Volume 17 Number 12 1989 Nucleic Acids Research

Nucleotide sequence of the *Euglena gracilis* chloroplast genes for isoleucine, phenylalanine and cysteine transfer RNAs and ribosomal protein S14

Jac A.Nickoloff+, David A.Christopher[2], Robert G.Drager[2] and Richard B.Hallick[1,2]*

[1]Department of Biochemistry and [2]Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721, USA

Submitted May 15, 1989 EMBL accession no. X15240

The nucleotide sequence of 780 bp of *Euglena gracilis* Z chloroplast DNA encoding three tRNA genes, *trnI*-CAU, *trnF*-GAA, *trnC*-GCA, and the gene for ribosomal protein S14 (*rps14*) has been determined. The genes are located in the chloroplast DNA Eco-A fragment within the overlap of fragments HaeII-E and PvuII-E and C (1). The genes are all of the same polarity, organized as *trnI*-39 bp spacer-*rps14*-52 spacer-*trnF*-1 bp spacer-*trnC*. The *trnI* is unusual in having a mismatched base pair (A--A) in the DHU stem and a long variable loop. Although *trnI* has a methionyl-tRNA anticodon (CAU), it shares the most sequence identity (67%) with the spinach isoleucyl-tRNA (also-CAU), which is charged with isoleucine (2). The Euglena *rps14* locus (101 codons) is unique in having a single 106 bp intron (positions 220-325), similar to previously described introns (5). The derived amino acid sequence is 48% and 37% identical to the corresponding sequences from liverwort (3) and E. coli (4), respectively.

```
                trnI
GCGTCTATAATTTGTAGGCGCATTTATGGCAGAGTGGACGATAGCACGGGACTCATAATCTCGCTCCGGAAGGACGTCGCTGGTTCGAATCCAGCTGAAT  100

                                rps14  exon 1
GCATCTTATTCGAGCATTATTTAACCTAAAAATTTTTTATTTATGTCAAAAAAAAGTCTTATAGCAAGACAGAGAAAACGAATAATATTAGTATTAATAC  200
                                  M  S  K  K  S  L  I  A  R  Q  R  K  R  I  I  L  V  L  I

ATTCCCATAACCGTTATGTATACAGGACAAATGGAAAAGACGAAAAATCTTTTGAGAAAAAATTGCGGATTTACTCTTTTTTGCAAAAATTACCTAGAAA  300
H  S  H  N  R  Y  V  Y  R  T  N  G  K  D  E  K  S  F  E  K  K  L  R  I  Y  S  F  L  Q  K  L  P  R  N

TAGTTTGCGTTGCCGATTAATTTGATTTTCTATTTGTCATTCAAATTTTATAGCTGTTAATAATTTGAAAAATTTATAAAATTTTTATATTTATTTTTTT  400
  S  L  R  C  R  L

                rps14.exon 2
GAGTATTTTTATTTGTTAACTCTTTCGAAATCGTTGTTACGTAACAGGGAGATCACGTGGATATTTTAGGACTTTTGGATTATCACGACATATTCTTCGA  500
                  R  N  R  C  Y  V  T  G  R  S  R  G  Y  F  R  T  F  G  L  S  R  H  I  L  R

GATATGGCTCATTATGGTTTACTTCCAGGTGTAACCAAAGCGAGTTGGTAAACTTAAATTTACTTATAATAATATAAGTATAAATTAATTTATAATAATA  600
 D  M  A  H  Y  G  L  L  P  G  V  T  K  A  S  W  *
   trnF                                                  trnC
ATAGCTGGGATAGCTCAGTTGGTAGAGCGGAGGACTGAAAATCCTTGTGTCACCAGTTCAAATCTGGTTCCTAGCATGGCGGCATGGCCAAGCGGTAAGG  700

CAGAAGATTGCAAATCTTTTATTCCCCAGTTCGATTCTGGGTGTCGTCTTTAAGAGAATTGAAGAAACAATAAAACAAAA  780
```

*To whom correspondence should be addressed

+Present address: Life Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

References

1. Cushman, J.C., Hallick, R.B., and Price, C.A. (1988) Curr. Gen. 13, 159-179.
2. Kashdan, M.A. and Dudock, B.S. (1982) J. Biol. Chem. 257, 11191-11194.
3. Umesono, K., Inokuchi, H., Ohyama, K., and and Ozeki, H. (1984) Nucleic Acids Res. 12, 9551-9565.
4. Ceretti, D.P., Dean, D., Davis, G.R., Bedwell, D.M., and Nomura, M. (1983) Nucleic Acids Res. 11, 2599-2616.
5. Christopher, D.A., Cushman, J.C., Price, C.A., and Hallick, R.B. (1988) Curr. Genet. 14, 275-286.

