Nucleic Acids Research, Vol. 18, No. 1

Nucleotide sequence of the mitochondrial genome of *Paramecium*

A.E.Pritchard, J.J.Seilhamer, R.Mahalingam, 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

Received August 16, 1989; Revised and Accepted November 15, 1989 EMBL accession no. X15917

ABSTRACT

The nucleotide sequence for 40,469 bp of the linear *Paramecium aurelia* mitochondrial (mt) genome is presented with the locations of the known genes, presumed ORFs, and their transcripts. Many of the genes commonly encoded in mt DNA of other organisms have been identified in the *Paramecium* mt genome but several unusual genes have been found. Ribosomal protein genes *rps14*, *rps12*, and *rpl2* are clustered in a region that also contains two other genes usually found in chloroplasts, but *rpl14* is over 16 kbp away. The ATP synthase gene, *atp9*, is encoded in this mt genome, but the *atp6*, *atp8*, and COIII genes have not been identified. All of the identified genes are transcribed. Many mono- and poly- cistronic transcripts have been detected which cover most of the genome, including large regions where genes have yet to be identified. Based on sequence comparisons with known tRNAs, only those for phe, trp, and tyr are encoded in *Paramecium* mt DNA.

INTRODUCTION

With the increasing number and variety of organelle genomes whose complete sequences are known, it is becoming possible to catalogue the occurrences of certain genes in different cellular locations. This gene distribution pattern—the assignment of particular genes to specific organelles in various organisms—provides a functional genealogy that can supplement phylogenetic trees based on rRNA sequence data. This information will be important in answering questions on the evolution of organelles, such as the polyphyletic vs. monophyletic origins of mitochondria (1, 2).

Complete genome sequences available for mitochondria include numerous representatives of vertebrates such as human (3), mouse (4), cow (5), and frog (6), as well as representatives of invertebrates such as *Drosophila* (7) and sea urchin (8). Much of the sequence of fungal mt genomes, *Podospora* (9), yeast (10), *Neurospora* (11), and *Aspergillus* (12), have also been published. Protozoan data are relatively lacking with the exceptions of the sequences for much of the maxicircle DNA of the mt genome in kinetoplastids *Leishmania* and *Trypanosoma* (13). In addition to *Paramecium* mt DNA sequences, the protozoan ciliates are also represented by partial sequences from *Tetrahymena* (14–18). These protozoan genomes are particularly interesting because they are so divergent from all other known genomes, reflecting an early branching and a rapid evolution of this DNA at both the primary sequence level and the level of genome content and organization.

Approximately 40% of these *Paramecium* mt sequences have been published previously (19–25). Many of the genes commonly encoded in mt DNA, including those for NADH dehydrogenase subunits 1–5, are newly identified in this and the accompanying communication (26). Other genes, not usually found encoded in mt genomes, are also located in *Paramecium* mt DNA such as those for an additional NADH dehydrogenase subunit and ribosomal proteins (22). These genes, and presumably other open reading frames (ORFs) in *Paramecium* mt DNA, may be found in nuclear, mt or plastid genomes of different organisms.

MATERIALS AND METHODS

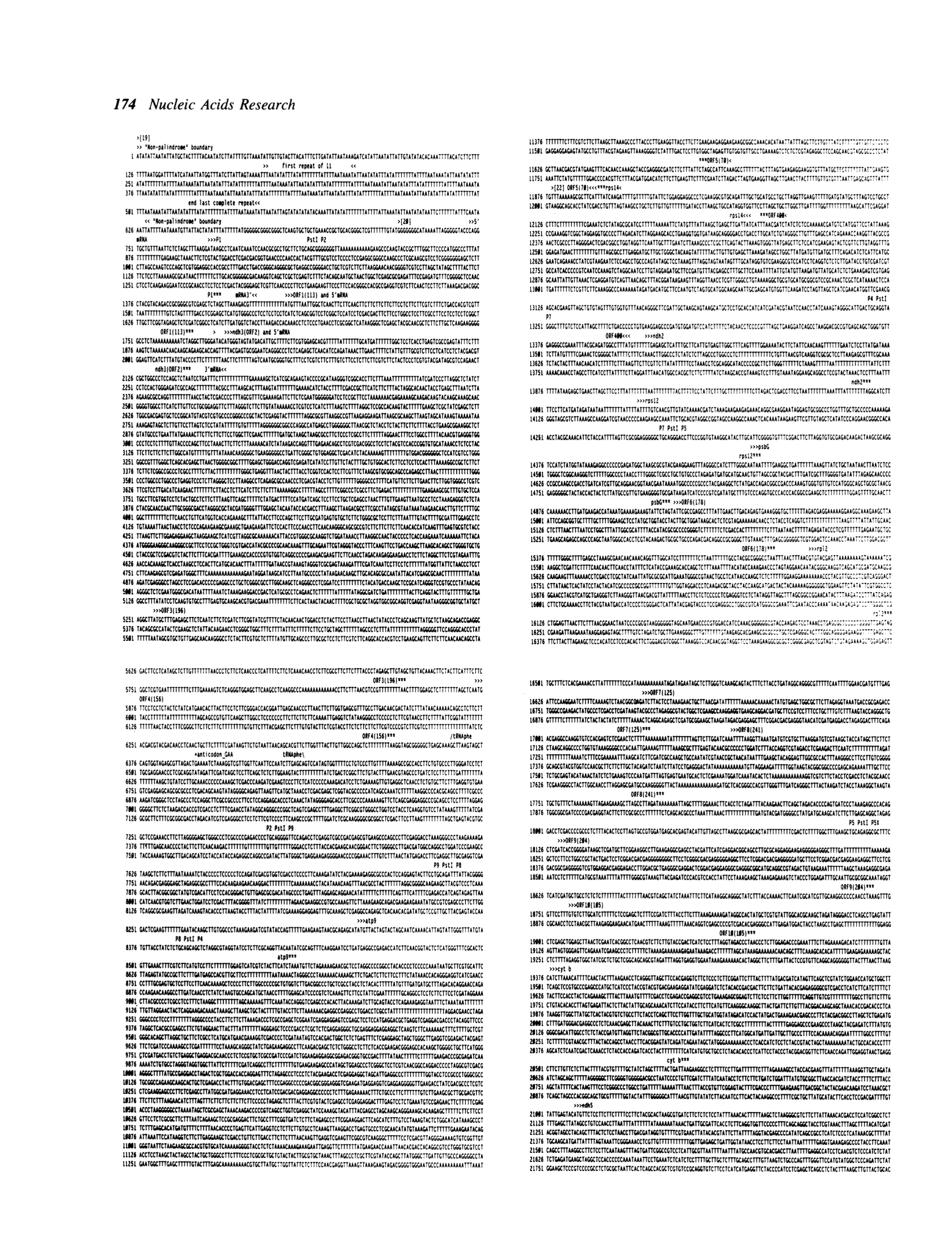
Nucleotide sequencing

The subclones of the species 4 stock 51 *Paramecium aurelia* mt DNA that were used in this project have been described previously (27). The sequencing strategies, using both the methods of Maxam and Gilbert (28) and of Sanger et al. (29), for the approximately 16.5 kbp of previously published sequences have been discussed elsewhere (19–25). The remaining DNA was sequenced in a similar manner using subclones of the *Pst*I and *Eco*RI fragments shown in Figs. 1 and 2. These major fragments are numbered by size following a previous convention (27) except for P5X which is about the same size as P5. The restriction enzyme sites at the boundaries of these major fragments were always sequenced across, usually with *Hind*III clones that have been described elsewhere (27). In most but not all regions, sequences were obtained for both strands. The order and contiguity of all sequenced subfragments were determined by overlapping sequences.

Transcript analysis

The transcript map shown in Fig. 2 was contructed from data obtained by hybridization of 32P labelled DNA fragments to Northern blots of total mt RNA (obtained from axenically grown cells) electrophoresed on methyl mercury agarose gels. In a few

