

Research article

Open Access

## Gene organization and sequence analyses of transfer RNA genes in Trypanosomatid parasites

Norma E Padilla-Mejía<sup>1</sup>, Luis E Florencio-Martínez<sup>1</sup>, Elisa E Figueroa-Angulo<sup>1</sup>, Rebeca G Manning-Cela<sup>2</sup>, Rosaura Hernández-Rivas<sup>2</sup>, Peter J Myler<sup>3,4</sup> and Santiago Martínez-Calvillo\*<sup>1</sup>

Address: <sup>1</sup>Unidad de Biomedicina, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Av de los Barrios 1, Col Los Reyes Iztacala, Tlalnepantla, Edo de México, CP 54090, México, <sup>2</sup>Departamento de Biomedicina Molecular, Centro de Investigación y de Estudios Avanzados del IPN, Apartado Postal 14-740, 07360, México, DF, México, <sup>3</sup>Seattle Biomedical Research Institute, 307 Westlake Avenue N, Seattle, WA 98109-5219, USA and <sup>4</sup>Departments of Global Health, and Medical Education & Biomedical Informatics, University of Washington, Seattle, WA 98195, USA

Email: Norma E Padilla-Mejía - qfedime@yahoo.com.mx; Luis E Florencio-Martínez - luise@campus.iztacala.unam.mx; Elisa E Figueroa-Angulo - efiguer2000@yahoo.com; Rebeca G Manning-Cela - rmanning@cinvestav.mx; Rosaura Hernández-Rivas - roherman@cinvestav.mx; Peter J Myler - peter.myler@sbri.org; Santiago Martínez-Calvillo\* - scalv@campus.iztacala.unam.mx

\* Corresponding author

Published: 18 May 2009

Received: 17 December 2008

BMC Genomics 2009, 10:232 doi:10.1186/1471-2164-10-232

Accepted: 18 May 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/232>

© 2009 Padilla-Mejía et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** The protozoan pathogens *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi* (the Trityps) are parasites that produce devastating human diseases. These organisms show very unusual mechanisms of gene expression, such as polycistronic transcription. We are interested in the study of tRNA genes, which are transcribed by RNA polymerase III (Pol III). To analyze the sequences and genomic organization of tRNA genes and other Pol III-transcribed genes, we have performed an *in silico* analysis of the Trityps genome sequences.

**Results:** Our analysis indicated the presence of 83, 66 and 120 genes in *L. major*, *T. brucei* and *T. cruzi*, respectively. These numbers include several previously unannotated selenocysteine (Sec) tRNA genes. Most tRNA genes are organized into clusters of 2 to 10 genes that may contain other Pol III-transcribed genes. The distribution of genes in the *L. major* genome does not seem to be totally random, like in most organisms. While the majority of the tRNA clusters do not show synteny (conservation of gene order) between the Trityps, a cluster of 13 Pol III genes that is highly syntenic was identified. We have determined consensus sequences for the putative promoter regions (Boxes A and B) of the Trityps tRNA genes, and specific changes were found in tRNA-Sec genes. Analysis of transcription termination signals of the tRNAs (clusters of Ts) showed differences between *T. cruzi* and the other two species. We have also identified several tRNA isodecoder genes (having the same anticodon, but different sequences elsewhere in the tRNA body) in the Trityps.

**Conclusion:** A low number of tRNA genes is present in Trityps. The overall weak synteny that they show indicates a reduced importance of genome location of Pol III genes compared to protein-coding genes. The fact that some of the differences between isodecoder genes occur in the internal promoter elements suggests that differential control of the expression of some isoacceptor tRNA genes in Trityps is possible. The special characteristics found in Boxes A and B from tRNA-Sec genes from Trityps indicate that the mechanisms that regulate their transcription might be different from those of other tRNA genes.

## Background

The parasites *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*, referred together as Trityps, are trypanosomatid protozoa that cause deadly human diseases known as leishmaniasis, African sleeping sickness and Chagas disease, respectively. Collectively, these pathogens cause millions of deaths in developing countries in tropical and subtropical regions of the world. Analyses of the recently reported genomic sequences of the Trityps revealed a striking feature: their genomes are organized into large directional gene clusters, *i.e.* tens-to-hundreds of protein-coding genes arranged sequentially on the same strand of DNA [1-3]. Transcription of the gene clusters is polycistronic, and mature mRNAs are generated from long precursors by trans-splicing and polyadenylation [4,5]. Most chromosomes contain at least two polycistronic gene clusters (PGCs), which can be either divergently transcribed (towards the telomeres) or convergently transcribed (away from the telomeres). Chromosome 3 from *L. major* contains two convergent PGCs (of 67 and 45 genes) that are separated by a tRNA gene. Interestingly, Pol II-transcription of both PGCs terminates within the tRNA-gene region [6]. The *L. major* nuclear genome is distributed among 36 relatively small chromosomes that range from 0.28 to 2.8 Mb. *T. cruzi* possesses ~28 medium-sized chromosomes, while *T. brucei* has 11 large chromosomes. Regardless of having diverged more than 200 million years ago, the genomes of trypanosomatids show a remarkable conservation of gene order (synteny) [7].

We are interested in the study of transcription by RNA polymerase III (Pol III), which produces small essential RNA molecules, such as tRNA [8]. All tRNAs have sequences of 74–95 bases that fold into a characteristic cloverleaf secondary structure with four constant arms. The acceptor arm binds to a particular amino acid, specified by the anticodon triplet located in the anticodon arm. The anticodon is complementary to an mRNA codon, specific for the amino acid carried by the tRNA. Therefore, tRNAs serve as adaptor molecules that mediate the transfer of information from nucleic acid to protein [9]. Organisms must have at least one tRNA for each of the 20 amino acids. Because different types of relaxed base pairings are allowed at the "wobble" position of the anticodon, certain tRNAs (known as isoacceptors) can read two or more synonymous codons differing by the third base. Consequently, cells do not carry tRNAs with anticodons complementary to all of the 61 possible codons in the genetic code. Interestingly, several organisms contain a large proportion of tRNA genes that have the same anticodon sequence, but differences elsewhere in the tRNA body [10]. The number of these tRNA genes, called isodecoders, varies from very low (10 in yeast) to very high (225–246)

in chimp and human. Thus, the diversity of tRNA genes is higher than originally thought [10].

