     28     28
%ID vs Mouse     22     28     32     28     64    100     49
SCORE vs Para    707   198    170    169    125    133    175
SCORE vs Mouse   133   215    233    189    464    623    290
SCORE vs LivCp   198   707    526    227    255    215    248


ND4
ORGANISM        Para  T.bru  LivCp  Droso  A.nid  Human  Starfish
SIZE            474    439    499    446    488    459    460
%ID vs Para     100     25     23     27     22     23     22
%ID vs Human     23     21     27     41     28    100     43
SCORE vs Para   2645    475    418    466    424    359    321
SCORE vs Human   359    258    521   1049    577   2378   1201


ND5
ORGANISM        Para  PrirMt  LivCp  TobCp  N.cras  Bovine  T.bru
SIZE            570    591    693    710    715    606    590
%ID vs Para     100     29     27     26     28     26     25
%ID vs LivCp     27     41    100     56     30    31.    26
%ID vs PrirMt    29    100     41     41     49     36     28
SCORE vs Para   3127    734    698    653    707    682    635
SCORE vs LivCp   693    827   3854   2226    953    722    532
SCORE vs PrirMt  734   3110    827    807   1203    796    437
```

Figure 4. NADH dehydrogenase similarity comparisons. Selected sequence alignments for proteins ND1, ND3, ND4, and ND5 from a variety of organisms are summarized by the % of the amino acids that are identical at corresponding positions for each paired alignment (%ID vs ————), and by the FASTP program similarity score obtained for each alignment (SCORE vs ————). The proteins that are not shown in the alignments in Figs. 3, 5, 6, 7 are ND1 encoded in human mt (10), mouse mt (15), Drosophila yakuba mt (Droso, ref. 16), and Neurospora crassa mt (N.cras ref. 21) DNAs; ND3 encoded in starfish mt DNA (25); ND4 encoded in Trypanosoma brucei mt (T.bru, hypothetical protein 8, ref. 26), Drosophila yakuba mt (16), Aspergillus mt (A.nid, ref. 27), human mt (10), and starfish mt (25) DNAs; ND5 encoded in primrose mt (PrirMt, ref. 28), tobacco cp (TobCp, ref. 17), bovine mt (22), and T. brucei mt (hypothetical protein C-590, ref. 26) DNAs. Para is for Paramecium, and LivCp is for liverwort cp.

in aligned sequences. In addition, there are, in cytochrome b, four invariant histidine residues that have been identified and are believed to bind two heme groups (18). These histidines are present in the Paramecium gene product as shown in Fig. 1.

In a comparison of sequences scored relative to that of the Paramecium gene product, Fig. 2 shows uniformly low scores and 20–25% identity. There is no significant difference between similarity scores of the Paramecium gene product compared to the analogous protein encoded in the mt DNAs of the protozoan, T. brucei (210), the vertebrate, mouse (207), and the plant (176). If the gene product from any other organism is used as the reference sequence, the comparative similarity scores are higher than those compared to the Paramecium gene product sequence. Shown in Fig. 2, for example, are scores relative to the yeast gene product. All scores are above 1000 (50–62% identity) in comparisons with the protein encoded in vertebrate, insect, fungal, and plant mt DNAs. The highest scores in this comparison are with the other fungal genes, as expected. The gene product from another protozoan, T. brucei, has low similarity scores in most comparisons including the comparison with the Paramecium gene product which shows only 22% identity.

The similarities relative to the purple bacterium, R. sphaeroides, are interesting because previous studies based on rRNA sequence comparisons suggest a close evolutionary link between mt and purple bacterium DNA (19) while cp DNA is more closely related to cyanobacteria (20). However, when cytochrome b from R. sphaeroides is compared with the cytochrome b encoded in Paramecium mt DNA and cytochrome b6 encoded in liverwort cp DNA, the liverwort protein has a higher measure of similarity. These results probably reflect a rapid rate of divergence of Paramecium and T. brucei mt genes.

NADH Dehydrogenase Proteins

The ND1 protein alignments shown in Fig. 3, are in agreement with features of a comparison in a previous study (21). Burger and Werner noted regions of high similarity which are also relatively conserved in the Paramecium gene as shown in Fig. 3. These conserved regions are the most polar in an otherwise hydrophobic protein and may be the functional domains. The sequence comparisons, summarized in Fig. 4, show that a variety