* To whom correspondence should be addressed



P5X PstI P6
21876 TCGTAGGAGCCCTAACAGCTCTCGCTGGCGGCCTCTCTGCAGTGTTCCAAACAGACCTAAAAAAAATACTCGCCTACTCTACTATAAGCCATTGTGGATTCCTAATCTTCTTGTGCAGCTTTGGA
22001 AATTTTAAGCTCGTCATAGTGTATCTTTTCGTGCATGGATTTTTTAAGGCTATTTCGTTCCTCTGCGTCGGAAACTTAATACGCTTTTCAAAGAGCTATCAAGACTTGAGGAGGATGGGATCTTT
22126 CTTTAAGTATCTTCCAGCTGAATTTTTTTTTTTAGTCTTCTCATTACTTAACCTCAGCGGGCTCCCCTTCTTTTTTGGATTCTATTCGAAGACACTCTTGTTCATGATTTCAGACGTGCTTTATT
22251 TTAGAGATGCTATCTTTTGCATGATTTTGCTAAGCTGCATCACAGGCCTTTTCTATTCTTTCAACATACTTTACTATTCTTTTTTTGATTCCAAAAAGGCAAGGAAGAGCATATACGCTGGCGTC
22376 ATTAGCGAGTACCTTAGGTCTTACTACTATAGCAACACCACGATGGCCTCCAACATTGCTATCTTTTTATTGATAGTCTCGTCCTGCCTGCTCTGCGCCTACCTCATAAACTTCTACCTCCTCAG
22501 CCTCTCAACTGCAACTGATTTTTATCTCGTCTACGTTAAGACCTTCTCTTTCACCCTCGCCCCCTTGAGCGAGGCCGCCCTCTTAAACTATTCTTTTTTTTATTGGATCATAGCAATCTTCTTCG
ndh5***
22626 TGATCTTGGTGCTCTTTTCTTACTATCAAAAAAAAACAACAGCAGAAGTAAGTCTAGCAGGGTTTTTTGATTTTTTTCTTGGAGGGTTTTTTTTTTAATTTTACGCCCCTTCCTACTTCTTTTTT
>>>ORF11(367)
22751 GTGAGGTTGTATTTCCACAGATTTTGTATTTACCCCCAACTTTAGCACAGGCTTCTCTTGTTTGAAGTCCGACGTCGTCATCCACATTGCTCAATGGCAGTATTGGTGATGATTTTGGTTTACAT
22876 ACTTGTGATCCCTGTATTTTTTTTTCATTAGCAGGTCCATCAGGGCTAGAACCTTTAAAATGCGCCCAAAAATCTATACTAGTTTTAGGTCTCACGGAAAGTGGGGTGACTTTTTAGCTTGCATC
23001 GTGCCCGTTATTTGATGCTTTAACATTCTCGTAAATTCTAATTTTATCTTAAGGCTCATGGAGTGGCAAAACGAGTCCACCATCTTTACCAGTTGCATTCGTGCTAGGCAGTGATACTGGATTTA
23126 TAAGTTTGAGCTGAAGAACATCCTCGACCTGTTGACGGTCCCAAAGAAGGTTGGTTGAAACAAGTGGCTCATCCATACGGGAAACTCAATCGAGGTTGCTGATGACTACTTCTATGCTTTAAGGA
23251 TTAGGGCTCAAAATAGCTGAACCTCAAAGTACTGAAAAACCTTCGTGAGGAGCTTAAAGAAGTTTAAAGTTAGCAATAACATCCACTTTATTGATGACGTAGTGACTGCTAAGAAGTTTAAGGCC
23376 TCCGTCCTAAGCTTCATGCCCTACAACCAGTTTGACGACAAGACCCTAGGCTTTAACTTAGCTAAAACAAACGACTCTACTGTACTTCTAGAGAACGACATCTTCTGTGATAACTTCTTATTTAA
23501 GGGCGGCGATTCTTGATTGTGCGAAAGCATCGTCTTCAATCACTTTGGTAAGAAGAAGATTGGGGGCAAGAAGAGCGACACCTTGTTCAATGCCCTAAATAAGACTATGTTTTTAAACAGAAAAT
>[21]
23626 TTTTTAAAAACTTCTTATACGACATCGATGCCCTCGTGAGCAATGCCTTCTTAGAAAAAAAACCTAAGTACGTAAATCTTACTACGAAGAGGACGGACTACGGCGACTTTTCTAGGTTCACTAAG
ORF11(367) no stop but overlap >>>COII
23751 AAGCGGGTGTTTGAGAAGAGGCCAATTTTAATTACCAAAGCCTTCTTCCCCCTAAACATGGAAGCTTTTGATAATGAAATCCTCAACGTCTCTTTCGATGCAACCGTTACGGTGGAGAAGAAGAT
23876 GAACACCGACCTCTTCTACTTAACCCTGAAGCAAAAAAGATACAAAAGAAAAAAAAACATCCCCTTAAGAATAAGGCTAGACAAGAATGACTTGAGCGCAAATACTGCCTCCATCAAGTTTACGG
24001 ACAAGCCTTATCTGGTTGCTAACAAAATCATAAAGAGCCTGGAGTTTAATGCTACGAGCCTTTACAGGTGCATTAAAAAAAATAAGCTTAGAAGTGAAAATTTTTCAGTTCAGCTTTCTAGAAGA
24126 CTCTTGAGAACTAAAAAAACTCTAGTGCTCCCTTCCCACGTAAACATCACCCTTATTTCAAACTCTTACGACGTGATCCATTCTTGGTTCATTCCTGCCCTAGGAATAAAGATTGACTGCGTTCC
24251 TGGAAGGGCCACACACCATACTTTCTACTGCGATAGCGTAGGCTTTTACTATGGTCAGTGTGCTGAAATTTGTGGACGATACCATCACCACATGCCCATTAAGCTCTGCATCCTCCCCTTCGAGC
>[23] COII*** >>Small sub-
24376 ACTTCTTAATTTGATGACAGCACTTTGGCCTGCCCAAGCTCTTGTTTACAGAATCTAAGAATAGGTTTGAGACGGACTACGGACTTAAAAAATTTTGCTGGTAGGCCAGCTCTTTAGTTGCAGCC
unit rRNA 5' E1 EcoRI E6
24501 AGGCTGCATACTTGGATGTAAGAATTCTTTTTTAGGAGTGTGATCTTAGCTTTGATTTAACGCTAATCAAATGCATTACACACGCAAGTTTTATTTTACACAAGAACCTTCTTCTTTGCTTAAAA
24626 TGCAAAGATCTAGTTTTAAAATAGCGTATTGGTGCGTAAAATATGTCTTTTTATTCGTATTGTACCTTAGATAGGCTAGGTTTTCTTCCTTTTAAGAAGTTTAATGCCACCCAAAAAAAATTTTT
E6 EcoRI E2
24751 TAAAAAAATCGATACTTAAGTATTGAACTCTTTTTTTGATCCAATAGGAAAAAAGATATAGAGCGCTTCGAATTCGAAGGGGGGCCACTTGGGCCTCCTGCGCTTAGCTGATTTGTGAGAAGAAT
P6 PstI
24876 CAGCCACATACAGGTATGTGAAAGACCTTATCTCGGACAGAAGTGGGGAATTTTGGGCAATGCGCTTAAGCGTGACCCAGCAAATTGATGTTGCGAAGAAAATAAGTGAGGTGCCTGCAGCAATG
25001 TGGGATATGCCTCTTTGACTTTTTTTAAACGCAGGCCAGAAACGAGTATAATTAAGCTAAGTTTTGTTAGAAGCGATGGCCAATACATGTGCCAGCAGCCGCGGTAATACATGAGTAGCTAGCGT
25126 TAATCGTCCTTATTGAGCGTACAGAGTGAGTCGGCGGTCATAACAATTAATAACTATTTTTATATAAAAAGATATTAATAAAAAGGACTAGTGATTTTTTTTTATGTAAGTTGAGCTTGTAAACT
25251 AGTGATAAAATACAGGGAGTTTACAGGAACATTCGACGCGAAAGCGACTAACAATTAAAAAATGACGCTATATCACGAAAGTATAGGTAATGAAAAGGCATTTTAAACCTAGTAGTCTATACTGT
25376 AAAAGATGAGTATTTAATACTAAGTTAACACAAAAATACTCCGCTTGGGTAGTAAAGCCGCAAGTCAGAAACTTAACAGAATTGGCGGGGACTTGTTCAAACGGTGGAGCATGTGGTTTAATGCG
25501 ATAATCCACGTAAAACCTTAGCAGCGTTAAATTTTTTTCTTTAGTTAAGCTTCATGGCTAAAATTTTTAGAAAAAAAGAGATGGTATTGCATGGCTGTCGTCAGTTCGTGTTTTGAAATTTAGAA
25626 TTAAGTTTTATAAACGAACGCAACCCTTGTTTTGTATGTTTAATACAAAATAAATGATTAAGGAATTATTCATTTTAGCTAGAAGGCTGAAGTCAAAGTCCTTACGGTCTGCGTACGCTGGGCTA
25751 CACACGTGTTACAATGGTAAAAACAATAGAAAAACAAAGCTCTAACCTTTAGTTTTTAAAAAATTACCACAGTACGAATTGTTTTCTGAAACTTGGAAGCATGAAGAAGAAATCGTTAGTAATCG
small rRNA discontinuity
>> <<
25876 CAAATAAGTATGTGCGGTGAAATATAAGTCAAGTCCTGCACACACTGCCCATCACGCTCAAAGAGTTTTAAATCTAGAAGCTGAATTTAGGATGTGACCCTACCCATTTTTTTTTTTAAAAAAAC
26001 TCGCCGGTGCACCAAACACCGGTGTTCTAAACAAGCTACTTGTTGTCAGCACAAAGAAGAGTAAAGGCGAAACCATTTGAAATGATCGGAGTGAAGTTGACACAAGGTACCGGTATGGGAACTTG
Small subunit rRNA 3'<<
26126 CCGGTGGAGCAAGGTGAACAGACCAATGCATGCACCCAGAACGGGGCGCGGTTAGAAATGTTTTGAGACTTTTTAAATTTCTTAAACGAATTTGTCGTTGAACATTTTTTGTATTTTTGATTTAT
26251 CTTTGTTTTTTTTAATTTTTTTAATTTTTTGCTAAATGCCGTCTTGAACGAGTTCATGTTTTTAATTCATTTTTTTAACAAGAACCAGATCAAGACGTCTAACTATTTTTATTATTTAGATTTCT
>>>ORF12(265)
26376 GATCCATTTTTTCTAAAAAATTTTTGTTTAACCTCTTCTTCTAATGGTGATTTTTTTTTTTATATGAAGTCCCTTTTTTATTTTTTTGAAAGGATAAATTTTTTGACTGCCGATTTTTTATTAAAG
26501 AAGATTAGGATCGGTGCCGATAACTTAAAAGACGTCAGCTCCCTTGATCTTTTTTTCGTACATATGCACCACAGGGGCGTGGCTGTTCACAACATCTTTTTATTTTTGTTAACCTACGTTCCTTT
26626 CTTTAAGAGAGCGGTCCTTGGGTGCTACTTAAAAAAAAAGCTCTTCATCGACCATGACCTAAAGAATAAAAACTATTTTAAGATGAAGCGTGCCAATGCTCTTTTTTTTAAAGGTGACTTCCTCA
26751 CCTCGAATAACTTTTTAAGGGCCAGCACTACTTTTTATAATTCTTTTTTTTTTTCCTAATAAGGAACCACGTCATAACTAAGACAACGCTCAATGCTATTTATTTTAGTAAGGGCTCTTTTAAA
26876 CCCGCCGAGACCAGCTACTATTTTAATGCCTACGTCCTTCATTTTTTTAAAAAAAACAAGGTTAACGACTTCATCGTCTTCTTAAATGAGAACAAAATAAAGGACTATTGGGCGTACCTAACGAC