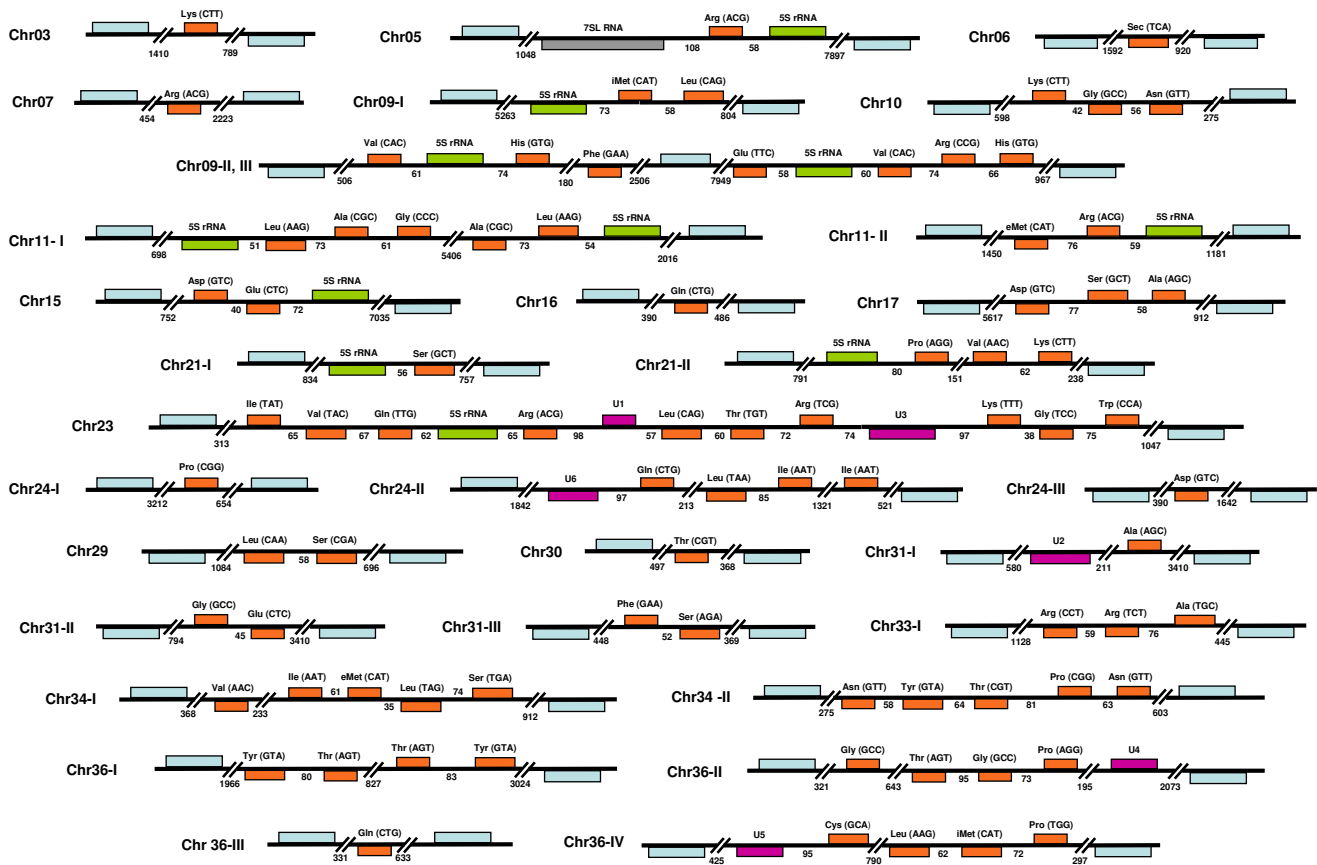
Most organisms usually contain several hundred tRNA genes distributed randomly over their entire genome. One of the distinctive features of most genes transcribed by Pol III is that their promoter sequences are located within the transcribed region. In the case of tRNA genes, the promoter consists of two conserved elements: Boxes A and B. While Box A is normally positioned close to the transcription start site, the location of Box B is variable, partly because some tRNAs have short introns within the coding region [8,11].

Here we report the *in silico* analysis of tRNA genes in trypanosomatids. We found that, unlike in most other organisms, the distribution of genes in the genomes of *L. major* and *T. brucei* does not seem to be totally random, being confined to a subset of chromosomes. In addition, 14 out of 39 convergent strand-switch regions found in *L. major* contain at least one tRNA gene, which suggests that the use of tRNA genes as signals for termination of transcription of PGCs might be a common process in this parasite. Our analysis also indicated that the majority of the tRNA clusters do not show conservation of gene order among Trityps. Analysis of the putative transcription termination signals in all the tRNA genes showed an average of 5 Ts (+/- 1) in *L. major* and *T. brucei*, and 6 Ts (+/- 2) in *T. cruzi*. Also, special features were found in promoter elements from tRNA-Sec genes from Trityps. Finally, we have identified several tRNA isodecoder genes in the Trityps.

## Results and discussion

### Number of tRNA genes

Analysis of the GeneDB databases from *L. major*, *T. brucei* and *T. cruzi* (Trityps) revealed the presence of 82, 65 and 115 tRNA genes, respectively (see Table 1 and Additional File 1). By using the tRNAscan-SE program, we confirmed the identity of all the annotated tRNA genes in *L. major* (Fig. 1). However, we found a few discrepancies in the *T. brucei* and *T. cruzi* annotated genomes. In the case of *T. brucei*, it was observed that the tRNA-Sec (Tb04\_tRNA-Sec1) gene annotated on chromosome 4 actually corresponds to the sRNA76 (see cluster chr04-V in Fig. 2), which is a tRNA-like molecule that associates to the 7SL RNA in trypanosomatids [12]. By doing a search of the *T. brucei* genome with the tRNA-Sec gene sequence reported previously [13], we located two copies of the genuine gene on chromosome 9 (see clusters chr09-II-III in Fig. 2). One of these tRNA genes is located within an open reading frame (ORF), annotated as a "hypothetical protein, unlikely" (Tb09.160.1080). We also identified one tRNA-Sec gene in *L. major* (see cluster chr06 in Fig. 1). Interestingly, eight tRNA-Sec genes were found in the *T. cruzi*