```
ND3                                                    +++++                                      +++
                 10        20        30        40        50        60        70        80        90       100       110       120
Paramecium: MGSNTLLFFVEHVFIFCNIFWLLTWVAEYFFKSKNNKQKHQFYECGIRALSELNIQINLNFSIVCVFLILYDVEFIFNYPFFFNFFLVNAGAFLVFFVFLFFVFYSLVYDSVQNSLALQL
Liverwort Cp
[204]    11      FV*LLIIS*FSIL**S*SK*I*P--IN-*-GPE*FTS**S**EPNG*AC**FQIRYYNFALVFVIF***TV*L**WANS*YNFGISS*IEALI*ILILIIG***AWRKGA*
                 ^^^^^ ^ ^^^ ^ ^^^     ^ ^^  ^^ ^^  ^ ^^ ^ ^ ^ ^^^^^^^ ^^^^^  ^^   ^^ ^^^^^ ^  ^^^       ^
Tobacco Cp
[170]     9      **WAFLI*SILVPI*AFLISGVLAPISKGPE*LST**S**EPNGDAWL*FRIRYYNFALVFVVF***TV*L**WANS*DVLGVSV*IEA*I*VLILIIG***AWRKGA*
                 ^^ ^ ^ ^^ ^ ^  ^  ^       ^^ ^^ ^^ ^ ^^ ^ ^ ^ ^^^^^^^ ^^^^^  ^ ^^   ^^ ^^^^^ ^  ^^^       ^
P. Anserina                                            F
[169]     1    *S***-**IL-F*S*IALL*-*F--INLI*APHNPYQE*YSIF***FHS*GQNRT*FGVK*F*FALVYL*L*L*ILLTF**AVSEYVN*IYGLIILLG*ITIITIGF**ELGKSA*KID
               ^ ^^^ ^^ ^ ^ ^ ^^^ ^^ ^   ^     ^^  ^^^^ ^ ^ ^  ^ ^ ^ ^ ^^ ^^^ ^^ ^^^     ^^ ^  ^^^^^ ^^ ^    ^^^^     ^ ^^
Mouse                                                                                                                         A
[139]     3      *YT*--I**NILLSLT*IL**-FWLPQN*LYSEANP****FDPT*SARLPFSNK*FL*AITFL*F*L*IALLL*LPWAIQTIKTSTNNIN*ILVTILSLG*A*ENT*KG*
                 ^^ ^ ^^^ ^^ ^ ^^ ^ ^    ^^^^     ^ ^^ ^^^^^^^^ ^^ ^^ ^ ^ ^^^^^^^ ^^ ^^   ^^ ^^^^^ ^^ ^   ^^^^ ^ ^^
Human                                                                                                     L
[133]     1    MNFA*ILNINTLLA-LLLNIITF*LP--QLN--GYNE*STP****FDPN*PARVPFSNK*FL*AITFL*F*L*IALLL*LPWALQTT*LPLNVNSSLL*IIIAL**A*EWL*KG
               ^^^ ^ ^^ ^^ ^ ^^    ^     ^^ ^^^^ ^^ ^^ ^ ^ ^^^^^^^ ^^ ^^   ^^ ^^^ ^ ^^ ^^ ^^
```
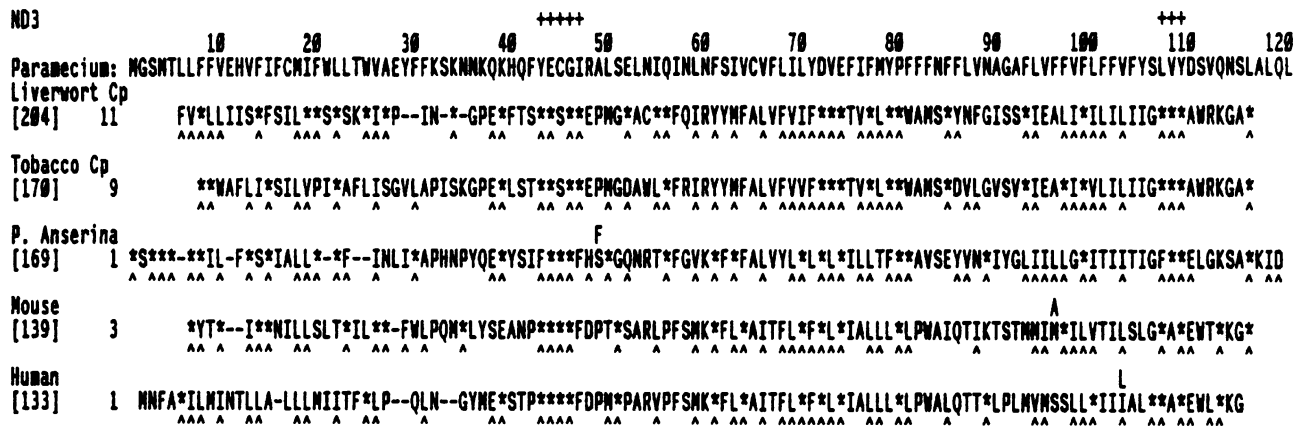
Figure 5. Amino acid sequence alignments with ND3 encoded in liverwort cp (23), tobacco cp (17), Podospora anserina mt (29), mouse mt (15), and human mt (10) DNAs. The + marks above the Paramecium sequence denote conserved regions.

of organisms have a uniform and moderately low (31−35% identity) degree of similarity with the *Paramcium* gene product. When comparisons are made, for example, relative to the bovine ND1 protein, high degrees of similarity (79% identity) are seen with other vertebrate genes, as expected. Genes from non-vertebrates show smaller similarity scores but even the liverwort cp gene has a slightly higher similarity score than *Paramecium*.

The *ndh3* gene sequence was previously published as *Paramecium* ORF2 (1), but has since been identified as encoding the ND3 protein after the liverwort and tobacco cp *ndh3* gene
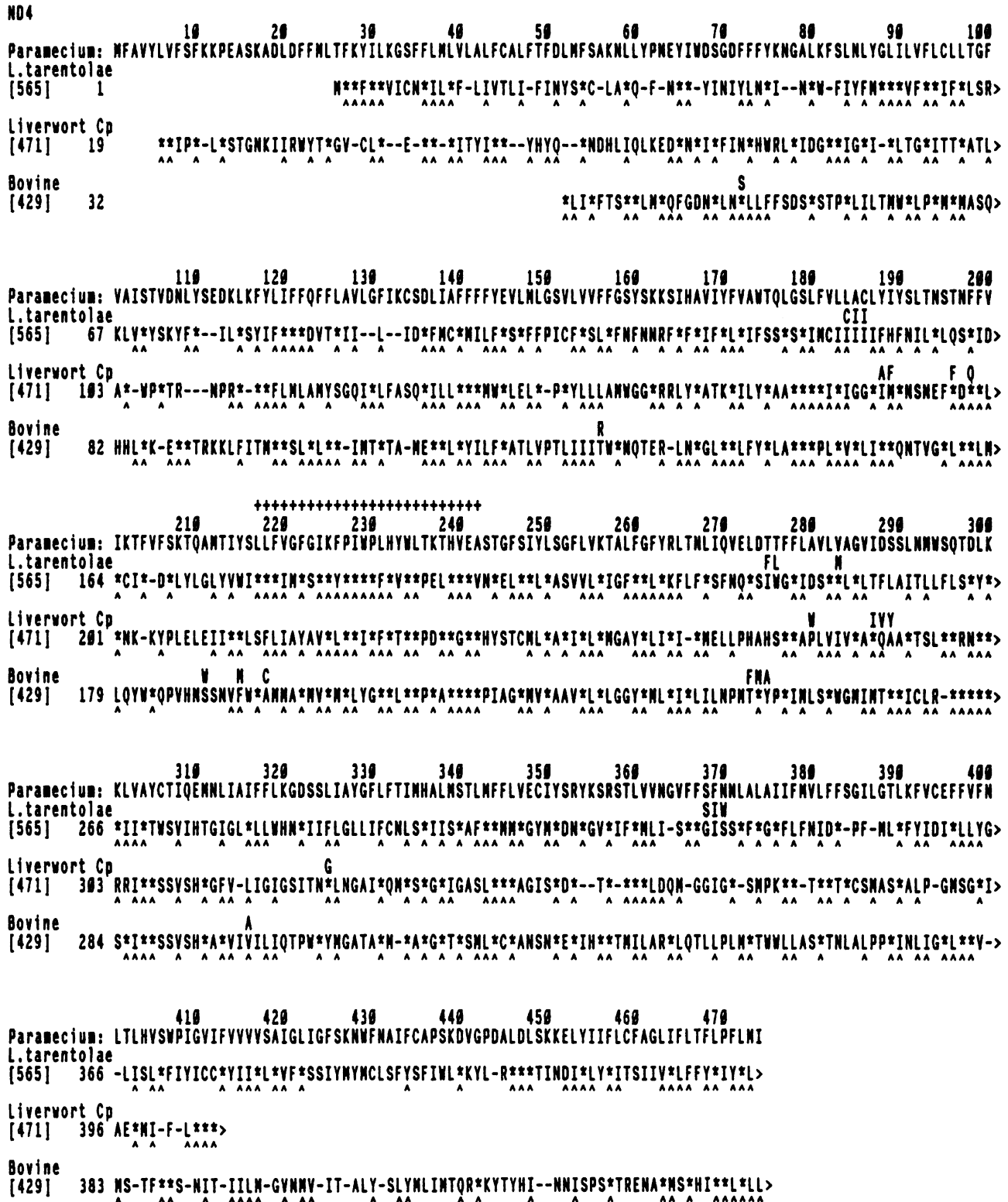
```
ND4
                10        20        30        40        50        60        70        80        90        100
Paramecium: MFAVYLVFSFKKPEASKADLDFFHLTFKYILKGSFFLHLVLALFCALFTFDLHFSAKNLLYPMEYIWDSGDFFFYKNGALKFSLNLYGLILVFLCLLTGF
L.tarentolae
[565]    1                            M**F**VICN*IL*F-LIVTLI-FINYS*C-LA*Q-F-N**-YINIYLN*I--N*W-FIYFN***VF**IF*LSR>
                                      ^^^^^  ^^^^ ^  ^   ^   ^   ^^    ^^ ^  ^   ^ ^ ^^^^ ^^ ^^