>>>ORF13(169) overlap ORF12, different frame
27001 CATATATGCAAGCCCCGAGGCCCAAAACTTAAAACTTATCGACCCCGTAGGGTCCTTGATTGAACAGTCAAACTCCTATGCCGACAGGCTGCCGACGAGAAGGACTTTAAGAAAACCAGCACCTA
ORF12(265)***ORF13 continues
27126 CTCAAAGCTCGATGTCGAGGCTATGCATTCTTCTTTTTTCTTTAACAACGCCCCCATTGAAGAAGCTGAGGCCTCCTCTCAGCCTGAGGCTCCATTCCTGCTACCCTTAAAAAAAATTTCAATTT
27251 TTTTTGCAAAAAAAAACTACCTCTATAACAAGGGTAAGTTTTCTAGGAATAAGCAGACATATAGGACGGGGGTTTATCTCTGCATCTGATTAACGGTACTGACGGTCGTGGGCCTTTATTTTTAT
27376 TTCTACCTTATGAGCATGAAGTTTACCTACAACTACCTACTCTTTTTATTCTTCCTAGGCCTTTTTTTTTATAAATTTTTTATAAAGAAAAATAACAAAAAATTTGAGGTGCATACTGATCTCTT
>[25]
ORF13(169)*** /tRNAtrp anticodon TCA tRNAtrp \
27501 TGAAAATTTTTAAGATTCGGGCACTACACTCTAGGGGAGTAGTTCAACGGAAAAACTTTAGTCTTCAAAACTGACATCGTGGGTTCGACTCCCGCCTCCCTTGTACTTTTTTTTGTTTTTTTAA
>>>ORF14(387)
27626 AATGCATAGAAGAGACCACAGACTTTCTATTTTTTAATAAGTATTCACGAGAATTTTCATACCACTTTTTTGGTAAGTTTTTTTTTTTTAAAGAAATAAATACAAAGCTTATTTTTTTTTTTGAG
27751 AGGTTTAACCTCTTTAGTTTCTTTACCTTCTTTAGTATTTTTATCTTATCGAGGAACCTTGTCCACACAAAAAACTTTATTACAGCACTCTTTACCTACACTAAAGCAACCCTAAAGAAACAAAA
27876 AAAAAACTTCTTTATTTTTTTTAGGATCTCCTTTTTCTTTGTCGTCTTTTTATTTTTTTACCTTCTATTCTCATATACGCTAGAAAACATTCCCGTATCCAAACTCTTTTTTGTTTGAGGTAGCT
28001 GAGGTGCTTTTTTATACATCTTTATCTCTGGCTTTAATTTTTTTGGAAAAAAGTATACCTATGGAAAGTATACTGAGGCTCTGCAAAGGTTTTGGAAGAGGAGTTTTAGCATCTTTTGACTTATC
28126 GAGGGCTTTGTTTTTTCAGCCTTCATCTTCCTGACCTTCAATGCAAGCTCTGAGGTCGTATACAGCTACGACCCCCAGGCTTTTTTTAAACTCCACCTGATCTCTTTAAGATTTTTTTTTTTTAA
28251 AATGCTCGCCCTCACTTTTTTAATTCTATGCTTTAGCATGGTCTCCTCTCTAAACTTAAGGAAGCGCTTTAACATGTTTATTTTTAACACAACATCACTTTTCGTCATCATTATTTTTTTAATTG
28376 AGTCTGACCAGTATATTAGTATTGTAAACTACTGTGGCTTCTTTGAATGATCCTTTAATCACAGCGACTTTGCACTCGATTCTGACTTTAGGAAATCCCGAACTGTTAACAGCTATGTGCTCTTA
28501 ATCGGCATTGCAAAGTACCTTCACATCCTCTTTATCGTATTTGTTTGATTTTTTAATTTTGCAAAAAACCTTGAAAACAATGAGTCCAGAGACTACATCGCCGGTACCTGCACACAAAATGCCAT
28626 CATCTTGTACCTCCTAAATTGGTTTGCCATCTATCCTTACATTAAGTACTTCTTCAGAACTTATTACTATAGCACTTTTTCTTGATTTTTTTTTGACTTTAAAAATGAGTCTGCTTTTAACTTCG
ORF14(387)***
28751 CAAGGTTTTTGTTTAACTTTTGGAGCAGTGCCTTTTTAAACTAGAAATAATAAAAAAAGTAACTTTTGGCTAGTTTTTTACTCGTTAATGTGTTCTTCATGAATTTAGATTTTCTAACGGGGGTC
28876 CTTATAAAGTACCATGTACTTTTTTATTACCTTGCCGTTGGGTTGAATACCTTCTTTATACGCAGCAGTGCAACAGTGGCTTCGGTGGATTCGTAGTGGTGTTCTGGGTTGTTTGCCTAGAGCTA
>>5' mRNA >>>ORF15(85)
29001 GTCTTGGTGACTAGCCTAGACAAACTTCCTTGATTCATCTTCTAGGGATACCAGGCACTTTGTCTCTTTTCACGTCGATTTTCGTAAGTTTAAAACTTCCTACGAAAGCGGCCCTGAGCAAACTC
29126 TTTTTACAATGCAAGGCTTTGCGTGGCTTCTTTTTAAAAGAAGCCAATTACAATTTTCAGCACTCTTGACTTTACTTTATTTTTGTGTTAAACCAACTTGCTTACATGCGCGGAAATTTTTCAGT
>[20] ORF15(85) no stop but overlap with >>>COI
29251 TATTGCCCTACGAGAATTGAAATGAGAAGTTAAGCGAGCCTTTGATAGATACCGTTATTTTTTTAAAAAGCATGTTTATACCATAAACCACAAAAGGATTGCATTAAACTATTTTTACTTTAGCA
29376 TGTGGACTGGCCTTTCAGGAGCTGCCTTAGCTACCATGATCCGTCTTGAGATGGCCTATCCAGGAAGCCCTTTTTTTAAGGGGGATTCGATAAAGTATCTTCAAGTAGCAACTGCTCATGGCCTC
29501 ATCATGGTATTCTTTGTGGTTGTCCCTATCTTCTTTGGAGGGTTTGCTAACTTTTTGATACCCTACCACGTAGGCTCAAAAGATGTTGCATTCCCCAGACTAAATAGTATTGGCTTTTGGATCCA
29626 ACCTCTCGGCTTCCTTTTGGTTGCAAAAATTGCGTTTTTGAGGACGACGTCTTGGAAGTACTACGACAAGACCTCTTTTTTCTTACAGCCCTACAACAAATCTCTTTACAGGGACTTCTTTAATT
29751 TTTTGACGGGGGAGCTTAGCTTTAACCCCTTTAAAAAAAGTTTAGATGAGTCCTTGTTCCTCTTTTTATGGAAGCCCAGAAAAAAGATAACCAACACGGAGTATACATCCTTCTTCTTCAACCCC
29876 CTCAACCTAAGCTTTTTGGACTCTTTTTTTTATTATAGTGACAACCTATGATCCCTGGCTAACAAGGTTGTTTCCTCTAGGAGGAAGAAAATTTACGTTACAAAGTGTTCGAATCGAGCTGCTGT
30001 AACAGCAGGTTGGACCTTTATAACACCCTTCTCTTCTAATATGAAATATTCTGGTTTTGGGGCTCAAGACGTCCTCTCAGTTGCTGTTGTGCTAGCTGGCATTAGTACAACCATATCTTTACTAA
30126 CTCTCATAACGAGGAGGACTCTAGTTGCTCCCGGACTTAGAAATAGGAGGGTGCTAATCCCTTTTATAACAATTTCTCTACTTCTTACACTCCGCCTCCTTGCCATCGTAACCCCAATTTTGGGA
30251 GCTGCGGTGCTTATGTCTCTCATGGACAGGCATTGACAAACTTCCTTCTTTGATTTTGCTTATGGAGGGGATCCTATACTATTCCAACATCTTTTTTGATTTTTTGGACACCCCGAGGTCTACAT
30376 TCTAATAATCCCAAGCTTTGGAGTTGCAAATATCGTCTTGCCTTTCTACACCATGAGGAGGATGTCTTCAAAACATCACATGATTTGAGCTGTCTATGTTATGGCTTACATGGGTTTTGTTGTTT
30501 GAGGCCACCACATGTATCTTGTAGGGCTAGATCATAGGAGCAGAAATATCTACAGTACCATTACCATCATGATTTGTCTTCCAGCAACTATTAAGCTTGTTAATTGAACTTTGACGCTAGCTAAT
30626 GCTGCTATTCATGTAGATCTGGTATTCTTGTTCTTCTGTTCATACGTCTTCTTCTTCTTAACTGGAGGTTTTACTGGAATGTGGCTCTCTCACGTAGGATTGAATATTAGCGTTCACGATACCTT
30751 CTATGTCGTGGCACACTTCCACCTTATGCTTGCTGGAGCTGCAATGATGGGAGCCTTTACAGGCCTCTATTATTACTACAACACCTTCTTTGATGTTCAGTACTCTAAGATCTTTGGCTTCCTTC
30876 ACCTTGTTTACTACTCTGCAGGTATTTGAACGACGTTCTTCCCCATGTTCTTCTTAGGCTTTTCTGGGCTTCCCAGAAGGATTCATGACTTTCCAGCCTTCTTCCTAGGTTGGCACGGACTTGCC
31001 TCCTGTGGCCACTTCTTAACCCTTGCCGGGGTCTGCTTCTTCTTCTTCGGAATTTTTGACTCCACAAGCGAGAATAAGTCTTCGATACTAGCTAACTTTGGTATACCAAAAATCGCTAAGAGGGC
31126 CCACCTTTATTTCTTTAAGATCAGCTACAATAACTATACAAATGAAATTGCTAGCGAGCTTCCAAAGGTCGAGGTAAGGAAGTTTATCATCGAAAACACTTTTGGAGAGTACGAGTGCGTTAAGT
COI***
31251 TGGTGCCCGTCACTAAGTAAACGACTAATATTGATTTTTATCTTTTTATGGTTGATGCTTTTTTGCACCTCTTAAAATGCTACTTTATTAGCTTTCAGTACTTCACTAAGAAAAATTCTAAGCTT
>>5'mRNA
31376 ATTTGAAAGATAGTCTTCTTTTGCGTTCTTTTGAACCTCGGACTTTTTTATATTTCTGTATCTTTTTAACACTAGCCTAGGCACCCCCTTTTTGAGTGTTTTAGGTTTTGATAGCCTCGGTTCTC
31501 TTTTAGCTTATAGCACAGCCTCGATTACCTATTTGGTTGACTTAGCTCTGTACCAGGGTTACAAGCTCGCACACCACTCCCTCCGCTTTTACCTAGAGTCGAGCATGCTCTTACTTAGGAACAAA
>>>ndhl
31626 GTATTCTTCTTACCCTTTTTTAGTATTTTAATCTATTCCATCGTTTTGATGCTGGTCGTTACCTTGATAATCGCATCCATAACTCTCCTAGAGCGAAAGCTGCTTTCCTTGGTTCAGAGGCGCGT
31751 AGGGCCTAACTTTGTAGGATACAAAGGTCGTCTTCAGTACCTTGCCGATGCACTAAAACTATTTTTAAAGGGGGTTGCAATCCCTTCTGGCGCTAATTCTTTTTTTTTTGTAGCCATGCCCTCTC
E2 EcoRI E4
31876 TTGCGGGAGCGGTTTGCTATACTTTTTGAATGAATTCCATCTGGGGCCCAAGCTTAAGCATGTTCGATGTTGAGTACAATATCGTCTATGCCTCCCTCCTCTCCATCCTCTTTGGACTCTGCGTC
32001 ATGCTTACCGGCTACTTTAGTAAAAATAAGTATTCGGTGATGGCTGGCCTACGAGCAGCTATCCTCATGCTAAATCTCGAAATTTTCTTGGGCATCGTCTTTTTAAATGTATGCTTTCTCGTCGA
32126 ATCTTTTTCTTTTGCAGCCTTTGCTGTTTATCAGGAAATTTTTTGACTTATTTTCTTATTTTTTTTCCTCCTATCAAACATCCTGCTCGTTTTCTTATTGGAGGTTAACAGGACCCCCTTTGACC
32251 TTGCTGAGGCCGAGTCCGAACTCGTTACAGGGTACACTACAGAATACGGGGGCTTCTACTTTGCACTATTTTACCTAGGTGAGTACTTTCACCTCTTCTTCTTTTCGTGCCTGATCAGTGTAGTT