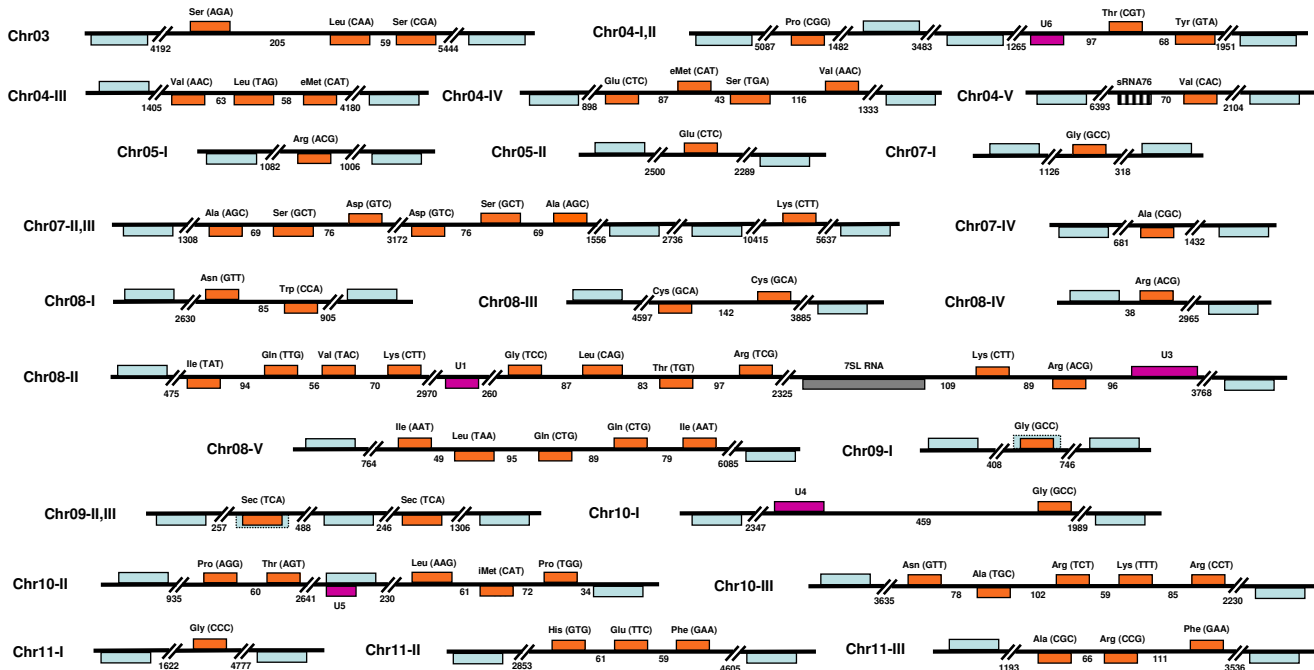


**Figure 1**  
**Organization of tRNA genes in *L. major*.** The 83 tRNA genes found in the genome of the parasite are shown in orange. The predicted anticodons are indicated in parentheses. 5S rRNA, snRNA and 7SL genes are shown in green, purple and gray, respectively. Genes are drawn to scale, and the sizes of intergenic regions are indicated (in base pairs). Protein-coding genes that flank Pol III-transcribed genes are shown in blue (not to scale). The tRNA-Sec gene on chromosome 6 is located at positions 69,586 to 69,673, in the complement strand. Putative pseudogenes are not shown. For practical purposes, we regarded protein-coding genes as the limits of a particular Pol III locus. For that reason, we considered cluster chr09-II, III as two independent Pol III loci. More of such cases are shown in Fig. 2.

genome (see Additional File 1); they all are organized as independent genes, not clustered with other Pol III genes (data not shown). The tRNA-Sec is a component of a translational mechanism that reads UGA (normally a stop codon) as a selenocysteine codon in selected mRNAs that contain a specific *cis*-acting RNA regulatory sequence in their 3' untranslated regions (3'-UTRs) [14]. The presence of selenoproteins, and all the machinery required for its synthesis, has been demonstrated in *L. major* and *T. cruzi* [13,15].

In *T. cruzi*, it was found that three tRNA genes annotated as Val-CAC (Tc00.1047053457717.10, Tc00.1047053483321.10 and Tc00.1047053506321.220) do not seem to correspond to the assigned amino acid (or to any other).

They showed only 61% identity with Tc00.1047053506459.249, which we consider is the "real" tRNA-Val-CAC gene, since it is 100% identical to the tRNA-Val-CAC gene from *T. brucei* (and shows 98% identity to its orthologue in *L. major*). Interestingly, we observed that they show 75% identity to the sRNA76 from *T. brucei* (data not shown), which suggests that they might actually encode the orthologue of this gene in *T. cruzi*. Alternatively, they may correspond to tRNAs with undetermined or unknown type. tRNA genes with undetermined type have been found in several species, including *Caenorhabditis elegans*, yeast and human, and their function is unknown. Additionally, Tc00.1047053507579.16, annotated as an Ile-TAT gene, seems to be a pseudogene (or an undetermined tRNA); since it shows only 22%



**Figure 2**  
**Organization of tRNA genes in *T. brucei*.** The 66 tRNA genes, distributed into 23 loci, are indicated in orange. The predicted anticodons are shown in parentheses. snRNA and 7SL genes are shown in purple and gray, respectively. The sRNA76 (misannotated in GeneDB as a tRNA-Sec gene) is shown as a striped box in locus chr04-V. The two tRNA-Sec genes that we found are located in loci chr09-II, III (at positions 505,606 to 505,693 and 509,443 to 509,530 of chromosome 9). The gene from locus chr09-II overlaps a putative Pol II gene (Tb09.160.1080, dotted blue box), annotated as hypothetical protein (unlikely). Also, the tRNA-Gly from locus chr09-I overlaps a putative Pol II gene (Tb09.211.4080, dotted blue box), annotated as hypothetical protein (unlikely). The U5 snRNA gene (locus chr10-II) also overlaps a Pol II gene (Tb10.6k15.2990, sequence orphan), but located on the opposite strand. Genes are drawn to scale, and the sizes of intergenic regions are indicated (in base pairs). Protein-coding genes that flank Pol III-transcribed genes are shown in blue (not to scale). In locus chr08-V, the Leu-TAA gene (Tb\_08\_TRNA\_Leu\_2) and the first Gln-CTG gene (Tb\_08\_TRNA\_Gln\_2) are annotated in the wrong (opposite) strand in GeneDB. Maps of an incomplete repertoire of tRNA genes from *T. brucei* were previously reported [16].

identity to the other two annotated Ile-TAT genes in *T. cruzi* (Tc00.1047053504427.231 and Tc00.1047053508043.11) (data not shown). Also, we identified an extra copy of tRNA-Ala-TGC on contig 8001 (see Additional File 1).

Thus, our analysis indicates the presence of 120 tRNA genes in *T. cruzi*, excluding four genes that might be undetermined tRNAs or encode orthologues of the sRNA76, and including the eight tRNA-Sec genes and the tRNA-Ala-TGC gene (Table 1 and Additional File 1). In *T. brucei* the number of identified tRNA genes is 66, including the two newly identified tRNA-Sec genes and excluding the gene of the sRNA76 orthologue (Fig. 2 and Table 1). In *L. major* there are 83 tRNA genes (Fig. 1 and Table 1), in addition to a pseudogene that we do not include in our analyses. The number of tRNA genes in trypanosomatids is relatively low, considering that eukaryotic organisms usually contain several hundred tRNA genes. For instance, *C. ele-*

*gans* has 568 tRNA genes, *Homo sapiens* presents 497 tRNA genes and *Saccharomyces cerevisiae* contain 271 tRNA genes [9,10]. In an extreme case, *Danio rerio* (zebra fish) has ~6000 predicted tRNA genes. On the other hand, the microsporidian parasite *Encephalitozoon cuniculi* has only 44 tRNA genes. Bacterial genomes usually have between 29 and 167 tRNA genes in their genomes [9,10].

tRNA genes from eukaryotes typically contain introns, which are usually located between bases 37 and 38 of the anticodon loop. Archaeal tRNA genes also have introns that can be found at the same location of the anticodon loop or in other regions of the tRNA gene. The size of introns is variable, ranging from 7 to 121 bases [9]. In bacterial genomes, a very small number of tRNA-gene introns have been reported, but they correspond to self-splicing introns (group I autocatalytic introns). Analysis of the tRNA genes in *Trityps* indicated that only the tRNA-Tyr genes contain an intron; which was previously reported in