Liverwort Cp
[471]   19         **IP*-L*STGNKIIRWYT*GV-CL*--E-**-*ITYI**---YHYQ--*NDHLIQLKED*N*I*FIN*HWRL*IDG**IG*I-*LTG*ITT*ATL>
                   ^^ ^ ^   ^      ^  ^ ^   ^ ^    ^^     ^^ ^^ ^^^ ^ ^^   ^     ^^ ^  ^ ^ ^  ^^^ ^ ^^ ^^  ^ ^

Bovine                                                                     S
[429]   32                           *LI*FTS**LH*QFGDN*LN*LLFFSDS*STP*LILTHW*LP*N*NASQ>
                                     ^^ ^  ^^ ^^^  ^^ ^^^^^  ^   ^ ^  ^ ^^ ^ ^^

                110       120       130       140       150       160       170       180       190       200
Paramecium: VAISTVDNLYSEDKLKFYLIFFQFFLAVLGFIKCSDLIAFFFFYEVLNLGSVLVVFFGSYSKKSIHAVIYFVANTQLGSLFVLLACLYIYSLTNSTNFFV
L.tarentolae                                                                                          CII
[565]   67     KLV*YSKYF*--IL*SYIF***DVT*II--L--ID*FNC*NILF*S*FFPICF*SL*FNFNNRF*F*IF*L*IFSS*S*INCIIIIIFHFNIL*LQS*ID>
               ^^    ^^    ^ ^^^^^^  ^ ^  ^   ^^^ ^ ^^^ ^^  ^   ^  ^^ ^^  ^  ^ ^ ^^ ^^^    ^ ^^ ^^  ^ ^

Liverwort Cp                                                                                        AF      F Q
[471]  103   A*-VP*TR---NPR*-**FLHLANYSGQI*LFASQ*ILL***NW*LEL*-P*YLLLANWGG*RRLY*ATK*ILY*AA****I*IGG*IN*NSNEF*D**L>
             ^   ^    ^^ ^^^^ ^      ^^^ ^^^ ^^^ ^^ ^^  ^ ^^^^  ^^ ^  ^^ ^ ^^^^^^^ ^ ^^^*^^^ ^  ^^^^^

Bovine                                                                   R
[429]   82   HHL*K-E**TRKKLFITH**SL*L**-IHT*TA-NE**L*YILF*ATLVPTLIIITW*NQTER-LN*GL**LFY*LA***PL*V*LI**QNTVG*L**LN>
             ^^ ^^^   ^  ^^ ^^^^^ ^^ ^  ^^^ ^^ ^^  ^^ ^^ ^^^^  ^   ^ ^^^ ^^^^ ^  ^^^ ^^^^ ^^^    ^ ^^^^

             +++++++++++++++++++++++++++++
                210       220       230       240       250       260       270       280       290       300
Paramecium: IKTFVFSKTQANTIYSLLFVGFGIKFPIWPLHYWLTKTHVEASTGFSIYLSGFLVKTALFGFYRLTNLIQVELDTTFFLAVLYAGVIDSSLNNWSQTDLK
L.tarentolae                                                                            FL    N
[565]  164   *CI*-D*LYLGLYVWI***IN*S**Y****F*V**PEL***VN*EL**L*ASVVL*IGF**L*KFLF*SFNQ*SIWG*IDS**L*LTFLAITLLFLS*Y*>
             ^ ^ ^    ^ ^ ^  ^^^^ ^ ^^^^^^^^ ^^   ^^^ ^ ^^^ ^   ^^^ ^^^^^^^ ^  ^^   ^^ ^^ ^^    ^ ^   ^ ^

Liverwort Cp                                                                            W     IVY
[471]  201   *NK-KYPLELEII**LSFLIAYAV*L**I*F*T**PD**G**HYSTCHL*A*I*L*NGAY*LI*I-*NELLPHAHS**APLVIV*A*QAA*TSL**RN**>
             ^  ^   ^  ^^^ ^^^ ^ ^^^^^ ^^^ ^^ ^^ ^^   ^ ^ ^^^^  ^^^ ^^ ^^    ^ ^^^ ^ ^^  ^  ^^ ^^^

Bovine          W    N C                                                                FNA
[429]  179   LQYW*QPVHNSSNVFW*AHHA*NV*N*LYG**L**P*A****PIAG*NV*AAV*L*LGGY*NL*I*LILNPHT*YP*INLS*WGHIHT**ICLR-*****>
             ^ ^   ^    ^^ ^ ^ ^^ ^ ^^^ ^ ^^^ ^^  ^ ^^ ^^^  ^^^ ^^ ^^^ ^^ ^  ^ ^   ^^ ^^  ^^ ^ ^ ^^^^^