ndhl***
32376 TTTTTGGTTCATGGGAACTCTTAAAGCCCTTCTTGTTCTTGCATAACTATACGGTTTAGGCTAGAATAGTGATTTTTTTTTTTTTTTGTCTTCACTGGGTACCTTTTTAGATCTATGATTTCCAT
32501 GAACGAACTTTTTATCCGAGGTAAGCTTGTTGCAGGTATTGACCTTGGTAAACTTTTTTTTTTTTTTTAGCTCTAGAAATGTTTTTTACTCTTGTGTTTATTTTTTTTTAAATATTCTAACCCTT
32626 GGGGTTTTCCTTGCATTCTTTAATTTTGAATTTTTAACTGGATTCTTCTGAGTCGTTGAGTTTACTGTTTTTTTATATTCCTGGTATTTTACTTTTATTTTTCTAGCAACGGAACCATTGTATT
32751 CAATAAGAACTTCTTTCTTTTTTTTTATTTTTTTTTTATTTTTTTTCTTTGTTTTCTTTTCTTCAATTTGAACCTTTTCATCCTTTGAAAAAACTAGCCTAAACTTTTTTTTCTTATGAGACGACT
32876 ACTACGACGCCCTCAACTTTAATGTTATGAACGATCTCAACTCTTTTTTTCTAAATTTTTTTATCTTTAACTCCTTCTTCTTTTTTTTTATTTACGGTCTTAATCTTTTTAACGTCTGTGGTTTGC
33001 ATTAACCTGTACCTAAACTCTAAAAAAAATAGCTTGGGAATGGTAGCTTCTTTTTTAAGGATTTTTAATTTTTTTAAAGACTTCAGCGCGTTCTCCTTTATGAGAAAGCAAAGCTTAGTCATTCA
33126 AAACTTTAGGAAGCCCGTGAATCGCCTAGTCTCTAAAGAGCAGTTTACTCCTAAGTTTTACAAGAAGGAGGAGGACAAGAAGGAGGAGGGCCAGAAGGAGGAGGGCCAGAAGGAGGAGGGCTCCC
>>>rpl14
33251 TAAGCTAGTCTCTTACCAGACTAAGAATGATCCAGAAAGAAACACAGCTCAATGTTTGTGATACCAGCTCTGTTTGGGTGGTCAACACCTTTCATATTTATAGGGGATTTAGGCATCGTATAGGA
33376 CGCCTAGGGGACTATATCAAGGTCTCCATACGTAGTACAAAGCCAGAGTGCACGATTAAGCGTGGTAAAAAAAAAAAAGCTATAATAGTTCGACACGCCTTTGGAAGATTGAAGAGGGACGGCTC
33501 TTTTTCTAAGTTTAGTAGCAACGTTTGCGTTTTATTAAAGAAAAGAACTGCTCCCCTAGGGCGAGAGATCAAAGGCCCTATCTTTTATGGTGTAAAAAAAAAAAAATTTGTTGCTTCCTTTCCCG
rpl14***
33626 GACGGGTGTAGCAAAGCTAGAGCTATAACCTCTAGCTTTGCAAATGTTGAGTAGCTTTAGCTTTGTTGAGTCGCGAACGAGGCTTGTAAATGTACTAACTGCTAAACTCGGCTTCTCTCTGAATC
33751 TTTTTGATCTCGCAGAGAAGACGCAAGCGATAAAAAACATTTCCTTGTTTTTTTATAATATAAATAACTTGAAGGTTCAAAACATTGTTAACCTTAATTTTTCTTGAAATAGCTTTTTGAGCATT
33876 TTCATATTTTTTCTATGTTTCTTCACAGTCCTCTTTACGTTCTTATTATTATTTTTTTTTTAAGTCTAAATAATTTTTTTATGCATGTCTGGAGTACCTCAAGCTTGCCTTCCTTCCTCCATGCT
>>>ndh4
34001 CGGAACGTGCTCTTCCTAGACTTTTTTTTTACAAGTCTAGTGTACTCGTCCATCATCTCCACCCTCACGATGTTTGCAGTTTATCTTGTTTTCAGCTTTAAAAAACCCGAGGCCTCTAAAGCCGA
34126 CCTTGATTTCTTTATGCTTACCTTTAAGTATATTTTAAAGGGCTCTTTCTTTTTAATGCTTGTGCTGGCCCTTTTCTGTGCACTCTTTACCTTCGATCTTATGTTTTCTGCTAAAAATCTTCTCT
34251 ATCCAAACGAATACATATGGGACTCGGGTGACTTCTTTTTTTACAAAAATGGAGCCCTTAAGTTCTCTTTAAATTTATACGGCCTTATCCTAGTCTTTCTCTGCCTCCTTACTGGGTTTGTTGCC
34376 ATCTCAACAGTAGACAATCTCTATTCTGAAGATAAGCTCAAATTTTATTTAATTTTTTTTCAATTCTTCCTCGCTGTCCTCGGCTTTATAAAGTGTAGCGACCTCATTGCTTTTTTTTTTTTTTA
34501 TGAGGTGTTGATGCTCGGCTCAGTCCTCGTTGTTTTCTTTGGAAGCTATTCTAAAAAATCAATACATGCAGTTATCTACTTCGTAGCATGGACGCAGCTCGGATCCCTCTTCGTCTTATTAGCAT
34626 GCCTTTACATCTACAGCTTGACAAACTCTACGAACTTTTTTGTCATTAAAACTTTTGTCTTCTCAAAGACCCAAGCAATGACGATTTACTCGCTCCTCTTCGTGGGCTTTGGTATTAAGTTCCCC
34751 ATCTGACCACTTCACTATTGGCTTACCAAAACTCACGTAGAGGCCTCTACTGGCTTTTCTATATACTTGAGTGGCTTTTTAGTAAAAACAGCTCTCTTTGGATTTTATAGGTTAACAAACTTAAT
34876 TCAAGTAGAGCTGGATACAACTTTCTTTTTAGCAGTTCTTGTTGCAGGCGTTATTGACTCCTCCCTCAACATGTGGAGCCAGACCGACCTTAAAAAACTTGTAGCTTACTGTACGATCCAAGAAA
35001 TGAACCTCATTGCCATCTTCTTTTTAAAAGGGGATTCCAGCCTCATCGCCTACGGCTTCCTTTTTACCATCATGCACGCGCTCATGTCTACGCTCATGTTTTTCCTAGTTGAGTGCATCTACTCA
E4 EcoRI E5
35126 AGGTACAAATCTAGGTCCACTCTAGTTGTTAACGGTGTTTTCTTCTCGTTCAACAACCTCGCATTAGCAATTATTTTTATGGTTCTTTTTTTTTCTGGAATTCTAGGAACCCTAAAGTTTGTTTG
35251 TGAATTTTTTGTCTTCAACCTTACCCTTCACGTTTCGTGACCTATAGGCGTTATTTTTGTGGTTGTTGTGAGCGCTATCGGCTTGATTGGCTTTTCTAAAAATTGATTTAATGCTATCTTTTGTG
ndh4***
35376 CACCAAGCAAAGACGTAGGCCCCGACGCCCTCGACCTCTCAAAGAAGGAACTTTACATTATCTTTTTGTGCTTCGCAGGCTTGATCTTCTTAACCTTCCTCCCTTTCTTAATGATTTAATCTTAG
35501 GGGCTTTATATTATCTTTTTGTGCTTCGTAGGCTGGGCTTCTTAGCCCTCTCTTAATAATTTAATTTTAGTATGCTTTTCAATCCTCTTTTTATTTTTAGAGTTTTGAATTTCTTCAAGAGTGGT
35626 TTTTATATAGACTACATCATAAAGAAGGTTACTGAAGCTATTGTAAAGAATTTATTTATTTTTTCTTCATTCTTCTTTGCCGAGAAGTTTTTAATTGAGTTTTTAACTAAAAAAAGCGTAGACAA
35751 CGTAATGATAAAACTAAACTCCTTTGCACTTGAAGATTTTAACTTTAATACCTTCTTCTCTCTCTTTGTTGTTTTTTTCTTGTATTTTTTTTTTTTTTTTTAGAGATGCTTTTCTTGTTTTTTTAAT
>>>ORF16(189)
35876 TTTTAGGACGCCTAATTCTTTTTTGGAAGGTAGTTACCTGGGACACCTTAGTATTTTTACCCTTGTATGAGACTTCTCCAATATGACTTATGTCGCATTGCAAAGCTTTCTTTGCATTCGAGATG
E5 EcoRI E3
36001 AGCTCAAGGGTTTCCGAAGTCTCTACATACTTCGCAAGCAACGAATTCTTCGTTCTACATGAAATTTTCTCCAAAGCATCTTATTTCTTTTTAGTTTTCGAGCTCTTTTTTAGCAATCTTATGGA
36126 CGCTTTTTTAAACAATATCGTGGTTTTTTTTGAGTTTGATGAGCCTACTCTTAGGAGGTACATTAGCTTCATGGACCCCACCCTCTTTTCTATCTACCACCCCGAGGCCGTTTTTATTCAAAAGA
36251 GCCTTCTTACCACAAATCATCTAAGTTATTTTGCAAATTTTAAGCTTAACGTAGCTGTACTCTTGGATAAGAGCAATGCTATAGAGCCCCTAAACTATATGCTACAGGTCAGCTTCTTACTATAC
ORF16(189)***
36376 TTCATTGTTTCTTTTCAATTCTTTTTTTTTCTTTCTTTTCTAAAAACAACGAGGAGTACCAAATCGATGTTGAGCACTCTGTAATTAACCTCTCCTCTGAGGCTGAAAAAGAAATCTTTTCGGTA
>>>ORF17(221)
36501 GACTATGCCCTTAATCTTTTATTCTTCTTGTTGTTTTTTTTTGGTGTCTACTTCGGCTTTTTGTTCTTGAGTATGTTCTCGTCCTTCACAGAGCTCTTTGTGTTCTTTGTGCCTCTCTTGTTTAT
36626 TTTTTATTTTATTGTTTTGATCCCCTTCAACCTCTTATTTGATTTTGGTCTTTTTTTTGTTGTATACTTGCGTGGTTCTTCAAACACAACTAGCCTATTCTTTGAGTGCGTTTATGAACTACATC
36751 GGGGTCTTAGCATTCTTCACCAGACTTATTGTTCAGTTTCTTAGGCTTGTTCTAATGTTTGTTGTTTACTGCATGATGCACGATACCGTCATGTTACAAAACTATTCTCAAAAAAATTTCTTGGT
>[24]
36876 TGGAGATAGCTTTTGAGAGGAGTTAATGAGCGTTCAGCCAAACGGAACCTCGATCTTCTTTTTCTTGTTTGCAACCTTCCCTCTAAGGTTGTTTTACTGATGCTATGAGTGCATGCATACTTTTT
37001 TTGTAGTTACTGTACAGTTTTCAGCATTCTTTACTATTGTTTTTGATTGTTCTTACTCTTCTATACTTTTTTTGTTTACGAGAAGTATGAGCACCACTTCGACAATATAACAAAGATGCATAAA
ORF17(221)*** >>5.8S rRNA 5'
37126 AAGCTCATTGAAGAGCTTAAGTCCTTAAAGAAGGACTCCTTTTAATTTTGATTTTTTTTTTTTTTTGCAGTCTAACAAAGCACTAGACGGATGCCTAAAAATCTTGGTTGAGGGCGTAAATCTAAA
37251 TACGATATTTGTATGTAAGCTTTCTTTTGATATATAGAATTGCTTAGGATGTACAAGGGCCCTTGAAAATTTTATGAAGCGAAACATCGTAGTAATAAATCTAAAACATAATTCAACTGAGATGT
5.8S rRNA 3'<< >>Lrg
37376 TAAAGTAACGGTGAGTGAAACAAAGTAGCTCAAAAATTAGAAAAGAGGGCTGAATTACTTGGAAAAGTAAGCTACAGTGGGTGATAGCCCCGTAAGCTTTTTGATTTAATTATTTTTTTTTGTGC
subunit rRNA 5'