**Table 1: Repertoire of tRNA genes and codon usage in Trityps.**

| Amino acid | Codon | tDNA anticodon | <i>L. major</i> gene copy number | <i>T. brucei</i> gene copy number | <i>T. cruzi</i> gene copy number | Anticodon <sup>2</sup> | <i>L. major</i> Codon usage % | <i>T. brucei</i> Codon usage % | <i>T. cruzi</i> Codon usage % |
|------------|-------|----------------|----------------------------------|-----------------------------------|----------------------------------|------------------------|-------------------------------|--------------------------------|-------------------------------|
| <b>Ala</b> | GCT   | AGC            | 2                                | 2                                 | 2 (98%) <sup>1</sup>             | <b>IGC</b>             | 1.83                          | 2.08                           | 1.61                          |
|            | GCC   | GGC            | 0                                | 0                                 | 0                                |                        | 3.61                          | 1.82                           | 1.62                          |
|            | GCA   | TGC            | 1                                | 1                                 | 2 (97%) <sup>1</sup>             | UGC                    | 2.02                          | 2.32                           | 2.16                          |
|            | GCG   | CGC            | 2                                | 2 (97%) <sup>1</sup>              | 2                                | CGC                    | 3.62                          | 2.07                           | 2.36                          |
| <b>Arg</b> | CGT   | ACG            | 4                                | 3                                 | 4 (98%) <sup>1</sup>             | <b>ICG</b>             | 1.05                          | 1.59                           | 1.60                          |
|            | CGC   | GCG            | 0                                | 0                                 | 0                                |                        | 3.18                          | 1.45                           | 1.75                          |
|            | CGA   | TCG            | 1                                | 1                                 | 2                                | UCG                    | 0.73                          | 0.9                            | 1.58                          |
|            | CGG   | CCG            | 1                                | 1                                 | 2                                | CCG                    | 1.38                          | 1.21                           | 1.85                          |
|            | AGA   | TCT            | 1                                | 1                                 | 2                                | UCU                    | 0.27                          | 0.68                           | 1.56                          |
|            | AGG   | CCT            | 1                                | 1                                 | 2                                | CCU                    | 0.55                          | 1                              | 1.86                          |
| <b>Asn</b> | AAT   | ATT            | 0                                | 0                                 | 0                                | <b>GUU</b>             | 0.53                          | 1.82                           | 1.37                          |
|            | AAC   | GTT            | 3                                | 2                                 | 4                                |                        | 2.07                          | 1.92                           | 1.45                          |
| <b>Asp</b> | GAC   | GTC            | 3                                | 2                                 | 2                                | <b>GUC</b>             | 3.42                          | 2.27                           | 1.60                          |
|            | GAT   | ATC            | 0                                | 0                                 | 0                                |                        | 1.44                          | 2.81                           | 1.57                          |
| <b>Cys</b> | TGT   | ACA            | 0                                | 0                                 | 0                                | <b>GCA</b>             | 0.42                          | 1.09                           | 1.78                          |
|            | TGC   | GCA            | 1                                | 2                                 | 2                                |                        | 1.46                          | 1.13                           | 2.34                          |
| <b>Gln</b> | CAA   | TTG            | 1                                | 1                                 | 2                                | UUG                    | 0.76                          | 1.69                           | 1.66                          |
|            | CAG   | CTG            | 3                                | 2                                 | 4                                | CUG                    | 3.32                          | 2.1                            | 1.8                           |
| <b>Glu</b> | GAA   | TTC            | 1                                | 1                                 | 1                                | UUC                    | 1.14                          | 3.17                           | 2.17                          |
|            | GAG   | CTC            | 2                                | 2                                 | 4 (95%) <sup>1</sup>             | CUC                    | 3.32                          | 3.81                           | 2.21                          |
| <b>Gly</b> | GGT   | ACC            | 0                                | 0                                 | 0                                | <b>GCC</b>             | 1.25                          | 2.27                           | 1.51                          |
|            | GGC   | GCC            | 4                                | 3                                 | 4                                |                        | 3.36                          | 1.49                           | 1.95                          |
|            | GGA   | TCC            | 1                                | 1                                 | 2                                | UCC                    | 0.64                          | 1.56                           | 2.16                          |
|            | GGG   | CCC            | 1                                | 1                                 | 2                                | CCC                    | 1.19                          | 1.39                           | 1.64                          |
| <b>His</b> | CAT   | ATG            | 0                                | 0                                 | 0                                | <b>GUG</b>             | 0.65                          | 1.13                           | 1.37                          |
|            | CAC   | GTG            | 2                                | 1                                 | 4 (98%) <sup>1</sup>             |                        | 2.04                          | 1.3                            | 1.64                          |
| <b>Ile</b> | ATT   | AAT            | 3                                | 2                                 | 4                                | <b>IAU</b>             | 0.82                          | 1.9                            | 1.42                          |
|            | ATC   | GAT            | 0                                | 0                                 | 0                                |                        | 1.88                          | 1.16                           | 1.10                          |
|            | ATA   | TAT            | 1                                | 1                                 | 2                                | UAU                    | 0.27                          | 1                              | 0.75                          |
| <b>Leu</b> | TTA   | TAA            | 1                                | 1                                 | 2                                | UAA                    | 0.16                          | 0.98                           | 0.89                          |
|            | TTG   | CAA            | 1                                | 1                                 | 2                                | CAA                    | 1.1                           | 1.96                           | 2.33                          |
|            | CTT   | AAG            | 3                                | 1                                 | 4                                | <b>IAG</b>             | 1.11                          | 2.23                           | 1.59                          |
|            | CTC   | GAG            | 0                                | 0                                 | 0                                |                        | 2.5                           | 1.56                           | 1.27                          |
|            | CTA   | TAG            | 1                                | 1                                 | 2                                | UAG                    | 0.47                          | 0.82                           | 0.54                          |
|            | CTG   | CAG            | 2                                | 1                                 | 2                                | CAG                    | 3.83                          | 1.85                           | 1.95                          |
| <b>Lys</b> | AAA   | TTT            | 1                                | 1                                 | 2                                | UUU                    | 0.54                          | 2.06                           | 1.94                          |
|            | AAG   | CTT            | 3                                | 3                                 | 4                                | CUU                    | 2.78                          | 2.66                           | 1.88                          |
| <b>Met</b> | ATG   | CAT            | 4                                | 3                                 | 6                                | CAU                    | 2.25                          | 2.34                           | 2.12                          |
| <b>Phe</b> | TTT   | AAA            | 0                                | 0                                 | 0                                | <b>GAA</b>             | 1.04                          | 2.05                           | 2.17                          |
|            | TTC   | GAA            | 2                                | 2                                 | 4                                |                        | 1.91                          | 1.59                           | 1.49                          |
| <b>Pro</b> | CCT   | AGG            | 2                                | 1                                 | 2                                | <b>IGG</b>             | 0.86                          | 1.11                           | 1.05                          |
|            | CCC   | GGG            | 0                                | 0                                 | 0                                |                        | 1.24                          | 1.11                           | 1.03                          |