                310       320       330       340       350       360       370       380       390       400
Paramecium: KLVAYCTIQENNLIAIFFLKGDSSLIAYGFLFTINHALHSTLNFFLVECIYSRYKSRSTLVVNGVFFSFNNLALAIIFNVLFFSGILGTLKFVCEFFVFN
L.tarentolae                                                                            SIW
[565]  266   *II*TWSVIHTGIGL*LLWHN*IIFLGLLIFCNLS*IIS*AF**NN*GYN*DN*GV*IF*NLI-S**GISS*F*G*FLFNID*-PF-NL*FYIDI*LLYG>
             ^^^^  ^    ^  ^^^  ^  ^ ^^^ ^ ^ ^^ ^ ^^ ^^^  ^^ ^^ ^  ^ ^   ^^^  ^ ^^^^  ^ ^ ^  ^ ^  ^^ ^^^^

Liverwort Cp                    G
[471]  303   RRI**SSVSH*GFV-LIGIGSITN*LNGAI*QN*S*G*IGASL***AGIS*D*--T*-***LDQN-GGIG*-SNPK**-T**T*CSNAS*ALP-GNSG*I>
             ^ ^^^  ^ ^  ^^ ^  ^  ^   ^  ^^   ^  ^ ^ ^^  ^^^^  ^ ^  ^ ^^^^^ ^  ^  ^ ^^ ^^ ^ ^ ^  ^  ^

Bovine                          A
[429]  284   S*I**SSVSH*A*VIVILIQTPW*YHGATA*H-*A*G*T*SHL*C*ANSH*E*IH**THILAR*LQTLLPLH*TWWLLAS*TNLALPP*INLIG*L**V->
             ^^^^  ^  ^ ^^  ^ ^^   ^  ^       ^  ^ ^ ^ ^^^  ^  ^^ ^^ ^^  ^   ^ ^^ ^^ ^    ^  ^^ ^^ ^^^^

                410       420       430       440       450       460       470
Paramecium: LTLHVSWPIGVIFVVVVSAIGLIGFSKNWFNAIFCAPSKDVGPDALDLSKKELYIIFLCFAGLIFLTFLPFLNI
L.tarentolae
[565]  366   -LISL*FIYICC*YII*L*VF*SSIYNYNCLSFYSFIWL*KYL-R***TINDI*LY*ITSIIV*LFFY*IY*L>
             ^ ^^     ^ ^^^ ^^ ^       ^    ^ ^    ^^^  ^^^^ ^^  ^^^^ ^^ ^^^

Liverwort Cp
[471]  396   AE*NI-F-L***>
             ^ ^   ^^^^

Bovine
[429]  383   MS-TF**S-NIT-IILN-GVNNV-IT-ALY-SLYNLINTQR*KYTYHI--NNISPS*TRENA*NS*HI**L*LL>
             ^   ^^  ^ ^^^^ ^ ^^    ^  ^^    ^ ^   ^ ^   ^^ ^  ^^^^^^
```

Figure 6. Amino acid sequence alignments with ND4 encoded in *Leishmania tarentolae* mt (30), liverwort cp (23), and bovine mt (22) DNAs. The + marks above the *Paramecium* sequence denotes a region discussed in the text.

ND5

```
              10        20        30        40        50        60        70        80        90       100
Paramecium: MFSFFFSFYALSVIFSLLFKHFLSSKGVFLLNTTSIGLFWAYSLSNLNLFFIKNKLIAIHLFRWFPLSAGYLVNFSFYIDTVAYSFTLLTLTIGVFVNLY
N.crassa                                                           GF
[837]    1 *YLSIIILPL*GS*VAGF*GRKVGVS*A-Q*-I*CLSVIITTG*AI*AF*EVN*IPVT*N****ID-*EV*NILWG*QF*SLTVAMLIPV*I*SSL*HI*>
           AA         A    A    AA  A  A   A       A  A AA A A  AAA              A A         A    A    AAAAA

Liverwort Cp           I         F         A         W
[769]    9 W*VPL*P*L-A*ILLGG*-FF*PN*IKKRR*SSFISIM*LNIM*LSFHF*WQQITGSP**RYLSWV*YKNFVLEIGYLL*PLTSIMLV*VT*VA*M*MI*>
           A  AA A    AAAA  A  A        A    AAAA      AA    A   AAA  A  A  AA A   AA  AA A  AA

L.tarentolae            S                                           C    C                              C
[739]    1 **N***N*FFNFGFICGI*--*IGRHIL*W*SVV-LCI*LVM*-TIFSC*CLSICIYGYCFYDFLLIMLDFFIWLV**CNGF-*I*I*YLID*VF*IVF*>
           AA AAA AA A        A    AA    AA A   AA A   A            AA  A  AAA A AA   A  A   A  A A   AA AA

                                  ++++++++++++++++++
              110       120       130       140       150       160       170       180       190       200
Paramecium: TYSYFRYEPHISRLISLINAFIASNIILVNSGNLVVFFFGWELIGITSFFLINFWGERAPTFKSAFKAFSFNKFSDSAVLIALILIYANVHDLNFEAILN
N.crassa
[837]  100 SI**NSHD**NQ*FF*YLSL*TFN*****TAN*YLLN*V***GV*VC*YL*VS**FT*IAANQ*SNS**LT*RVG*CFLT*GNFVVLWTLGN*DYATVFS>
           AA   AAA AA A A    AAAAA      A  AA A AAA AAA AAAAA AA  A   A     AA AA  A  A A A   A AAAA  AA

Liverwort Cp                                I
[769]  111 SD**NF*DEGYIKFFCYLSL*T***LG**L*P**Q*YI-F***V*NC*YL**G**FT*PSAANACQ***VT*RIG*FGL*LGILGF*WITGSFD*Q-Q*->
           AA AA    AA   A A   A AAAA AA A A     AAAAAA AAAAA AA  A   AA      A  A A  AA AA A A       AAAAA A

L.tarentolae                                                Y
[739]  100 AFY*NYFDNLLA*FFHIFWW*VLC*NFFIL*YDFLTAYC****L*LF*****S**-Y*FYAL*FG****FIS*VG*--**--*L*VF*-I-TFLNNG-YC>
           A A  AA A AA A    AA  A  AA A AAA  A AAAAAA AAAAA AA  A   AA  AAAA    A  A  AA  AAAAAA A  A