37501 TGCTAGGGTGGTCTCAACCTGGGTTCATTCCCCAGCTAATAAAGTTCGATTCTTTGTAGCGGCTCCGAGTATATATTATTAACAAAATAATATAGAAATGTTTTTCAAAAAAAAAAGCTAATAAA
37626 ACCAAATTTTGATAGAAAATAAGTACCGTGAGGGAAAGGTGAAAAGAAATTTGTAGAATGTTTAAAGATTCTGAAATCTAGTGCAGTGAAACAGTTAAAGCGTGTTGTTTTAACGTACCTTTTG
37751 TATAATGGGCCAACTAGTTTATAAAATTAGCGAGCTTATCAAATCGCGTAATGAAAATGTTATAGTTAATTTTATAAGACCCGAAGTCAAGTGATCTAATCATGGCTAGGTAGAAGATCGAACCC
37876 ATAAATGTTGCAAAATTTCGGGAGAAGCTGTGATTAGGGGTGAAAGGCTAATCAAACTTGACGATAGCTGGTTTTTCGCGAAATCTATCTACGTAGAGTATTTTTTTTCTTTTGTGCGGGGTAGT
38001 GTAATCTTTTCTTAAGAAGATCTGATCTTTTTCGTGAAGAATCTTCCAATACGCATTAAAATAATAAATAAAACGGACTGAGAGTGCAAAGATTCTTGGTCGAGAGGGAAACAGCCCAGACCGT
38126 ACGATAAAGTGCATAAACAATGCGAAGTAAAAGAATTGTTTTTAAAAAAATATCGGGAGGTAGGCTTAGAATCAGCCAGCCTTTAAAGAAAGCGTAACAGCTCACCAATAGCTACGATTATTTGA
38251 ACAAACATCGTATTAAAAATAATTTTTAAATTAAAGCGCTAAGCATTGTACTGAAGCGGCGGGTAACTCGGTAGCGAAACGTTTTGTAGGTTGTTGAAGGTTTATTGAGAAATAGGCTGGAGATA
38376 TCAAAATTGATAATGTTGGCATGAGTAATAGACAAAATGTTCAAATCATTTTTGTTTGATAAGTTAGGGTTGCTTTGTTTTGATCATCTTACAAAGTGTGATCCGGTCTCTAATTTTTTAAATAG
38501 ATATTTTTAAAGATGAGCTCTGTATTTTTTTCTGGCCTCCTTTGTGTTCTTTTTAATTTTTTTAGAAGAGTGCGTGGAGGCGGACTGGAAACATTTTACATGGTGTACCGTACTAACACTAACGC
38626 AAGTACTTTAGTCGAGCAGATGACGACAAAAGAGCTAATGATATTGAAGGAACTCGGCAAAATTACTTTGTAACTTCGGGATAAAAAGTGCGTCAGCAAACAAAAAAATGGGGGGTAGCGACTGT
38751 TTACTAAAAACATAAGATTTTGCAAAATTTAATTATGATGTATAAAATCTGACTCCTGCCCGGTGTTGTCATGCAAAGTTTTAACGGTTAGCGCTGTTGATAACAAGCAGCAATAAACGGCGGCC
38876 ATAACTCTGATGGTCCTAAGGTAGCAAAATCCCTTGACGGGTAAGTTCCGTCCTGCACGAATGGAGTAACGACTGCCCTACTGTCTCCAATATCAGCTCTATGAAATTGAATTTGCTGTGAAGAT
39001 GCAGCTTTTTACAACTAGACGGAAAGACCCTATGCACCTTTACTGATGCTAGGAACTAAAGGGATATACTGGAGATAAATTAAGGTAGGAGTTAAATCGCAATTGAAAAACTACTTCTCTTTTAC
39126 GTCTCTTCATAAAAATGACTAGGTTATTTTTACTTTTCTGGCTGTTAGTTTAACTGGGGCGGTTGCCTCCTAAAAAGTAACGGAGGTGAGCATAAAGTTACGCTTTGTGAGGAATTTTTCTTTGC
39251 AAAATGAGTTAATAAAACTGCGTAATTTGATTAAATTACAAACTAGTAATTTAGGGGCTAATTGCCTGCTATAATGATCCGGTGTTTTTTTTTGAATAAGACATCGCTCAACGAATAAAAGGTAC
39376 GCTAGGGATAACAGGCTTATAAATTCTGAGAGTTCCTATTAAAGAATTTGTTTGGCACCTCGATGTCGGCTCATCACATCTTGGTGGTGCAGAATCTGCCAAGGGTTTGGCTGTTCGCCAATTAA
39501 AGTGGTACGTGAGCTGGGTTTAAAACGTTGTGAGACAGTTTGGTCCCTATCTGTTGTAGAATTAAGAAACGAGGCTGAGATCGACGCCAGTACGAGAGGACCGACTAGAACTAGCCTCTGGTTCG
39626 GGATCTACTTTAAATAAAGGTATAGATGCTATGCTAAGCTGGATAAGCTAATTTTTTAATTAAAATTTTTTAGCTTAAAGAATTAGAAGCTTTTTAGCAACTCGTTTCTAATAGTATTTTAGAAG
Large subunit rRNA 3'<< /tRNAtyr
39751 ACTACTAAGTTAATTTGAAGGTAGCTGTTATAGGGCCCTTAACGGGCGTACCAGCTCCCCTTAAAAAAATACTAAGACTCTTATTATTATTATATGAAGTAATGGCTGAGTGGTTAAAGCGGCAG
anticodon GTA tRNAtyr\
39876 ACTGTAAATCTGTTGGTAGTACCGTCGTTGGTTCGAATCCAACTTACTTCATACTAAGAAATTAGATTAACTTTTGTAAACAATAAAATTACAAAAATTTTTTTTATAAAAAAAAATTTAGAACA
40001 TCTTTGTTTTTTTTTTTTAGGAACTAAATTAGCTCCTGATTTCTTAAAGCTTAGAACGCGATTGTTAAAAATCTTTAATAAGCGCATGTCTATCTTTAATTTTTTTTTTTATAAAAAGATTTGTTT
40126 CTTTTTTAATTTTTTTTTTTAAAAACATAAAAAAATCAAAACTCCTTTTTATTGCAGATAAGAAAATATACTATAAGTACGAGTACGTTAAGTACTTTAACTACCTGTTTGATAGAGAAGATTTT
40251 TATAAAAAATATAGATTGTATAGGCTATATATTCGCGAACGTATGAACATATTTATTCAGCATCCAGAGCTTTTTTTTTTTTTTTTTTACTAAAAACAAAACCTTGTTAAATCAATTAAACACCAT
E3 EcoRI
40376 GCCCGTATGCACAAATATTATTTTATTTTCAGATAATAACATGTCCTTTTCTAATTTTTTTTTTAACGTACCGCTGATCATTAGTTATGAATTC

Figure 1. Nucleotide sequence of species 4,51 *Paramecium* mt DNA starting at a 'non-palindrome' point, indicated by double brackets >> above the sequence line, of the single strand loop at the replication initiation terminus (19). Within this terminal loop there is a repeated sequence element, also bounded by double brackets, that is variable in number; 11 repeats are arbitrarily shown here. Other features are represented by symbols above the sequence line as follows: (1) >>> above the translation initiation codon indicates the direction of the gene or ORF which is named, *** is above the stop codon; (2) ORFs are numbered consecutively followed, in parentheses, by the number of amino acids that are encoded in each; (3) >> and << define the labelled 5′ and 3′ termini of the known boundaries of mRNA and rRNA; (4) >[n] gives the reference number for previous publications; (5) tRNA boundaries are shown with / and \ , all are read 5′ to 3′ on the strand printed; (6) selected *Pst*I and *Eco*RI restriction enzyme sites show the major fragments, denoted Pn and En numbered in order of size, used for cloning and sequencing. The gene labelled P1, starting at bp 777, is a *Paramecium* gene previously identified (20, 21).