**Table 1: Repertoire of tRNA genes and codon usage in Trityps. (Continued)**

|              |     |     |                      |    |                      |            |      |      |      |
|--------------|-----|-----|----------------------|----|----------------------|------------|------|------|------|
|              | CCA | TGG | 1                    | 1  | 2                    | UGG        | 1.05 | 1.39 | 1.44 |
|              | CCG | CGG | 2                    | 1  | 2                    | CGG        | 2.61 | 1.18 | 1.57 |
| <b>Ser</b>   | TCT | AGA | 1                    | 1  | 2                    | <u>IGA</u> | 1.02 | 1.26 | 1.37 |
|              | TCC | GGA | 0                    | 0  | 0                    |            | 1.69 | 1.25 | 1.24 |
|              | TCA | TGA | 1                    | 1  | 2 (98%) <sup>1</sup> | UGA        | 0.73 | 1.35 | 1.41 |
|              | TCG | CGA | 1                    | 1  | 2                    | CGA        | 2.17 | 1.16 | 1.2  |
|              | AGT | ACT | 0                    | 0  | 0                    | <u>GCU</u> | 0.73 | 1.51 | 1.21 |
|              | AGC | GCT | 2 (98%) <sup>1</sup> | 2  | 2                    |            | 2.6  | 1.35 | 1.60 |
| <b>Thr</b>   | ACT | AGT | 3                    | 1  | 2                    | <u>IGU</u> | 0.68 | 1.3  | 1.13 |
|              | ACC | GGT | 0                    | 0  | 0                    |            | 1.77 | 1.21 | 1.18 |
|              | ACA | TGT | 1                    | 1  | 2                    | UGU        | 1.04 | 1.74 | 1.74 |
|              | ACG | CGT | 2                    | 1  | 2                    | CGU        | 2.48 | 1.48 | 1.98 |
| <b>Trp</b>   | TGG | CCA | 1                    | 1  | 2                    | CCA        | 1.07 | 1.09 | 2.33 |
| <b>Tyr</b>   | TAT | ATA | 0                    | 0  | 0                    | <u>GUA</u> | 0.4  | 1.13 | 0.84 |
|              | TAC | GTA | 3                    | 1  | 2                    |            | 1.99 | 1.41 | 1.04 |
| <b>Val</b>   | GTT | AAC | 2                    | 2  | 2                    | <u>IAC</u> | 0.86 | 2.29 | 1.49 |
|              | GTC | GAC | 0                    | 0  | 0                    |            | 1.92 | 1.14 | 1.26 |
|              | GTA | TAC | 1                    | 1  | 2                    | UAC        | 0.54 | 1.26 | 0.83 |
|              | GTG | CAC | 2                    | 1  | 1                    | CAC        | 3.82 | 2.87 | 2.65 |
| <b>SeC</b>   | TGA | TCA | 1                    | 2  | 8                    |            |      |      |      |
| <b>TOTAL</b> |     |     | 83                   | 66 | 120                  |            |      |      |      |

<sup>1</sup> One isodecoder tRNA was identified in *L. major*, one in *T. brucei* and six in *T. cruzi*. The percentage of identity between isodecoders is shown between parentheses. The rest of the genes with the same anticodon are 100% identical.

<sup>2</sup> Underlined anticodons are those that recognize more than one codon; for example, the Tyr anticodon GUA recognizes two codons, TAT and TAC, since anticodon Tyr-AUA is missing in trypanosomatids. A1 of the anticodon is modified to inosine (I).

*T. brucei* [16]. The intron is 11 bases long in *L. major* and *T. brucei*, and 13 bases long in *T. cruzi*; and as in other organisms, it is located between bases 37 and 38 (data not shown). Thus, introns are very rare in Trityps, not only in protein-coding genes, but also in tRNA genes.

**Isoacceptor tRNA species**

Analyses of the anticodon sequences of the tRNA genes in Trityps showed the presence of 46 isoacceptor types in each of the three species (Table 1) [17]. These 46 isoacceptor types are able to read the 61 codons that specify the canonical amino acids, in addition to Sec, the 21st amino acid. The number of isoacceptor species found in Trityps is similar to that found in other organisms (from 41 to 55 isoacceptors) [10]. It is important to mention that the two methionine isoacceptors, the initiator and elongator, have been identified (see below), but for practical purposes these two isoacceptors will be considered as one.

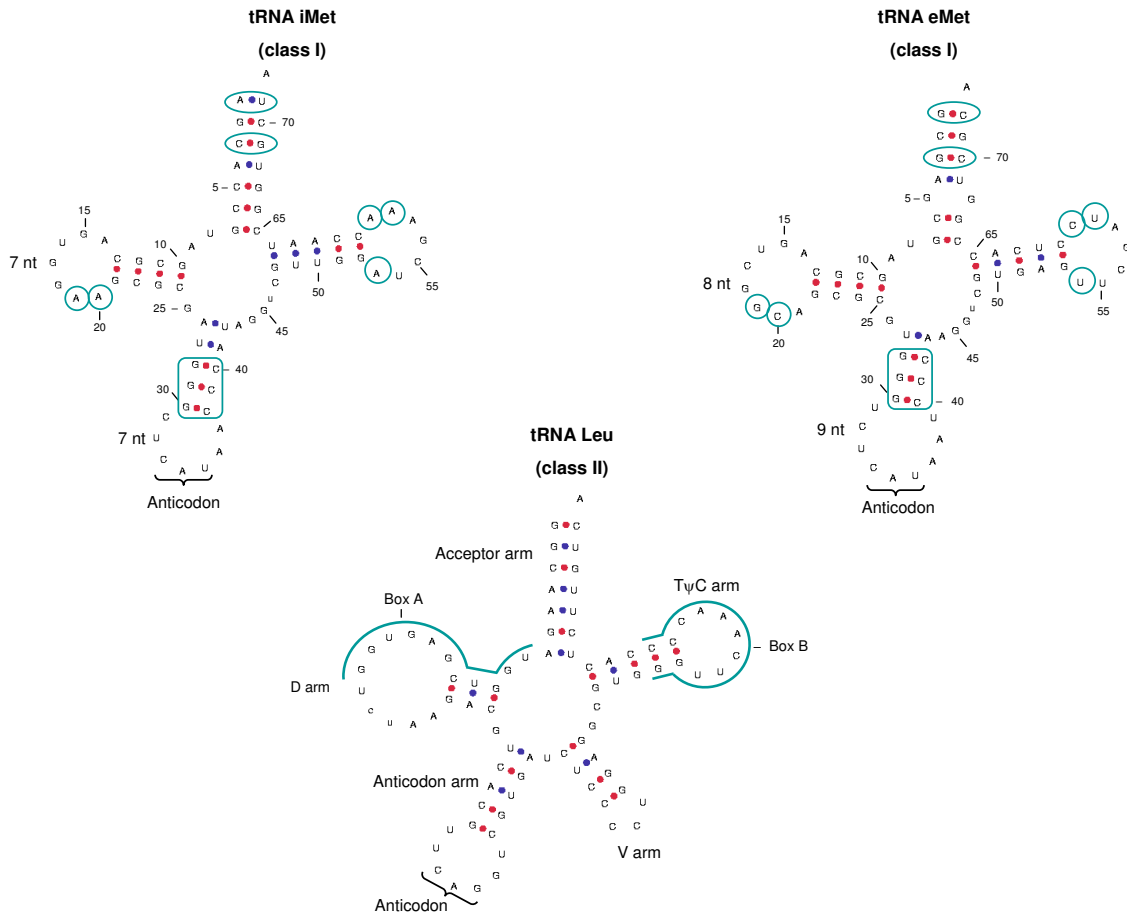
Sixteen anticodons were not found in the tRNA genes of trypanosomatids, even though their corresponding codons are present in the protein-coding genes of these organisms [17]. For example, the tRNA with anticodon Ile-GAU is not present in the genome of trypanosomatids,