                                  ++++++++++++++++++++++++++++++++++++++++++
              210       220       230       240       250       260       270       280       290       300
Paramecium: VSNLYSENKLGSTPQINSWNLISFCLLFAAFVKSAQFGFHVWLPDSNEAPVPASALIHSATLVSAGVFLINRFYPILELSLYFKLVTALVGALTALAGGL
N.crassa
[837]  200 LAP-*--IN--*--D*-A-TI*GI***IG*NA**S*V*L*****NA**G*T*V*****A**N*T***Y*L**SS*LI*Y*STVL*LCLWL**I*TVFSS*>
           A    A    A    A  AA  AAA A  AA A AAAAAA  AA A AA  A AAAAA AAAA AAAAAAA AAAA A A     A  AA AAAAA A A

Liverwort Cp
[769]  209 SKRFFELLSYNQINLVF-AT*CALF*FLGPVA****PL*I****A**G*T*I*****A**N*A**I**VA*NF*LFQNLPFVNSIISWT**I***L*AT>
           AAA  A     A  A     A  A    A AAA  AAAAA AAAAAA AA A  A AAAAA AAAA AAAAAA A AAAAAA  A  A   AAAAAA A

L.tarentolae                                                                                        C
[739]  193 *ITF*F-LTFLCVDY*FI-VF*V*L**TCG*T**T***L*I****A**G*I*V*****A****VC*II**SFVFNCFDFWF**YSLIGW-SS*ILVMNS*>
           A  AA A            A AA  AAAAAAAAA AA AAAA AAAAA  AA AA  A   A  AAA AAA  A      A         A      A

           +++++++++++++++++++
              310       320       330       340       350       360       370       380       390       400
Paramecium: SAVFQTDLKKILAYSTISHCGFLIFLCSFGNFKLVIVYLFYHGFFKAISFLCVGNLIRFSKSYQDLRRNGSFFKYLPAEFFFLVFSLLNLSGLPFFFGFY
N.crassa
[837]  291 IGL**Q*I**VI****N*QL*NNVIAIGLSSYNVALFH*IN*A*Y**LL**GA*SV*HAVADN**F*KF*GLKN***LTYSVNLIAS*S*VAF*YNT***>
           AAA AAAAAAAAAAAA A A   A A A  A  A AAAAA AA  A  AA  A   AA A AAA A A  AAA  A  AA   A A  AAA  AAA

Liverwort Cp                                                            NSNEPIV
[769]  308 I*LA*K****G****N*QL*YNNLALGI*SY*AGLFH*IT*AYS**LL**GS*SV*GHPNKS*NNIF**GLRQ*N*ITAITFL*GT*S*C*I*P*AC*W>
           AA A AAAA AAAAAAA AAAAA       A A A  A A AAA AA  A  AA       AAA A  AAA      AAA A  A AAA A A

L.tarentolae                                   CY                                    ARN
[739]  291 CVFYNF*V*RYV*F***CQIS*SN*-*CL-SLD*YVGC**F*N*Y**TL*IVL*VW*H*FFGL**V*CY--**T*FCGCILL*I*AI**SCS*W*LC***>
           A  AAAAA AAAAAA  A  AA A A   AA    A A  AA A AAAA  AA AA  AAA      AAAA    AA AA     AAAAA AAA   A AA AAA

              410       420       430       440       450       460       470       480       490       500
Paramecium: SKTLLFNISDVLYFRDAIFCNILLSCITGLFYSFNILYYSFFDSKKARKSIYAGVISEYLRSYYYSNTTNASNIAIFLLIVSSCLLCAYLINFYLLSLST
N.crassa           S  V
[837]  391 **DFILESAYGQF*SG*VYIIATIGA*FTTL**VKV**LT*LANPNGYIHF*RHF*-L*E*L*V*VSY*-GKE-E-*Y*P-KH--NSKEIN*L-PR*V*G>
           AA AAA      AA  AA AAA A A  A   AAA AAA AA  A       A  A AAA A A A   AA   A A    A A AA   A  A AA AAA

Liverwort Cp                                                                                      PK
[769]  416 **DE-ILVNSW*N*-PILGSIAFFTAGLTA**N*R-I*FLT*E-GDF*GHFFD-DVK-K*S*ISI-WGSLEF*KEQ*K*DKK*T*YE*NN*NLFP*IIL*>
           AA      A  A      A A AA A   AA A  AAA A     A  A  A   A  A A  AAA      A   A   A   A AA   A AA A AAA

L.tarentolae   CL                           L V
[739]  392 C*D**LLNLTSFF*ILEFL*VC*FFIFFTVI*NYFL*FFC**KCFCLVDTLF--LLFDFECCLV*CTFCLYNCF-VL*FF*LD-F*YVFIFSS*C*-FW->
           A AAAAA  AAA    AAA A A  A A AAAA AA  A AAAA  AA AA    A   A  A    AAAA A  AA A     A A A AA

              510       520       530       540       550       560       570
Paramecium: ATDFYLVYVKTFSFTLAPLSEAALLNYSFFYWIIAIFFVILVLFSYYQKKTTAEVSLAGFFDFFLGGFFF
N.crassa                                                           IH
[837]  485 EGG*F*S-LPLVILA*FSIF-FGFITKDI*IGLGSN**IDNS**PIHEINIDT*FAVPTL*K-L*-P*I*>
           AAA A A      A  A  AA    AA  A  AA  A    AA AA    A  A A A   AA   A

Liverwort Cp
[769]  511 IPTVFIGFIGI-L*DENKNNVD-S*S*WLTLS*NSFNYSNSEK*LEFLFNAIPS**I-A**GIL-IA*Y>
           AA AA    A A   A    A A A  A A A  A   A     AAA AA A  AA

L.tarentolae
[739]  490 -S-***Y*ASF*DIA*FTVFIIDIIKF-YILSGVIFY*FNIDCINFF-WRVFLFITN-**LF*IFTTWY*>
           AAA A  AA A A     A  A         AA A    A A AAA A  AA
```

Figure 7. Amino acid sequence alignments with ND5 encoded in *N. crassa* (31), liverwort cp (23), and *Leishmania tarentolae* mt (30) DNAs. The + marks above the *Paramecium* sequence denote conserved regions noted in ref. 28.

