Nucleotide sequence of the genes for ribosomal protein S4 and tRNA^{Arg} from the chlorophyll *c*-containing alga *Cryptomonas* Φ

Susan E.Douglas and Dion G.Durnford
Atlantic Research Laboratory, National Research Council of Canada, 1411 Oxford Street, Halifax, Nova Scotia, Canada B3H 3Z1

Submitted February 20, 1990 EMBL accession no. X51511

Cloning and mapping of the plastid genome from *Cryptomonas* Φ has shown that the gene arrangement in this alga differs significantly from land plants (1). Sequence analysis of the region upstream of *rbc*L revealed the presence of the genes for ribisomal protein 4 (*rps4*) and tRNA^{Arg}. The location of these genes approximately 1.5 kb from *rbc*L is different from the arrangement found in land plants where the genes are separated by a least 5 kb of DNA which encodes several genes (2). Furthermore, in land plants, *rps4* is usually flanked by the genes for tRNA^{Ser} and tRNA ^{Thr} rather than tRNA^{Arg} as found in this alga. *rps4* is encoded on the same strand as *rbc*L but tRNA^{Arg} is found on the opposite strand. The *rps4* gene product from *Cryptomonas* Φ shows the highest amino acid similarity to that from *Marchantia polymorpha* (62%), followed by *Nicotiana tabacum* (57%) and *Zea mays* (55%).

ACKNOWLEDGEMENTS

The technical assistance of Jeff Gallant is gratefully appreciated. This is NRCC publication number 31135.

REFERENCES

1. Douglas,S.E. (1988) *Curr. Gen.* **14**, 591–598.
2. Ohyama,K., Fukuzawa,H., Kohchi,T., Shirai,H., Sano,T., Sano,S., Umesono,K., Shiki,Y., Takeuchi,M., Chang,Z., Aota,S., Inokuchi,H. and Ozeki,H. (1986) *Nature* **322**, 572–574.

```
        10         20         30         40         50         60         70         80         90        100
TAGGCTCAGT AGGATTCGAA CCTACATTAG AGACTTAGAA GGTCCCTGTC CTGATCCATT AGACGATGAG CCCTAAATTC GTATTTAAAT AAATTAGCGT
       110        120        130        140        150        160        170        180        190        200
GTCTTACAAT AAATTAAAAC TAATTTAAAG TGGGCGGTAC TGGATTTGAA CCAGTGACCA CCTGCTTGTA AGGCAGGCGC TCTACGCTGA GCTAACCGCC
       210        220        230        240        250        260        270        280        290        300
CGCTATTGTT TTACTTAACA TTAATATTAA CATTTATAGT CTATAGAATT CAACTCATTT ATTAAATAAA AATTTATTAA ATAATTATAA AACTTGTTAT
       310        320        330        340        350        360        370        380        390        400
TAAGTAAAAT AAAAATCAAG AGAAGCTGGT GAAGGGACTT GAACCCGCAA CCTACTGATT ACAAATCAGT TGCTCTACCA ATTGAGCTAC ACCAGCATTT
       410        420        430        440        450        460        470        480        490        500
AATATTTAAT AATGCATATA TTAAATATAC AATATATTAT TTAAGATAGT AAAGTATTTT TTTGTTTAAT AAATTAATAT ATATATATTA TACTAATTTT
       510        520        530        540        550    557
ATGCTTATAA AATAAAAATT AAAATAAATA AAAATAATAA GTGGAGAAAA AGTAAAT
558                                                                                                        641
ATG TCT CGT TAC AGA GGA GCA GTC ATA AAA ATT ATT CGT CGT TTA GGA GAA CTT CCA GGG TTA ACA CGC AAA ACA ACA ACA CGA
 M   S   R   Y   R   G   A   V   I   K   I   I   R   R   L   G   E   L   P   G   L   T   R   K   T   T   T   R
642                                                                                                        725
ACA TCT AGA CCA GGT CAA CAT GGT ACA CAG GCG AGA AAA CCA TCA GAA TAC GCA ATT CGA TTA GAA GAA AAA CAG AAA TTA CGT
 T   S   R   P   G   Q   H   G   T   Q   A   R   K   P   S   E   Y   A   I   R   L   E   E   K   Q   K   L   R
726                                                                                                        809
TTT AAT TAC GGG TTA ACT GAA AAA CAA TTA TTA CAA TAT GTT AGA ACA GCG AAA CGT ATA AAA GGT TCT ACT GGT GAA GCT CTA
 F   N   Y   G   L   T   E   K   Q   L   L   Q   Y   V   R   T   A   K   R   I   K   G   S   T   G   E   A   L
810                                                                                                        893
TTA CAA TTA CTC GAA ATG AGA CTC GAC AAC GTA ATA TTT CGT CTA GGA ATG GCA CCA ACA ATA CCT GCA GCT AGA CAA TTA GTT
 L   Q   L   L   E   M   R   L   D   N   V   I   F   R   L   G   M   A   P   T   I   P   A   A   R   Q   L   V
894                                                                                                        977
AAT CAC GGC CAT ATA AAA GTG AAT AAC ACT AGA GTG TCT ATA CCT AGC TAT CAA TGC AAA GCT GGT GAT ATG ATC TCA ATA CGT
 N   H   G   H   I   K   V   N   N   T   R   V   S   I   P   S   Y   Q   C   K   A   G   D   M   I   S   I   R
978                                                                                                       1061
CAG CAT CCG AAA TCA CAA AGT ATA GTA AAA AAT TAC TTA CAA TTC CCG GGT TTA GCG AAT ATG CCT AAT CAT TTA CAA ATT GAT
 Q   H   P   K   S   Q   S   I   V   K   N   Y   L   Q   F   P   G   L   A   N   M   P   N   H   L   Q   I   D
1062                                                                                                       1145
AAA GAT AAT TTA ACT GGT AAA ATT AAC GGT ATT ATT GAG CGT GAT TGG GTT GCA TTA AAT GAG CTT TTA ATC GTA GAG TAT TAC
 K   D   N   L   T   G   K   I   N   G   I   I   E   R   D   W   V   A   L   N   E   L   L   I   V   E   Y   Y
1146                                                  1200
TCA AGA AAA GGA TAA AAATTACCAA GAACTTTCTT CACTTGGCAA TGTATCTGAG
 S   R   K   G   Z
```

Nucleotide sequence of the region upstream of *rbc*L from *Cryptomonas* Φ containing the genes for *rps4* and tRNA^{Arg}. The gene for rTNA^{Arg} is indicated by underlining (positions 1–73) and the deduced amino acid sequence of *rps4* is shown beneath the nucleotide sequence. Both strands were sequenced for the entire region presented.