but the codon AUC is present in their protein sequences (Table 1). As mention above, this is possible because some tRNAs are able to recognize more than one codon by allowing flexible base-pairing between the first nucleotide of the anticodon and the third position of the codon (tRNA wobble recognition). Analysis of the data shown in Table 1 indicates that C3 or U3 in the codon are recognized by G1 or A1 of the anticodon (A is converted to the nucleotide inosine in the mature tRNA, which can pair with U3 or C3 in the codon). Thus, trypanosomatids use the A1- or G1-sparing strategy as a decoding mode [9]. This anticodon-choice pattern is similar to that of other eukaryotes such as *C. elegans*, *H. sapiens* and *A. thaliana*. Other eukaryotic organisms, like yeast and *D. melanogaster*, use the A1- or G1 and C1-sparing strategy [9]. As observed previously [17], the spared anticodons are used equally, since 50% (8/16) of the U3 and C3 codons are read by A1 (or I1) and the remaining 50% are read by G1. In most four-fold degenerate codon families (*i.e.* Leu, Val, Ser, Pro, Thr, Ala and Arg, but not Gly) A1 reads the codons containing U3 and C3, since the corresponding tRNAs with G1 are not present in the Trityps. We found the same for the Ile family, which contains three codons. On the other hand, all the two-fold degenerate families

use G1 to read U3 or C3, given that the tRNAs with A1 are missing. The families that use this strategy are: Asn, Asp, Cys, Gly (although it is fourfold-degenerated), His, Phe, Ser (AGU and AGC codons) and Tyr.

The genomes of *L. major*, *T. brucei* and *T. cruzi* contain four, three and six tRNA-Met genes, respectively [16,17]. Further analysis of these genes indicated that in *L. major* two of them (LmjF09.TRNAMET.01 and LmjF36.TRNAMET.01) correspond to initiator tRNAs (iMet) and two (LmjF11.TRNAMET.01 and LmjF34.TRNA.01) correspond to the elongator form (eMet) (Fig. 3). In *T. brucei*, one iMet (Tb10\_tRNA\_Met\_1) and two eMet genes (Tb04\_tRNA\_Met\_1, Tb04\_tRNA\_Met\_2) were found, although only one of each type was previously reported [16]. Finally, two iMet (Tc00.1047053508231.92, Tc.00.1047053506251.88) and four eMet genes

(Tc00.1047053504055.87, Tc00.1047053504055.91, Tc00.1047053506435.327 and Tc00.1047053506435.345) were located in *T. cruzi*. As shown in Fig. 3, the two types of tRNA-Met possess specific features, and most of them were found in the genes from the Trityps. One of the main characteristics is the highly conserved A:T base pair that is present in all tRNA-iMet in eukaryotes at position A1:U72 (A1:U71 in Trityps), whereas a G:C pair is found in tRNA-eMet [9]. In yeast, it has been reported that the A1:U72 base pair is the most important determinant for a tRNA-Met to play the role of iMet, since it is necessary for binding to Initiation Factor 2 (eIF2) [18]. When this sequence is mutated, the tRNA-iMet is able to bind to Elongation Factor Tu (EF-Tu) and participates in translation elongation. It is likely that this base pair has a similar function in trypanosomatids.



**Figure 3**  
**Secondary structures of initiator and elongator tRNA-Met in *L. major*.** Important features for iMet and eMet tRNAs function are indicated. These two molecules correspond to class I tRNAs. The secondary structure of a class II tRNA (Leu) is also presented. The names of the different arms are shown. The position of the internal control elements (Boxes A and B) is also indicated.

Other determinants for iMet function are: A53, A58 and A59 in the T $\psi$ C loop (U54, U59 and C60 in eMet); the base pair C3:G69 (G3:C70 in eMet); bases A19 and A20 (G19 and C20 in eMet); a D-loop composed of 7 nucleotides (8 nt in eMet); and an anticodon loop of 7 nt (9 nt in eMet) [9,19]. All these features are conserved in the Tritryps (Fig. 3). Another distinctive characteristic of the iMet tRNA is the presence of three consecutive G:C pairs at the bottom of the anticodon stem, which is conserved not only in eukaryotes, but also in eubacteria and archaeobacteria. In *E. coli*, it has been shown that mutations in these three consecutive G bases reduce the efficiency of initiation of protein synthesis, by affecting the interactions between the tRNA and the ribosomal P site; thus, these bases are essential to discriminate between initiator and elongator tRNA-Met [9,20]. Tritryps iMet tRNAs have these conserved G:C pairs but, surprisingly, we found them in the eMet tRNAs as well (Fig. 3). Thus, in the Tritryps these base pairs are not a discriminator between iMet and eMet tRNAs, and these organisms must use other features of the iMet to direct it to the P site of the ribosome.

In several organisms it has been observed that there is a correlation between tRNA gene copy number and codon usage [21,22]. Apparently, selection on synonymous codon positions causes co-adaptation of codon usage and tRNA content, in order to optimize the effectiveness of protein synthesis [23]. In Tritryps, it has been reported that bias in codon usage correlates with tRNA gene copy number and with protein expression level [17]. This conclusion was made after analyzing around 60,000 codons from highly expressed (tandem duplicated) protein-coding genes from the three parasites. We conducted a similar analysis, but including all the 8272 protein coding genes from *L. major* (5,249,748 codons), and 5119 randomly selected genes from *T. brucei* (2,620,035 codons), as well as 1779 genes from *T. cruzi* (986,435 codons). We plotted codon usage (see Table 1) against the number of tRNA genes for each isoacceptor, for the three species, and a possible correlation was evaluated by the Spearman test (Figure 4, panels A-C). The data indicated a positive correlation between these variables for *L. major* ( $r_s = 0.80$ ) and *T. brucei* ( $r_s = 0.64$ ), which indicates that, similar to other organisms, codon usage patterns seem to be co-adapted with the relative abundance of the corresponding tRNAs in these parasites. However, in the case of *T. cruzi*, the obtained Spearman value ( $r_s = 0.35$ ) indicated a low degree of correlation between the two variables. This may reflect the fact that the *T. cruzi* strain used for the sequencing project is a hybrid of two strains, and some of their genes might be duplicated, while others might not be; as shown in Fig. 4C and Table 1, in *T. cruzi* the vast majority of the isoacceptor species are encoded by either two or four genes (only two isoacceptors have one gene, and none of them has three genes), whereas in *L. major* and *T. brucei*