```
ATPase, subunit 9                             +++++      ++++++
                10        20        30        40        50        60        70
Paramecium: MLLVLAIKTLVLGLCMLPISAAALGVGILFAGYNIAVSRMPDEAETIFNGTLHGFALVETFVFNSFFFGVIVYFI
Yeast mt
[135]    1 *Q****A*YIGA*ISTIGLLG*GI*IA*V**ALING*****SIKD*V*PNAIL****S*ATGLFCLNVSFLLL*>
             A  AAAA A A AA  AAAA AAAA       AAAAA  AAAA  AAAAAA A A   AAA A
N.crassa nuclear
[119]    9 N*EVS*N*GN*SAAIGLTG*GI*I*LV**ALLNG*A***ALRGQL*SYAIL***F**AIGLFDLNVALNAK*>
            AA  A A  A AA A A AAAAAAA  A        AAAAAAAA A A  A  AA A A A
Bovine nuclear
[105]    5  *A*FIGA*AATVGVAGSGA*I*TV*GSLI*GYA***SLKQQL*SYAIL****S*ANGLFCLNVAFLIL*>
              A A    A AA  A   AA A AAA    A AAA  AAAAAAA AA A  AA A A A
Maize mt
[ 97]    1  NLEGA*SIGA*AATIALAG**V*I*NVLSSSIHS*A***SL*KQS*GYAIL****T*AIASFAPNNAFLIS*V>
              AA A  A   A A AA AAAAAA AA   AAA A AA  AAAAAA A AA AA
E.coli
[ 84]    7 D**YN-AAAVNN*-L-AA-IG**I*I***GGKFLEGAA*Q**LIPLLRTQFFIVNG**DAIPNIAVGL*LY*N>
             AA A   AAAA    AAAAAAAA   A    A AA   A  AA   AAA  A   AAA A
```

Figure 8. Amino acid sequence alignments with ATPase subunit 9 encoded in yeast mt (YstMt, ref. 32, sometimes called oli1 gene product), *N. crassa* (nuclear gene, ref. 33), bovine (nuclear gene, ref. 33), maize mt (34), and *E. coli* (lipid-binding protein, c chain, ref. 35) DNAs.

ATP9

ORGANISM	Para	YstMt	N.cras	Bovine	MzeMt	E.coli	LivCp
SIZE	75	76	81	75	74	79	81
Organelle	Mt	Mt	Nuclear	Nuclear	Mt		Cp
%ID vs Para	100	32	28	25	25	26	---
%ID vs LivCp	---	30	28	32	26	31	100
%ID vs N.cras	28	56	100	58	50	30	28
SCORE vs Para	384	135	119	105	97	73	---
SCORE vs LivCp	---	112	113	102	87	128	357
SCORE vs N.cras	119	210	375	215	194	93	113

Ribosomal protein L14

ORGANISM	Para	T.pyr	E.coli	B.stear	Myco	LivCp	TobCp
SIZE	119	119	123	122	122	122	123
%ID vs Para	100	42	36	31	36	28	29
%ID vs T.pyr	42	100	34	35	36	32	27
%ID vs E.coli	36	34	100	68	53	58	54
SCORE vs Para	622	264	196	194	163	137	119
SCORE vs T.pyr	264	600	202	206	196	177	145
SCORE vs E.coli	196	202	581	419	326	380	368

Figure 9. ATPase subunit 9 and ribosomal protein L14 similarity comparisons. Selected paired sequence alignments are summarized as described in the Fig. 4 legend. The proteins that are not shown in the alignments in Figs. 8 and 10 are ATPase9 encoded in liverwort cp DNA (LivCp, *atpH* gene product ref. 23), and L14 encoded in *Bacillus stearothermophilus* (B.stear, ref. 36), *Mycoplasma capricolum* (Myco, ref. 37), and tobacco cp (*atpH* ref. 17) DNAs. Other abbreviations are: Para, Paramecium; YstMt, yeast mt DNA (32); N.cras, *N. crassa* (33); MzeMt, maize mt DNA (34); T.pyr, *Tetrahymena pyriformis* (T.pyr, ref. 38).

sequences were published. As shown in the alignments in Fig. 5 and the comparisons in Fig. 4, the cp genes have the highest degree of similarity compared to the *Paramecium* gene (score 198, Fig. 4) which aided in the identification. Further confirmation is provided by the appropriate size of the *Paramecium* gene product, the two regions of conserved identities noted in Fig. 5, and the number of conserved amino acid replacements noted in the figure. The first block of conserved amino acids noted in Fig. 5 is within a region that is absolutely conserved in a previously published comparison (25) of the ND3 protein from six animal species. The similarity scores of the genes from fungal and animal sources, relative to the *Paramecium* protein, are only slightly lower than those of the cp genes and the difference is probably not significant. When the sequences are scored relative to the mouse gene product, a close similarity among the animal mt proteins is noted (>50% identity, Fig. 4). Relative to the cp sequence, another cp gene product has high similarity (72% identity) and the animal mt genes show much less similarity (28−30%), as expected. The *Paramecium* gene product has the lowest similarity scores in any of the comparisons (21−28% identity).

