Available online at www.sciencedirect.com
SCIENCE DIRECT
Molecular & Biochemical Parasitology 146 (2006) 24–29
MOLECULAR & BIOCHEMICAL PARASITOLOGY

Unique organisation of tRNA genes in *Entamoeba histolytica*

C. Graham Clark [a,*], Ibne Karim M. Ali [a,1], Mehreen Zaki [a,2], Brendan J. Loftus [b], Neil Hall [b,c]

[a] *Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK*
[b] *The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA*
[c] *The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK*

Received 4 August 2005; received in revised form 12 October 2005; accepted 14 October 2005
Available online 28 October 2005

Abstract

The genome sequence of the protistan parasite *Entamoeba histolytica* HM-1:IMSS has been completed recently. Among the findings has been a unique organisation for the tRNA genes in this organism. Forty-two of the tRNA isoacceptor types are encoded in tandem arrays that vary in unit length from 490 to 1775 basepairs and contain from 1 to 5 tRNA genes. In three cases a 5S RNA gene is also present in the unit. An estimated 10% of the genome is made up of these arrays. Interspersed between RNA-encoding sequences are short tandem repeats that are polymorphic between isolates and, in some cases, within isolates. The number and organisation of tRNA genes in *E. histolytica* is unprecedented. In addition to encoding the tRNAs of the organism we propose that the arrays may fulfil a structural role in the genome.

Keywords: *Entamoeba histolytica*; Genome organisation; Tandem repeat; Telomeres; Transfer RNA

1. Introduction

As more genome projects reach completion it is to be anticipated that novel ways of organising genes and genomes will be uncovered, such as the evidence for a whole genome duplication in *Saccharomyces* [1] and chromosome-sized transcription units in kinetoplastids [2]. The recently completed genome of *Entamoeba histolytica* [3], the parasite responsible for amoebic dysentery and amoebic liver abscesses, was already known to have some unusual features prior to the start of the sequencing project. Its ribosomal RNA genes are carried on 24 kb circular episomes with two transcription units in an inverted repeat, and these circles make up about 20% of the total cell DNA [4]. A number of less abundant nuclear DNA circles of varying sizes have also been described but their function is not known [5,6]. Large families of transcribed but apparently untranslated 'genes' have also been described [7].

We previously identified a small number of polymorphic loci in *E. histolytica* containing variable numbers of short (8–16 bp) tandemly repeated A + T-rich sequences (STRs) [8]. These clusters of short repeats were themselves organised into tandem arrays of larger unit size. Database comparisons identified the presence of tRNA-related sequences adjacent to the STRs in each case (Zaki and Clark, unpublished). We have extended our analysis of such sequences to the entire genome and in this report describe the unique organisation of tRNA genes in this organism.

2. Materials and methods

2.1. Genome sequencing

The *E. histolytica* HM-1:IMSS genome was sequenced using four distinct libraries [3]. Two contained small inserts (ca. 2 kb) and two contained medium-sized inserts (ca. 8 kb). Twenty-nine percent of all sequence reads were excluded from the genome assembly due to their repetitive nature. Approximately two-thirds of these were derived from the ribosomal DNA episome; the rest proved to be derived from the tRNA arrays.

Abbreviations: STR, short tandem repeat

* Corresponding author. Tel.: +44 20 7927 2351; fax: +44 20 7636 8739.
E-mail address: graham.clark@lshtm.ac.uk (C.G. Clark).
[1] Present address: Department of Microbiology, Stanford University School of Medicine, Stanford, CA 94305-5107, USA.
[2] Present address: School of Biosciences, Division of Molecular Cell Biology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

0166-6851/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2005.10.013

2.2. tRNA array identification

Both assembled and excluded sequences were scanned using tRNA-scan SE [9]. Reads encoding the same tRNA isoacceptor type were aligned (http://prodes.toulouse.inra.fr/multalin/multalin.html; [10]), manually edited, and a consensus sequence obtained. Consensus sequences of tRNA array units are available in the Third Party Annotation Section of the DDBJ/EMBL/GenBank databases under the accession numbers BK005648–BK005672.

2.3. STR, codon usage and S/MAR identification

Tandem Repeat Finder (http://tandem.bu.edu/trf/trf.html; [11]) was used to identify the STRs in between tRNA genes, although the precise repeat sequences and boundaries of the repeat regions are often ill-defined. Codon usage for *E. histolytica* was obtained from the Codon Usage Database (http://www.kazusa.or.jp/codon/). Scaffold/matrix attachment region (S/MAR) prediction was performed using MARSCAN (http://ngfnblast.gbf.de/cgi-bin/emboss.pl?_action=input&_app=marscan).