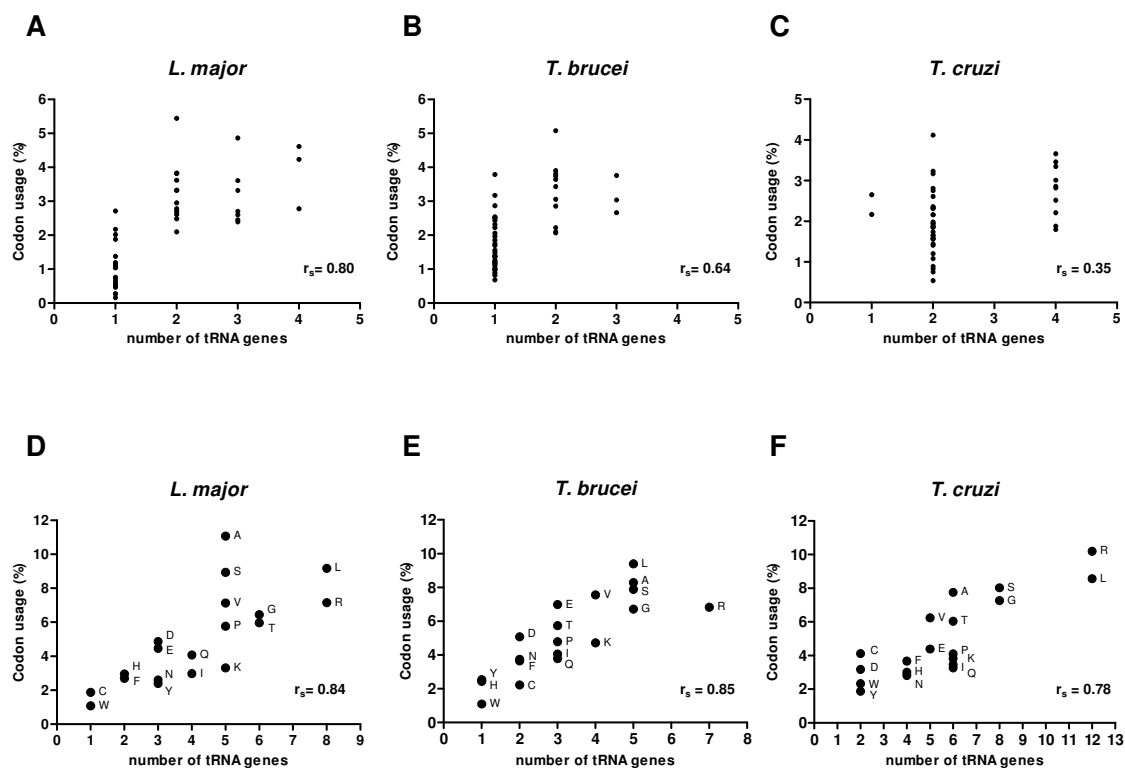
a high number of the isoacceptors are encoded by a single gene. The correlation analysis was repeated, but now plotting the percentage of codon usage *versus* the number of tRNA genes per amino acid (Fig. 4, panels D-F). This time, the Spearman value was high in *T. cruzi* ( $r_s = 0.78$ ), indicating a strong correlation between both parameters. As before, strong correlations were found in *L. major* ( $r_s = 0.84$ ) and *T. brucei* ( $r_s = 0.85$ ).

#### Organization of tRNA genes

In *L. major*, the 83 tRNA genes are distributed among 31 loci, on 19 different chromosomes (Fig. 1 and Additional File 1). Most tRNA genes are organized into clusters of 2 to 10 genes, on either top or bottom strand, which may contain other Pol III-transcribed genes. For example, in the locus located on chromosome 23 (chr23 in Fig. 1) there are 10 tRNA genes, a 5S rRNA gene and the U1 and U3 snRNA genes. Locus IV on chromosome 36 (chr36-IV) has four tRNA genes and the U5 snRNA gene. The eleven 5S rRNA genes found in the *L. major* genome are distributed in six chromosomes, and are always associated to tRNA genes (Fig. 1 and Additional File 1). Only eight loci contain single tRNA genes (chr03, chr06, chr07, chr16, chr24-I, chr24-III, chr30 and chr36-III). In most cases, intergenic regions that separate Pol III-transcribed genes are short, with an average size of 202 bases (Fig. 1). However, they can be as small as 35 bases (intergenic region between tRNA-eMet and tRNA-Leu on cluster chr34-I) and as long as 5406 bases (intergenic region between tRNA-Gly and tRNA-Ala on cluster chr11-I). Intergenic regions between the protein-coding genes that flank the clusters and the first or last Pol III gene of the cluster are normally longer than the ones that separate Pol III genes. The average length of such regions is 1490 bp, with a minimum of 238 bp (intergenic region between tRNA-Lys and the "right" Pol II gene on cluster chr21-II) and a maximum of 7949 pb (intergenic region between tRNA-Glu and the "left" Pol II gene on cluster chr09-III) (Fig. 1). The mean length of intergenic regions between protein-coding genes in the *L. major* genome is 2045 bp [1].

In *T. brucei*, the 66 tRNA genes are located on 26 loci, on eight different chromosomes (Fig. 2 and Additional File 1). As in *L. major*, in *T. brucei* the number of tRNA genes per cluster ranges from 2 to 10. Eleven of the 66 tRNA genes are single genes in *T. brucei* (loci chr04-I, chr05-I, chr05-II, chr07-I, chr07-III, chr07-IV, chr08-IV, chr09-I, chr09-II, chr09-III and chr11-I). Similarly to *L. major*, intergenic regions that separate Pol III genes in *T. brucei* are short in most cases (average length is 327 bp, ranging from 43 to 3172 bp) (Fig. 2). Regarding intergenic regions between the protein-coding genes that flank the clusters and the first or last Pol III gene of the cluster, the mean size is 2473 bp (showing a range from 34 to 10415 bp). In the *T. brucei* genome the average length of intergenic





**Figure 4**

**Correlation between number of tRNA genes and codon usage in *Trityps*.** Panels A-C show the correspondence between codon usage and the number of tRNA genes for each isoacceptor (45 types), for the three species (see Table 1). Panels D-F show the correlation between codon usage and the number of tRNA genes per amino acid. Correlation coefficients were evaluated by performing a Spearman test.  $r_s$  values are indicated in each panel;  $p < 0.0001$  for panels A, B, D and E;  $p < 0.02$  for panel C;  $p < 0.0001$  for panel F.

regions between protein-coding genes is 1279 bp [2]. Since it has not been possible to assemble fully adjacent sequences for the chromosomes of *T. cruzi*, at the present we are unable to determine the genomic organization of the tRNA genes in this parasite. In contrast to *L. major*, the 5S rRNA genes in *T. brucei* and *T. cruzi* are organized into tandem arrays, which are not associated to tRNA genes [24,25].