The alignments of the ND4 protein, Fig. 6, show few long stretches of amino acid identities with the *Paramecium* sequence, but the overall similarities over the long polypeptide shown in Fig. 4 are clear. One moderately conserved region, indicated in Fig. 6, contains significant similarity with the ND4 protein from a protozoan flagellate, *L. tarentolae*. The similarity scores shown in Fig. 4 include comparisons with another closely related flagellate, *T. brucei*, whose genes are highly similar to those of *L. tarentolae*. The scores, relative to *Paramecium*, are higher for the liverwort cp gene than for the human or starfish gene, probably reflecting only a high degree of divergence of the Paramecium gene. Similarities relative to the human ND4 show >40% identity with that encoded in other animal mt DNAs. The human mt ND4 has a greater similarity score compared to the liverwort cp gene product than compared to that encoded in the protozoan mt DNA.

There are three highly conserved domains shown in the *Paramecium* ND5 alignments, Fig. 7, which have been noted in a previous comparison (28) of the gene product from primrose mt DNA (*Oenothera*), *N. crassa*, and mouse. These domains contain long stretches of identities aligned with the *Paramecium* sequence, but overall (Fig. 4), *Paramecium* ND5 has only 25−29% identity in any of the comparisons ranging from the mt gene from the protozoan, *T. brucei*, to the cp gene from

liverwort. Even the primrose mt protein sequence has ca. 40% identity in an alignment with the cp ND5 polypeptide and 49% identity with the mt ND5 from *N crassa*.

ATPase, Subunit 9

Identification of the ATPase subunit 9 gene (EC 3.6.1.34) in *Paramecium* mt DNA is interesting because this gene is found in different organelles in different organisms. The mt proton translocating complex (ATPase) contains a hydrophobic membrane component, denoted F_0, and an F1 component consisting of relatively hydrophilic subunits. Most of the subunits are encoded by nuclear genes and imported into the mitochondrion from the cytoplasm. Subunit 9, of the F_0 component, is encoded in the nucleus of *N. crassa* and cow (33), but is mt DNA encoded in yeast (32), maize (34), and *Paramecium*. Alignments with the *Paramecium* gene, Fig. 8, exhibit two regions that are highly conserved regardless of the organism or organelle in the comparison. The scores given in Fig. 9 show that the *Paramecium* gene product has the greatest similarity with the mt DNA encoded yeast polypeptide (32% identity), but also has comparable similarity with the nuclear encoded *N. crassa* protein (28% identity). Although the NADH dehydrogenase subunits encoded in cp DNA are as similar to the *Paramecium* mt DNA encoded proteins as the animal or fungal NADH dehydogenase subunits are, the same is not true for the ATPase 9 protein sequence. The FASTP simililarity program does not yield a meaningful alignment or score in comparisons between the *Paramecium* mt DNA and liverwort cp DNA encoded gene products, although all other sequences tested could be aligned with the cp protein (*atpH* gene product, ATPase subunit III, refs. 17 and 23). Relative to the *N. crassa* gene product, the similarity scores are about the same in comparisons with the mt DNA encoded yeast gene product (56% identity), the nuclear DNA encoded bovine gene (58% identity), and even the mt DNA encoded gene in maize (50%). These similarities suggest a common origin for these genes which are now encoded in different organelles. There is a significantly lower similarity score in the comparison with the *Paramecium* gene product.
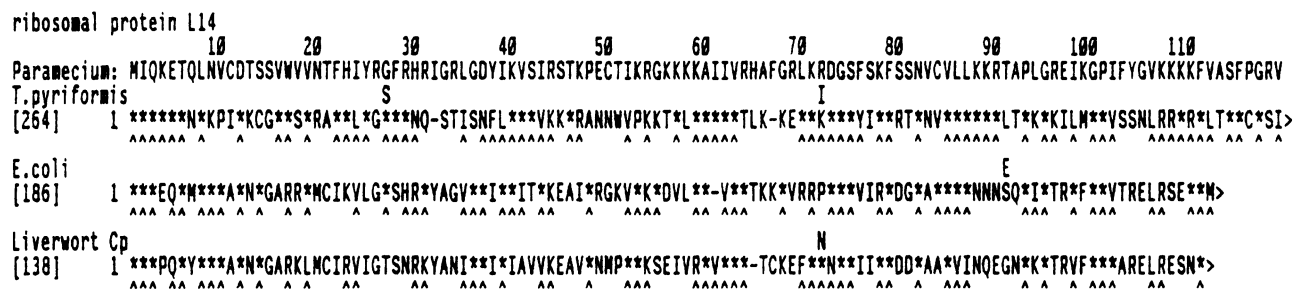
```
ribosomal protein L14
              10        20        30        40        50        60        70        80        90       100       110
Paramecium: MIQKETQLNVCDTSSVWVVNTFHIYRGFRHRIGRLGDYIKVSIRSTKPECTIKRGKKKKAIIVRHAFGRLKRDGSFSKFSSNVCVLLKKRTAPLGREIKGPIFYGVKKKKFVASFPGRV
T.pyriformis                          S                                                I
[264]     1 ******N*KPI*KCG**S*RA**L*G***NQ-STISNFL***VKK*RANNWVPKKT*L*****TLK-KE**K***YI**RT*NV******LT*K*KILM**VSSNLRR*R*LT**C*SI>
           ^^^^^^ ^  ^   ^^ ^ ^^^^ ^^^^   ^ ^^^^^^^^^ ^^   ^ ^ ^ ^^^^^    ^^^^^^^^ ^^ ^  ^^^^^^ ^ ^ ^ ^^^   ^^^^^^^ ^^^ ^ ^