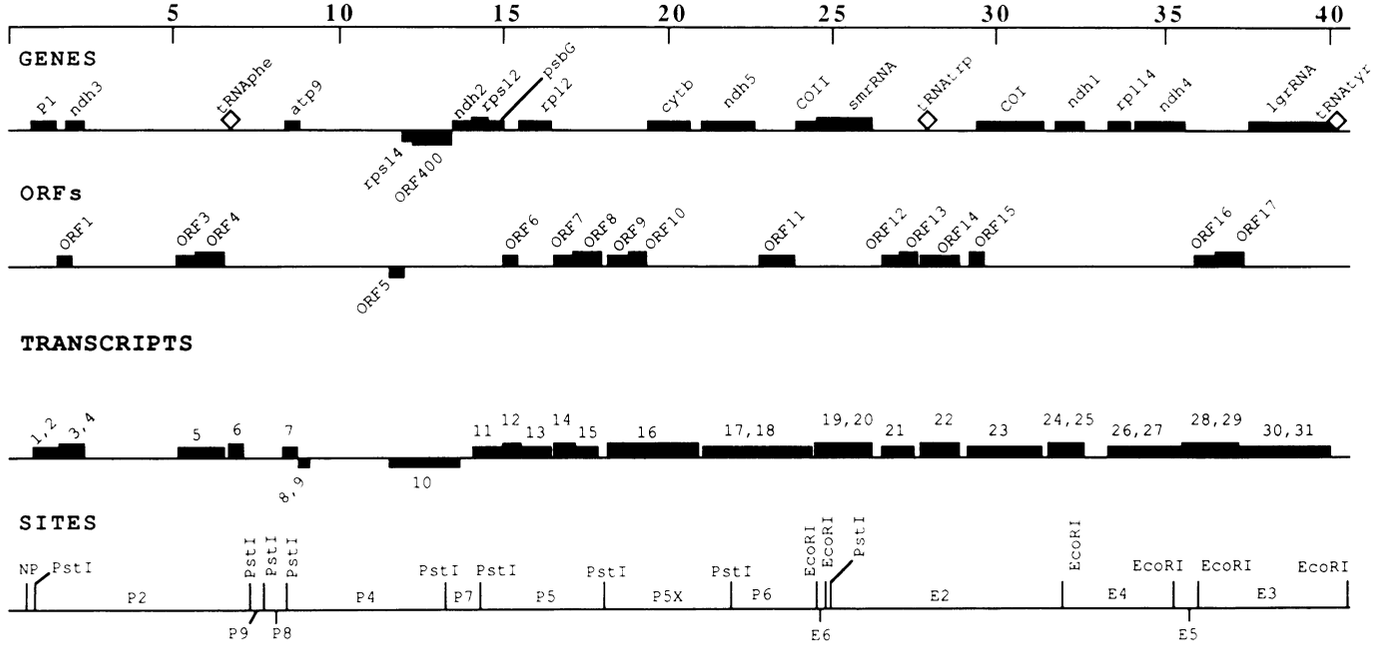


Figure 2. Genomic organization of the *Paramecium* mt DNA. Genes, ORFs, and transcripts shown by blocks above the line read left to right on the strand printed in Fig. 1; those shown below the line read right to left on the opposite strand. The major cloning and sequencing fragments are Pn and En (eg. P2) for those bounded by *PstI* and *EcoRI* restriction sites, respectively. NP, on the sites line, is for the 'non-palindrome' boundary of the linear duplex and the single strand DNA terminal loop.

cases, transcripts were more precisely mapped by S1 protection experiments. Details of the procedures used are described in previous publications on this genome (20, 23, 24). The probes were either nick translated cloned fragments isolated by preparative polyacrylamide gel electrophoresis, or strand-specific M13 clones. Sizes of specific-region probes varied from less than 100 bp to a few kbp depending on number and sizes of the hybridizing transcripts. The precise boundaries of most of the transcripts are therefore not known.

Sequence analysis

Searches of Genbank, EMBL, and National Biomedical Research-Protein Identification Resource (NBRF-PIR) databases using the Lipman-Pearson similarity programs (30,31) were performed as described in the accompanying paper (26). Other sequence analyses were performed using programs from the same computer resources.

RESULTS AND DISCUSSION

Gene content

The DNA sequences for *Paramecium* mt genes that have been identified by amino acid sequence comparisons with known genes from other organisms are shown in Fig. 1, and the locations are schematically represented in Fig. 2. The gene content is similar to that of many other mt genomes with a few significant exceptions. Of the three cytochrome *c* oxidase subunits (COI, COII, COIII) commonly encoded in mt DNA, COIII has not been identified in the *Paramecium* mt genome. The gene was also not detected initially in *T. brucei* by nucleotide sequence analysis but has since been identified by RNA and cDNA sequencing (32). The gene could not be identified in the DNA sequence because the RNA is edited resulting in the addition to transcripts of uridines which are not encoded in the genome and the deletion from the transcripts of some uridines which are encoded in the DNA sequence (32). There is no evidence for or against RNA editing in *Paramecium*.

Animal mt DNAs (33) encode 7 subunits of NADH dehydrogenase (proteins ND1−6 and ND4L). Many of these are also found encoded in the mt DNA of fungi (12), except that in yeast they appear to be lacking. Homologues of these subunits are also found in the chloroplasts of plants (34). Of the seven, we have identified only the genes *ndh1−nhd5* in *Paramecium*. Both of the other genes products, ND4L and ND6, are relatively small and are therefore difficult to identify in the divergent *Paramecium* mt genome. Most of the *Paramecium ndh* genes are approximately the same size as those in other organisms and in amino acid alignments, approximately 25−30% of the *Paramecium* residues are identical with those from other sources (26). However the *Paramecium ndh2* gene lacks approximately 450 nucleotides at the 5′ end compared to the gene in animals, plants, and fungi (22). Amino acid sequence alignments of the *Paramecium* gene product with others shows only 15−25% identity (22). In the mt genome of the protozoans *L. tarentolae* and *T. brucei*, only the genes coding for ND1, ND4, and ND5 have been identified so far (13).

In addition to these seven NADH dehydrogenase subunits, it has recently been shown (22) that at least one additional subunit is encoded in a number of organelle genomes: (1) the mt genome of *Paramecium* (ORF400), (2) the mt DNA of *L. tarentolae* (the overlapping ORFs 3 and 4 which, with a frame shift, code for the single gene) and (3) the chloroplast genomes of tobacco and liverwort (ORF393 and ORF392, respectively). An ORF with homology to all of these has been identified as a 49 kDa subunit of NADH dehydrogenase subunit which is encoded in the bovine nuclear genome (35). The *Paramecium* gene product, ORF400, is a long protein which shares approximately 34% sequence identity with that encoded in chloroplasts, a surprisingly high degree of similarity for *Paramecium* proteins. It is possible that other ORFs in the *Paramecium* mt genome encode homologues

of more as yet uncharacterized subunits of the enzyme complex.

Other genes commonly encoded in mt DNA include the gene for apocytochrome b, *cyt b*, which is found in all mt genomes including that of *Paramecium*, and subunits of ATP synthase: subunits 6 (*atp6*), 8 (*atp8*), and 9 (*atp9*) of the F0 moiety of the ATPase complex. All of these subunits are hydrophobic proteolipids which are required for the membrane component of the enzyme complex. The 'dicyclohexylcarbodiimide-binding protein', the 8 kDa gene product of *atp9*, has a corresponding polypeptide in eubacteria (the c chain, see ref. 36) and chloroplasts (subunit III), and the *atp6* gene product corresponds to chain *a* of the eubacterial enzyme complex. The mt subunits have been found encoded in either the nuclear or mt genome, depending on the organism. The only subunit so far identified in the *Paramecium* mt genome is *atp9*. This gene is encoded in the nucleus of some fungi, including *N. crassa* (37), and animals (37). It is encoded in the mt DNA of yeast (38), and plants (39), in addition to *Paramecium*. There is also evidence for a mt DNA encoded form of *atp9* in *N. crassa* (40), as well as the nuclear encoded subunit. None of the subunits has been identified in the kinetoplastid DNA of the flagellate protozoans (13).

The *atp6* and *atp8* genes are in adjacent overlapping reading frames (41) in the mt DNA of most animals. In many fungi, including yeast (10) where they are sometimes referred to as *oli2* and *aap1*, the same gene order is conserved, *atp8* followed by *atp6*, but there is a few hundred bp spacer between the genes. Since the *atp8* gene is relatively small, encoding about 66 amino acids, it may have been missed in the search of the rather divergent *Paramecium* mt genome. But the larger *atp6*, coding for about 250 amino acids, is more certainly not encoded in the protozoan mt DNA. Thus, the *Paramecium* mt genome appears somewhat unique in encoding *atp9*, which is not found in animal mt genomes, and not encoding *atp6* or *atp8*, which are found in most animal and fungal mt DNAs.

Other genes encoded in the *Paramecium* mt genome which are rarely found in other mt DNA are those for the ribosomal proteins L2 (*rpl2*), L14 (*rpl14*), S12 (*rps12*), and S14 (*rps14*)(26, 22). Although not encoded in animal, or fungal mt DNA, some of these genes have been identified in plant mitochondria (42), plant chloroplasts (43), and in the mt DNA of another ciliate, *Tetrahymena* (14). Three of these genes in the *Paramecium* genome are clustered in the left half, with two NADH dehygrogenase genes and the *psbG* gene, but *rpl14* is over 16 kbp away (Figs. 1, 2).

Another distinguishing feature of the *Paramecium* mt genome is the presence of the *psbG* gene but the significance of this is not entirely clear since the function of this gene is not known (44, 45). Other than in *Paramecium*, the gene has only been found in chloroplast genomes and there is evidence suggestive of a role for it in functions unique to chloroplasts (44, 45). On the other hand, it is co-transcribed with *ndhC*, the chloroplast equivalent of the mt *ndh3*, which suggests a close relationship with NADH dehydrogenase complexes. In *Paramecium* mt DNA, the gene is close to two NADH dehydrogenase subunit genes, ORF400 and *ndh2*. The gene is transcribed (see below) and the gene product is remarkably similar in sequence to its chloroplast homologues. These facts suggest an essential gene product which functions in mitochondria. Using antibodies raised against the *psbG* gene product, the location of the protein in various thylakoid membrane fractions of pea has recently been investigated (64). Based on these results and a sequence similarity between the *psbG* gene product and the *ndh* gene product of *E. coli*, it was proposed that *psbG* is not a photosystem 2 gene (64).