2.4. DNA analysis

E. histolytica HM-1:IMSS was cultured [12] and DNA was isolated, PCR amplification performed, and PCR products separated in 1.5% agarose gels as previously described [8]. For PCR analysis of unit variation, 30 cycles of 94 °C for 30 s, primer-specific annealing temperature for 30 s and 72 °C for 30 s were employed. Primers used were: W-I: W-I5 GCCGGTCAGTGGTTCAATCC and W-I3 GCGACCCTGGCGTTATTAGC, annealing temperature 63 °C; E-Y: E-Y5 TACATAAGTCGTGGTAAAGAGAAG and E-Y3 CTACATCTACAGTCCTCCGCT, annealing temperature 52 °C; V-F: V-F5 GGTTTCATGGTGTAGTTGGT and V-F3 GAACTTCAGTCTAACGCTGC, annealing temperature 55 °C.

2.5. Copy number estimation

Genomic DNA was extracted from a known number of cells and PCR products produced from defined regions of array units were quantified using fluorescence on a gel (Phoretix software) by comparison to a standard (Hyperladder II, Promega). Slot blots were produced carrying a range of dilutions of both genomic DNA (10^6–5×10^4 cells) and PCR products (10^6–10^4 copies). Portions of tRNA array units were amplified for use as a probe using species-specific primers [13]. These primers minimize the tRNA gene content of the probe and eliminate cross hybridization with arrays encoding related tRNA genes. In each experimental hybridization, cross-hybridization was evaluated by including a control slot-blot carrying 10^7 copies of every STR region. Hybridizations were exposed to a PhosphorImager screen and the results quantified using ImageQuant 5.0 software (Amersham BioSciences).

Table 1
tRNA array unit characteristics

Array name	Isoacceptor types	Unit length (bp)	5S RNA?	GenBank accession number
[A^{AGC}]	Ala^{AGC}	559	×	BK005648
[ALL]	Ala^{CGC}, Leu^{TAA}, Leu^{CAA}	1154	×	BK005649
[ASD]	Ala^{TGC}, Ser^{GCT}, Asp^{GTC}	1166	×	BK005650
[G^{GCC}]	Gly^{GCC}	813	×	BK005662
[G^{TCC}]	Gly^{TCC}	490	×	BK005663
[H^{GTG}]	His^{GTG}	634	×	BK005664
[LS]	Leu^{CAG}, Ser^{CGA}	961	×	BK005667
[LT]	Leu^{AAG}, Thr^{AGT}	935	×	BK005666
[MR]	$eMet^{CAT}$, Arg^{TCG}	1031	×	BK005653
[NK1]	Asn^{GTT}, Lys^{CTT}	1006	×	BK005655
[NK2]	Asn^{GTT}, Lys^{CTT}	1251	×	BK005656
[P^{TGG}]	Pro^{TGG}	761	×	BK005669
[R^{TCT}]	Arg^{TCT}	686	×	BK005654
[R5]	Arg^{ACG}	846	Yes	BK005651
[RT]	Arg^{CCT}, Thr^{AGT}	964	×	BK005652
[SD]	Ser^{TGA}, Asp^{GTC}	767	×	BK005657
[SPPCK]	Ser^{AGA}, Pro^{AGG}, Pro^{CGG}, Cys^{GCA}, Lys^{TTT}	1775	×	BK005659
[SQCK]	Ser^{AGA}, Gln^{CTG}, Cys^{GCA}, Lys^{TTT}	1403	×	BK005658
[TQ]	Thr^{CGT}, Gln^{TTG}	767	×	BK005660
[TX]	Thr^{TGT}	1107	×	BK005670
[V5]	Val^{TAC}	812	Yes	BK005671
[VF]	Val^{GAC}, Phe^{GAA}	965	×	BK005668
[VME5]	Val^{CAC}, $iMet^{CAT}$, Glu^{CTC}	1466	Yes	BK005672
[WI]	Trp^{CCA}, Ile^{AAT}	1135	×	BK005665
[YE]	Tyr^{GTA}, Glu^{TTC}	938	×	BK005661

tRNA isoacceptor types are given with their anti-codon. One unit length is given but other length variants are sometimes present, differing primarily in number of STRs. Consensus sequences and unit organisations are given at http://homepages.lshtm.ac.uk/entamoeba/units/units.htm.