E.coli                                                                                              E
[186]     1 ***EQ*M***A*N*GARR*MCIKVLG*SHR*YAGV**I**IT*KEAI*RGKV*K*DVL**-V**TKK*VRRP***VIR*DG*A****NNNSQ*I*TR*F**VTRELRSE**M>
           ^^^ ^^ ^^^ ^ ^  ^ ^      ^  ^ ^^^   ^^^ ^^^ ^^ ^    ^^^^   ^^ ^^^   ^ ^ ^^^ ^^ ^ ^^^^     ^^^ ^ ^^^   ^^ ^^^

Liverwort Cp                                                                          N
[138]     1 ***PQ*Y***A*N*GARKLMCIRVIGTSNRKYANI**I*IAVVKEAV*NMP**KSEIVR*V***-TCKEF**N**II**DD*AA*VINQEGN*K*TRVF***ARELRESN*>
           ^^^ ^^ ^^^ ^ ^  ^ ^       ^^     ^^  ^ ^^^ ^ ^^  ^^^ ^  ^^ ^    ^^^^^^  ^^^^^^ ^^ ^ ^^^  ^ ^ ^ ^^^  ^^  ^
```

Figure 10. Amino acid sequence alignments with ribosomal protein L14 encoded in *T. pyriformis* mt (38), *E. coli* (39), and liverwort cp (23) DNAs.

Ribosomal Protein L14

In the ribosomal L14 alignments shown in Fig. 10 and the comparisons provided in Fig. 9, the high degree of similarity (42% identity) of gene products from two ciliates, *Paramecium* and *Tetrahymena*, is seen. The fact that both genomes even have the gene is significant since such ribosomal proteins, corresponding to ones found in prokaryotes, have been previously identified only in plant mt and cp DNA. The gene's location in the linear mt DNA of *Paramecium* and *Tetrahymena* is also approximately equivalent.

The prokaryote L14 proteins have 31−36% identity with the *Paramecium* protein, and the plant cp gene products have 28−29%. It can be seen that similarities scored relative to L14 from *Tetrahymena* are almost identical to those relative to *Paramecium*'s ribosomal protein. When comparisons are made relative to *E. coli*'s L14, we see a high degree of similarity with the prokaryote and with the cp gene products, 53−68% identity, and a lower degree of similarity with the two protozoan mt proteins, 34 and 36% identity. Therefore, even though the ciliate gene products have a relatively high % identity compared to the plant cp and prokaryote proteins, the protozoan polypeptides are still, not surprisingly, the most divergent of those known. A similar situation exits with the 'chloroplast-like' genes that have been found in the *Paramecium* mt genome (2). The *Paramecium* psbg gene product has 44−48% identity compared to corresponding prokaryote and cp gene products. Genes encoding ribosomal proteins L2, S12, and S14 have also been previously identified in the *Paramecium* genome (2). Comparisons of these *Paramecium* gene products with corresponding proteins encoded in prokaryote and cp DNA show approximately 30% identity for L2, 30% for S14, and 40−45% for S12.

CONCLUSIONS

In all paired comparisons of these proteins, the *Paramecium* gene products are shown to be among the most divergent. Proteins from the protozoan kinetoplastids *Leishmania* and *Trypanosoma* are also very divergent from other known polypeptides including those of *Paramecium*. There is probably no significant difference between most scores of sequence alignments with any one *Paramecium* gene product and the corresponding protein from any of the organisms included in this study. The scores are low and the degree of error is significant compared to score differences between comparisons involving different organisms. However, proteins encoded in the mt DNA of a ciliate, *Tetrahymena*, show significantly greater similarities with the homologous *Paramecium* proteins (Fig. 10). Also, some comparisons seem to show unusual similarities between the

Paramecium mt and plant cp genomes (2), ribosomal proteins for example (Fig. 9). But, other *Paramecium* mt gene products, such as ATPase 9, show little or no similarity with the corresponding cp DNA encoded proteins (Fig. 9). From these results, it appears that the genes encoded in the *Paramecium* mt genome are equally divergent from those encoded in the DNAs from a number of different kingdoms.

ACKNOWLEDGMENTS

REFERENCES

1. Pritchard, A.E., Seilhamer, J.J., and Cummings, D.J. (1986) Gene 44, 243−253.
2. Pritchard, A.E., Venuti, S.E., Ghalambor, M.A., Sable, C.L., and Cummings, D.J. (1989) Gene 78, 121−134.
3. Mahalingam, R., Seilhamer, J.J., Pritchard, A.E., and Cummings, D.J. (1986) Gene 49, 129−138.
4. Fearnley, I.M., Runswick, M.J., and Walker, J.E. (1989) EMBO J. 8, 665−672.
5. Pritchard, A.E., Seilhamer, J.J., Mahalingam, M.A., Sable, C.L., Venuti, S.E., and Cummings, D.J. (1989) Nucl. Acids Res. 18, 173−180. (accompanying paper)
6. Lipman, D.J. and Pearson, W.R. (1985) Science 227, 1435−1441.
7. Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444−2448.
8. Lazowska, J., Jacq, C., and Slonimski, P.P. (1981) Cell 27, 12−14.
9. Gabellini, N., Sebald, W. (1986) Eur. J. Biochem. 154, 569−579.
10. 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.ASanger, F., Schreier, P.H., Smith, A.J.H., Staden, R., and Young, I.G. (1981) Nature 290, 457−465.
11. Benne, R., De Vries, B.F., Van den Burg, J., and Klaver, B. (1983) Nucl. Acids Res. 11, 6925−6941.
12. Dawson, A.J., Jones, V.P., and Leaver, C.J. (1984) EMBO J. 3, 2107−2113.
13. Waring, R.B., Davies, R.W., Lee, S., Grisi, E., Berks, M.M., and Scazzocchio, C. (1981) Cell 27, 4−11.
14. Citterich, M.H., Morelli, G., and Macino, G. (1983) EMBO J. 2, 1235−1242.
15. Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W., and Clayton, D.A. (1981) Cell 26, 167−180.
16. Clary, D.O., Wolstenholme, D.R. (1985) J. Mol. Evol. 22, 252−271.
17. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguc Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., De H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A Tohdoh,

N., Shimada, H., and Sugiura, M. (1986) EMBO J. 5, 2043–2049.

18. Howell, N. and Gilbert, K. (1988) J. Mol. Biol. 203, 607–618.
19. Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G.J., and Woese, C.R. (1985) Proc. Natl. Acad. Sci. USA 82, 4443–4447.
20. Margulis, L. (1970) Origins of Eukaryotic Cells. Yale University Press, New Haven.
21. Burger, G. and Werner, S. (1985) J. Mol. Biol. 186, 231–242.
22. Anderson, S., de Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F., and Young, I.G. (1982) J. Mol. Biol. 156, 683–717.
23. 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 Ozaki, H. (1986) Nature 322, 572–574.
24. Brown, T.A., Davies, R.W., Ray, J.A., Waring, R.B., and Scazzocchio, C. (1983) EMBO J. 3, 427–435.
25. Himeno, H., Masaki, H., Kawai, T., Ohta, T., Kumagai, I., Miura, K., and Watanabe, K. (1987) Gene 56, 219–230.
26. Hensgens, L.A.M., Brakenhoff, J., De, Vries, B.F., Sloof, P., Tromp, M.C., Van, Boom, J.H., Benne, R. (1984) Nucl. Acids Res. 12, 7327–7344.
27. Lazarus, C.M., and Kuntzel, H. (1984) submitted to the NBRF-PIR Protein Sequence Database.
28. Wissinger, B., Hiesel R., Schuster, W., and Brennicke A. (1988) Mol. Gen. Genet. 212, 56–65.
29. Cummings, D.J., and Domenico, J.M. (1988) J. Mol. Biol. 204, 815–839.
30. de la Cruz, V.F., Neckelmann, N. and Simpson, L. (1984) J. Biol. Chem. 259, 15136–15147.
31. Nelson, M. and Macino G. (1987) Mol. Gen. Genet. 206, 307–317.
32. Macino, G. and Tzagoloff, A. (1979) J. Biol. Chem. 254, 4617–4623.
33. Sebald, W., Hoppe, J. and Wachter, E. (1979) In Quagliariello, E., et al.(eds.). Function and Molecular Aspects of Biomembrane Transport. Elsevier/North-Holland, Amsterdam, pp. 63–74.
34. Dewey, R.E., Schuster, A.M., Levings, C.S., III, and Timothy, D.H. (1985) Proc. Nat. Acad. Sci. USA 82, 1015–1019.
35. Jans, D.A., Fimmel, A.L., Langman, L., James, L.B., Downie, J.A., Senior, A.E., Ash, G.R., Gibson, F., and Cox, G.B. (1983) Biochem. J. 211, 717–726.
36. Kimura, M., Kimura, J., and Ashman, K. (1985) Eur. J. Biochem. 150, 491–497.
37. Ohkubo, S., Muto, A., Kawauchi, Y., Yamao, F., and Osawa, S., (1987) Mol. Gen. Genet. 210, 314–322.
38. Suyama, Y. and Jenney, F. (1989) Nucl. Acids Res. 17, 803.
39. Ceretti, D.P., Dean, D., Davis, G.R., Bedwell, D.M., and Nomura, M. (1983) Nucl. Acids Res. 11, 2599–2616.