Transcripts

In Fig. 2 is shown a schematic representation of the transcripts that were detected by Northern hybridizations of total RNA with region-specific and strand-specific DNA probes. There are a large number of transcripts, at least 31 are shown, that cover almost all of the genome. For the regions where no RNA was detected, it is possible that the transcripts are in low copy number rather than being absent. The transcripts that are shown are almost all from the same strand, going left to right, with a few exceptions located in the left half of the molecule. Transcripts 8 and 9 are located just downstream of transcript 7 (for the *atp9* gene) but are encoded on the opposite strand. The transcripts 8 and 9 are probably overlapping since they are approximately 800 and 500 bases in length and were detected by DNA probes covering a range of only 800 bp. However, a search of the corresponding DNA strand for open reading frames showed the largest was only 212 bp long. None of the ORFs in the region has significant similarity with sequences in protein or DNA databases. It is likely these transcripts do not encode proteins but may function to regulate the transcription of the adjacent *atp9* gene.

As shown in Fig. 2, a number of transcripts are shown as a single line but are labelled with two numbers, as with transcripts 8 and 9 discussed above. Most of these represent overlapping transcripts such as transcripts 1 and 2 for the P1 protein gene or transcripts 24 and 25 for *ndh1* (20). This apparent pattern of two overlapping RNAs for a single gene has also been noted in another protozoan. Northern blot analysis of transcription in *T. brucei* has indicated the presence of double transcripts differing in size by 150−200 bp for all protein coding genes except *ndh5*. The size difference was not ascribed to polyadenylation since the RNA of both sizes were poly A+ (46). The four *Paramecium* transcripts named above have been studied in more detail (20) with S1 nuclease mapping of termini to show they are overlapping, but the functional significance of this arrangement is not known.

Transcripts are both mono-and poly-cistronic. ORF400, *rps14*, and ORF5, for example, are apparently on a single transcript 10, while COI, and the downstream *ndh1* are encoded on separate transcripts. There is no evidence for RNA splicing or trans-splicing but no transcripts have been sequenced directly for verification.

Both of the rRNA genes in *Paramecium* mt DNA are discontinuous (23, 24) and are represented by transcripts 19+20 and 30+31. This is also the case in *Tetrahymena* (17, 18). Discontinuities in the large subunit rRNA (lgrRNA) are known in eukaryotes where the 5.8S rRNA gene is separated from the 5′-end of the lgrRNA by a spacer that is excised during processing. But the discontinuity in the small subunit (smrRNA) is unusual. The discontinuity is at the 5′-end of the *Tetrahymena* mt smrRNA (18), and may be at the 3′-end of the *Paramecium* smrRNA (23) but there is some controversy about this point (18) and additional experiments are needed.

ORFs

ORFS were assigned based on the following criteria: (1) strand specific transcripts were detected for this region; (2) a minimum length of 300 bp between possible initiation codons ATG, ATA, ATT, ATC, GTG and stop codons TAG and TAA (see *Genetic Code and tRNAs* below); (3) Fickett's rule for codon bias was obeyed (47); (4) a general bias for codons ending in C was also considered (see *Nucleotide composition and codon usage* below).

In one case ORFs 12 and 13 (Figs. 1 and 2) are in different reading frames that are overlapping by about 200 bp. Although

there is no evidence for genes in this genome to overlap by as much as this, both of these ORFs satisfied the criteria listed above and were therefore included. Similarly, ORF15 is unusual in that it has no stop codon and is contiguous with the COI gene. The start of the COI gene was initially based on alignments with previously known genes even though the 5′ end of the mRNA was found to be about 275 bp upstream of the COI initiation codon (20). Subsequently, the *Tetrahymena* COI gene was found to have amino acid sequence similarity with that encoded in *Paramecium* mt DNA, upstream of the assigned initiation in *Paramecium* (15). It is therefore possible that the *Paramecium* COI gene starts upstream of the current assignment but, because of the uncertainty, this region is labelled as an ORF. Similar reasons apply to the assignment of ORF11 immediately upstream of the *Paramecium* COII gene whose initiation codon is uncertain (21).

Genome organization and comparisons

The gene content and organization of the compact mt genome of vertebrates is highly conserved (3−6). If invertebrates are added to the comparison, the genome size and gene content is still conserved but more variation is seen in the order of genes due to rearrangements and translocations (7, 8,). Fungal mt genomes are larger than those of animal and are less conserved in size, gene content, and gene organization (10, 12). Higher plants contain very large mt genomes that have a complex and multipartite structure because of intramolecular recombination events between repeated sequences (48, 49). The *Paramecium* mt genome, about 40 kbp, is larger than the animal mt DNA, but smaller than most fungal and plant mt DNA. Its gene organization is unlike any other known except for that of another protozoan ciliate, *Tetrahymena*, which also has a linear mt genome (14, 15, 50). These cilate mt genomes have little similarity with another protozoan—the kinetoplastid genomes of *Leishmania* and *Trypanosoma* (13) which have an actively transcribed informational region, about 15 kbp, and a nontranscribed divergent or variable regions containing a variety of repeated sequences of unknown function.

Although only a limited portion of the *Tetrahymena* mt genome has been published, it is known that the approximate location and direction of the two rDNA, the COI, COII and the *rpl14* genes are the same for the two ciliate mt DNAs (14, 15, 21). The tRNAtrp is also similarly located in the *Paramecium* and *Tetrahymena* mt genomes. However, tRNAglu, not found encoded in *Paramecium* mt DNA, is located in the *Tetrahymena* mt genome immediately downstream of *rpl14*, a gene common to both genomes. There are some other differences between the two genomes including total number of encoded tRNAs, the direction but not location of the tRNAphe gene, and the terminal inverted repeats of *Tetrahymena* which have not been found in *Paramecium*. The origin of replication is at a unique terminus of the *Paramecium* mt DNA, but near the center of the *Tetrahymena* mt genome (51).

Nucleotide composition and codon usage

As shown in Table 1, the *Paramecium* mt DNA sequence presented is 59% A+T. There is a significant asymmetric strand distribution of A and T, with a preponderance of T on the strand which corresponds to the sense mRNA strand of most of the genes. There are a few regions which are very A+T rich, such as at the replication initiation end of the molecule, but a plot of base composition (A+T) vs. nucleotide position over a window of 50 bp shows fluctuations which appear uniformly throughout the genome. The fluctuations, however, are rather extreme and frequent reflecting numerous runs of extreme composition such as poly T tracts. Such segments often occur at the boundaries of gene coding sequences but they are also found within genes. The known gene coding regions are not surrounded by large A+T rich, non-coding, spacer regions as in the mt DNA of yeast. Table 1 shows that the known gene coding region (17 genes, 13341 bp, including the 16 listed on the second line in Fig. 2 and ORF1 whose transcripts were previously studied, ref. 20) has a base composition almost identical to that of the whole genome.

A codon usage table that includes the 17 known *Paramecium* genes shows unique preferences for certain codons. This overall pattern of codon bias was compared to that of individual genes or a group of related genes (ribosomal proteins, for example) and all were, in general, consistent. No one gene or group of genes, including the 'chloroplast-like' genes (22) yielded a usage table that was significantly different from the rest. In particular, there is a marked preference for the use of codons ending in C as summarized in Table 1. A similar bias is seen with mt transcripts from human but not from *Drosophila* which has 93% of its codons ending in A or U. The reason for this bias is not clear, especially since the number of tRNAs that are utilized by the *Paramecium* mt genome, and therefore the importance of the third codon position, is not known (see below). Although 57% of the bp in the coding region are A or U, only 45% of the codons end in those bases. In many prokaryotes, on the other hand, there is a positive correlation between the A+T content of the genome and the relative use of A or U in the third codon position (52). Of the three tRNAs that have been identified in the *Paramecium* mt genome, one corresponds to the phenylalanine codon UUC and another to the tyrosine codon UAC, both of which end with C. After C, *Paramecium* mt genes use U next in frequency at the third codon postion. UUU is the most frequently used codon (6.6% of total codons), probably reflecting the common occurrence of poly T tracts throughout the genome. In some cases, U is the preferred in the third codon position rather than

Table 1

	% composition sense strand			% composition third codon position		
	human	Droso	Para	human	Droso	Para
A	29	32	25 (25)	36	45	18
T	15	44	32 (34)	15	48	27
G	12	12	19 (19)	5	3	19
C	33	11	24 (22)	43	3	36

The composition of the *Paramecium* (Para) mRNA sense strand for the 15 named protein coding genes, ORF1, and ORF400 shown in Fig. 2. The number in parentheses is for the compostion of the entire DNA strand printed in Fig. 1. The compostion for *D. yakuba* (Droso) is for all protein genes (7) and for human is all protein genes except *ndh6* (3).

C. The valine codon GUU is used more frequently than GUC, and the serine codon UCU is preferred over UCC.

Other significant features of codon usage in *Paramecium* are almost equal use of the tryptophan codons UGA and UGG; UAA is used as frequently as UAG as a stop codon. All codons are used but the CCG codon for proline is used only twice while CCC is used 70 times. Other rarely used codons include CGU and CGG for arginine, and GCG for alanine.

Genetic code and tRNAs

The genetic code that is utilized by *Paramecium* mt DNA is not precisely known, but a probable code has been proposed based on alignments, with other organisms, of corresponding gene products (20). Of the numerous variations to the universal genetic code that have been found in organelle DNA from different organisms, only the use of (1) UGA to code for tryptophan rather than for a stop, and (2) unusual start codons are found in *Paramecium* mt transcripts. Presumed translation initiation codons include AUG, AUA, AUU, AUC (*cyt b*, this work and ref. 26), GUG, and possibly GUA (20, 26). It is not known if these initiation codons are translated as methionine as in mammalian mt DNA (3−4). In the related ciliate, *Tetrahymena*, an unusual tRNAmet, encoded in the mt genome, is thought to translate AUG, AUA, and AUU as methionine (15). Such a tRNA may be encoded in the *Paramecium* mt genome (see below).

We have identified 3 tRNAs encoded in the mt DNA of *Paramecium*: tRNAtrp(anticodon UCA) (25), tRNAtyr(anticodon GUA) (53), and tRNAphe (anticodon GAA) identified in this communication. These tRNAs can be folded into the usual cloverleaf structure but, like other mt tRNA genes lack a CCA end. The tRNAphe gene, Fig. 1, has a normal primary structure with most of the usual conserved bases at predictable positions (54) as is the case with the other 2 *Paramecium* tRNAs (25, 53).

The number of tRNAs in a mt genome is less than the minimum of 31 required to translate all 61 codons according to Crick's Wobble hypothesis (55). Mammalian mt DNA encodes a set of 22 tRNAs (3−4), but tRNA genes appear to be entirely absent from the mt DNA of the trypanosoid protozoans, *T. brucei* (56) and *L. tarentolae* (57). Recent evidence strongly suggests a nuclear transcriptional origin of the 35−40 tRNAs found in mitochondria of *L. tarentolae* (63). *Tetrahymena* mt DNA encodes about 8 tRNAs (15), including those found in *Paramecium* mt DNA. Only three tRNA genes have been found in the mt DNA of the unicellular green alga, *C. reinhardtii* (15).

A new set of recognition rules has been proposed to account for some of these results (eg. 59) so that a tRNA species with an unmodified U at the first position of the anticodon is able to recognize all four codons specifying the same amino acid, and a tRNA with a modified U is unable to recognize U and C at the wobble position of a codon. In this scheme, the required number of tRNAs is 23. In those organisms with only a few tRNAs, there is evidence for importation of nuclear DNA-encoded cytoplasmic tRNAs into the mitochondria (15, 16, 60). With ciliates, this is particularly interesting because it is known that in the nuclear genetic code the universal stop codons UAG and UAA are translated as glutamine (61,62) which would leave the mt genome without a stop codon if the corresponding tRNA were imported. All of the tRNAs that are mt DNA encoded in the two ciliates, *Paramecium* and *Tetrahymena*, recognize the members of a codon family that contain two different amino acids. The tRNAs that recognize a codon family with one amino acid are presumably imported but precise numbers and types are not yet known. As Suyama has pointed out (15), for ciliate mt genomes, it is not known if more than one tRNA is required for a single amino acid codon family or if a single RNA with an unmodified U at the wobble position can translate all four codons as in most mt DNA. If most of the tRNAs utilized are nuclear encoded then one might expect that the universal code is more closely adhered to in ciliate mt DNA compared to other mt genomes. This appears to be the case since the translation of UGA as tryptophan and apparently some unusual start codons are the only known deviations from the universal code. The completely normal secondary structures of the three known *Paramecium* tRNAs are also in accord with this idea.

It is possible that there are more tRNAs in the mt genome than are reported here. The replication terminus end of the linear molecule has not been cloned and it is estimated that there are a few hundred bp beyond the end of the sequence shown here. Several tRNAs could be clustered there and a specific search for tRNAs has not been done in this laboratory. It is unlikely that we have missed some tRNAs because they are split by introns, but a number of mt genomes reportedly have highly unusual tRNAs and one, tRNAmet, has been suggested to be encoded in the *Paramecium* mt genome (17, 53) based on its similarity to the one in *Tetrahymena*. We cannot exclude the possible existence of this and other unusual tRNA genes in the *Paramecium* mt genome.

ACKNOWLEDGMENTS

We gratefully acknowledge Dr. M. Silliker for a critical reading of the manuscript. This work was supported by NSF grant DMB-86-05319. Computer resources for some of this work was provided by BIONET National Computer Resource for Molecular Biology which is funded by the Biomedical Research Technology Program, Division of Research Resources, National Institutes of Health, Grant Number P41RR01685.

REFERENCES

1. Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G.J., and Woese, C.R. (1985) Proc. Natl. Acad. Sci. USA 82, 4443−4447.
2. Gray, M.W., Sankoff, D., and Cedergren, R.J. (1984) Nucl. Acids Res. 12, 5837−5852.
3. Anderson, S., Bankier, A.T., Barrell, B.G., De Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperson, 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, 457−464.
4. Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W., and Clayton, D.A. (1981) Cell 26, 167−180.
5. 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.
6. Roe, B.A., Din-Pow, M., Wilson, R.K., and Wong, J. (1985) J. Biol. Chem. 260, 9759−9774.
7. Clary,, D.O. and Wolstenholme, D.R. (1985) J. Mol. Evol. 22, 252−271.
8. Jacobs, H.T., Elliott, D.J., Math, V.B., and Farquharson, A. (1988) J. Mol. Biol. 202, 185−217.
9. Cummings, D.J., and Domenico, J.M. (1988) J. Mol. Biol. 204, 815−839.
10. de Zamaroczy, M. and Bernardi, G. (1986) Gene 47, 155−177.
11. Dujon, B. (1983) In Schweyen, R.J. et al. (eds.). Mitochondria 1983. Walter de Gruyter, Berlin, pp. 1−24.
12. Brown, T.A., Waring, R.B., Scazzocchio, C. and Davies, R.W. (1985) Curr. Genet. 9, 113−119.
13. Simpson, L., Neckelman, N., de la Cruz, V.F., Simpson, A.M., Feagin, J.E., Jasmer, D.P., and Stuart, K. (1987) J. Biol. Chem. 262, 6182−6196.
14. Suyama, Y. and Jenney, F. (1989) Nucl. Acids Res. 17, 803.
15. Ziaie, Z. and Suyama, Y. (1987) Curr. Genet. 12, 357−368.
16. Suyama, Y. (1986) Curr. Genet. 10, 411−420.
17. Heinonen, T.Y.K., Schnare, M.N., Young, P.G., and Gray, M.W. (1987) J. Biol. Chem. 262, 2879−2887.

18. Schnare, M.N., Heinonen, T.Y.K., Young, P.G., and Gray, M.W. (1986) J. Biol. Chem. 261, 5187–5193.
19. Pritchard, A.E., Laping, J.L., Seilhamer, J.J., and Cummings, D.J. (1983) J. Mol. Biol. 164, 1–15.
20. Pritchard, A.E., Seilhamer, J.J., and Cummings, D.J. (1986) Gene 44, 243–253.
21. Mahalingam, R., Seilhamer, J.J., Pritchard, A.E., and Cummings, D.J. (1986) Gene 49, 129–138.
22. Pritchard, A.E., Venuti, S.E., Ghalambor, M.A., Sable, C.L., and Cummings, D.J. (1989) Gene 78, 121–134.
23. Seilhamer, J.J., Olsen, G.J., and Cummings, D.J. (1984) J. Biol. Chem. 259, 5167–5172.
24. Seilhamer, J.J., Gutell, R.R., and Cummings, D.J. (1984) J. Biol. Chem. 259, 5173–5181.
25. Seilhamer, J.J. and Cummings, D.J. (1982) Mol. Gen. Genet. 187, 236–239.
26. Pritchard, A.E., Sable, S.E., Venuti, S.E., and Cummings, D.J. (1989) Nucl. Acids Res. 18, 163–171. (accompanying paper)
27. Cummings, D.J. and Laping, J.L. (1981) Mol. Cell. Biol. 1, 972–982.
28. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 449–560.
29. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
30. Lipman, D.J. and Pearson, W.R. (1985) Science 227, 1435–1441.
31. Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448.
32. Feagin, J.E., Abraham, J.M., and Stuart K. (1988) Cell 53, 413–422.
33. Chomyn, A., Mariottini, P., Cleeter, M.J., Ragan, C.I., Matsuno-Yagi, A., Hatefi, Y., Doolittle, R.F., and Attardi, G. (1985) Nature 314, 592–597.
34. Matsubayashi, T., Wakasugi, T., Shinozaki, K., Yamaguchi-Shinozaki, K., Zaita, N., Hidaka, T., Meng, B.Y., Ohto, C., Tanaka, M., Kato, A., Maruyama, T., and Sugiura, M. (1987) Mol. Gen. Genet. 210, 385–393.
35. Fearnley, I.M., Runswick, M.J., and Walker, J.E. (1989) EMBO J. 8, 665–672.
36. Nelson, N. and Taiz, L. (1989) TIBS 14, 279–282.
37. 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.
38. Macino, G. and Tzagoloff, A. (1979) J. Biol. Chem. 254, 4617–4623.
39. Dewey, R.E., Schuster, A.M., Levings, C.S., III, and Timothy, D.H. (1985) Proc. Nat. Acad. Sci. USA 82, 1015–1019.
40. Brown, T.A., Ray, J.A., Waring, R.B., Scazzocchio, C., and Davies, R.W. (1984) Curr. Genet. 8, 489–492.
41. Fearnley, I.M. and Walker, J.E. (1986) EMBO J. 5, 2003–2008.
42. Wahleithner, J.A. and Wolstenholme, D.R. (1988) Nucl. Acids Res. 16, 6897–6913.
43. 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.
44. Steinmetz, A.A., Castroviejo, M., Sayre, R.T., and Bogorad, L. (1986) J. Biol. Chem. 261, 2485–2488.
45. Steinmuller, K., Ley, A.C., Steinmetz, A.A., Sayre, R.T. and Bogorad, L. (1989) Mol. Gen. Genet. 216, 60–69.
46. Simpson, L. (1987) Ann. Rev. Microbiol. 41, 363–382.
47. Fickett, J. (1982) Nucl. Acids Res. 10, 5303- 5318. 48. Palmer, J.D. and Shields, C.R. (1984) Nature 307, 437–440.
49. Palmer, J.D. and Herbon, L.A. (1987) Curr. Genet. 11, 565–570.
50. Suyama, Y., Fukuhara, and Sor, F. (1985) Curr. Genet. 9, 479–493.
51. Arnberg, A.C., Van Bruggen, E.F.J., Clegg, R.A., Upholt, W.B., and Borst, P. (1974) Biochim. Biophy. Acta 361, 266–276.
52. Bibb, M.J., Findlay, P.R., and Johnson, M.W. (1984) Gene 30, 157–166.
53. Seilhamer, J.J. and Cummings, D.J. (1981) Nucl. Acids Res. 9, 6391–6406.
54. Rich, A. and RajBhandary, U.L. (1976) Ann. Rev. Biochem. 45, 805–860
55. Crick, F.H.C. (1968) J. Mol. Biol. 19, 548–555.
56. Benne R. and Sloof, P. (1987) BioSystems 21, 51–68.
57. Simpson L., Neckelman N., de la Cruz, V.F., Simpson, A.M., Feagin, J.E., Jasmer, D.P., and Stuart, K. (1987) J. Biol. Chem. 262, 6182–6196.
58. Boer, P.H. and Gray, M.W. (1988) Curr. Genet. 14, 583–590.
59. Heckman, J.E., Sarnoff, J., Alzener-DeWeed, B., Yin, S., and RajBhandary, U.L. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 3159–3163.
60. Marechal-Drouard, L, Weil, J-H., and Guillemaut, P. (1988) Nucl. Acids Res. 12, 4777–4788.
61. Preer, J.R., Preer, L.B., Rudman, B.M., and Barnett A.J. (1985) Nature 314, 188–190.
62. Horowitz, S. and Gorovsky, M.A. (1985) Proc. Natl. Acad. Sci. USA 82, 2452–2455.
63. Simpson, A., Suyama, Y., Dewes, H., Campbell, D.A., and Simpson, L. (1989) Nucl. Acids Res. 17, 5427–5445.
64. Nixon, P., Gounaris, K., Coomber, S., Hunter, C., Dyer, T., and Barber, J. (1989) J. Biol. Chem. 264, 14129–14135.