In most eukaryotic organisms, tRNA genes seem to be dispersed randomly throughout the genome. However, in human cells the distribution is non-random, since more than 25% of the tRNA genes are located in a region of only about 4 Mb on chromosome 6. This region represents only 0.1% of the human genome, but contains an almost complete set of tRNA genes. Moreover, 280 out of 497 tRNA genes (more than half) are found on either chromo-

some 1 or chromosome 6 [26]. The distribution of genes in the *L. major* genome does not seem to be totally random, since half of the chromosomes do not contain even a single tRNA gene. Additionally, 60 tRNA genes (72%) are located in only 7 chromosomes (9, 11, 23, 24, 31, 34 and 36), which represent only 26% of the genome (Fig. 1 and Additional File 1). In *T. brucei*, 40 (61.5%) of the tRNA genes are found in just 3 chromosomes (4, 7 and 8), which is only about 24% of the genome (Fig. 2 and Additional File 1). tRNA genes in *S. cerevisiae*, though dispersed in the linear genome, co-localize with 5S rDNA at the nucleolus. Nucleolar localization requires tRNA gene transcription, because inactivation of the internal promoter eliminates its nucleolar location [27]. It remains to be tested whether tRNA genes in trypanosomatids show such a specific cellular localization.

In *Schizosaccharomyces pombe* and *C. elegans*, tRNA genes are often clustered in centromeres [28]. These tRNA genes contribute to centromere function by defining domain boundaries important for centromere activity [29]. Putative centromeric regions have been reported in a few chromosomes in *T. cruzi* and *T. brucei* [30], where they localize to strand-switch regions that separate divergent PGCs. While these regions do not seem to contain tRNA genes, two clusters of tRNA genes in *L. major* (chr09-II and chr10), and one cluster in *T. brucei* (chr04-I) are located in divergent strand-switch regions (Figs. 1 and 2), and thus might be candidates to contain centromeric regions. Therefore, it is possible that in trypanosomatids, like *S. pombe* and *C. elegans*, some tRNA genes might be important for centromeric activity.

#### **Spatial relation between Pol III and Pol II genes**

We have previously shown that transcription of two convergent PGCs on *L. major* chromosome 3 terminates on the convergent strand-switch area, within the tRNA-gene region [6]. Interestingly, 14 of the 39 convergent strand-switch regions (35.9%) in the *L. major* genome contain at least one tRNA gene (Fig. 1), representing 45.2% of the 31 tRNA loci. A similar situation was found in *T. brucei*, where 34.6% of the tRNA loci are located within convergent strand-switch regions (Fig. 2). This suggests that the use of tRNA genes as signals for termination of transcription of convergent clusters of protein-coding genes might be a common process in trypanosomatids. Indeed, recent evidence suggest that this is the case for tRNA clusters located within PGCs, since peaks of acetylated histone H3 are found immediately downstream of the tRNA cluster in all cases [31]. Acetylated histones are markers for open chromatin in all eukaryotes and have been found at the 5' end of all polycistronic gene clusters in *L. major*.

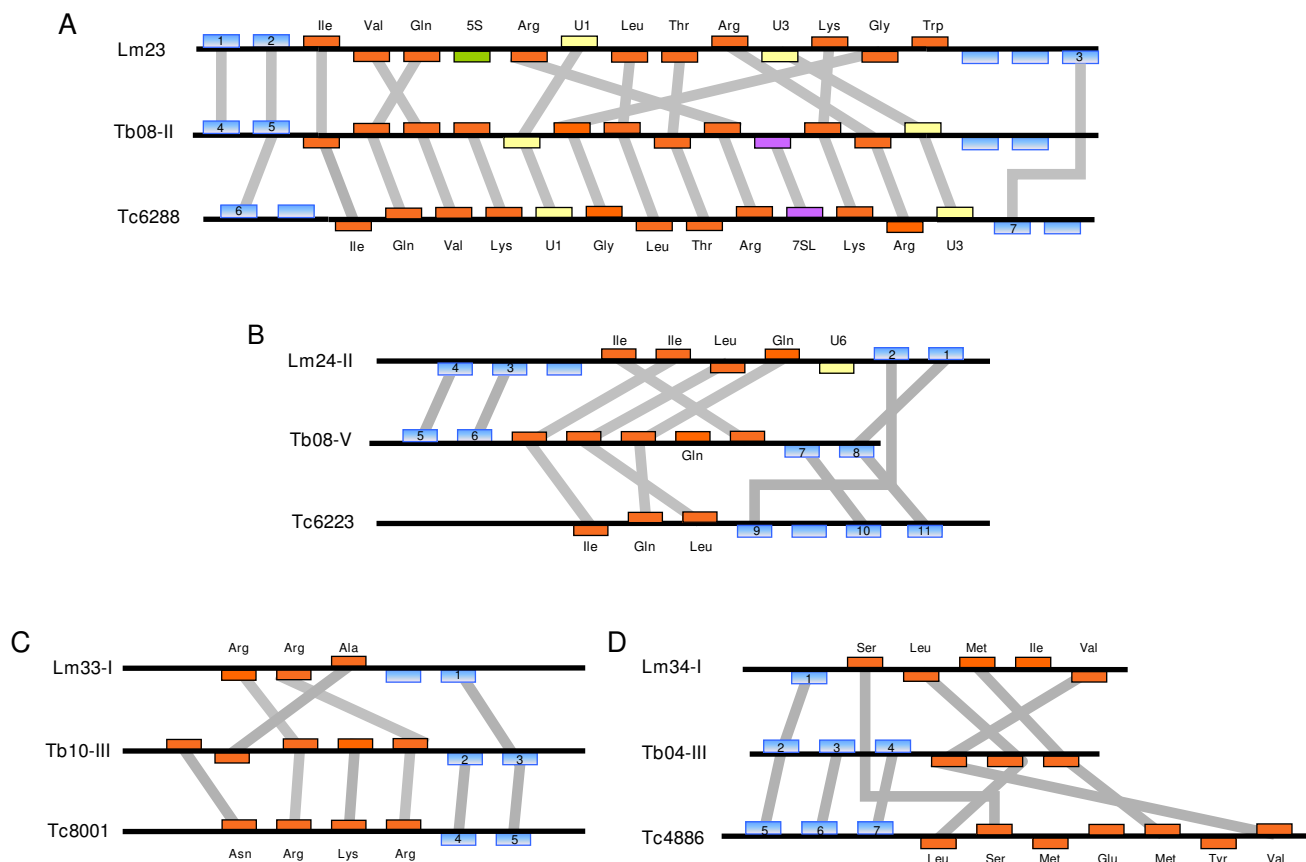
#### **Synteny of Pol III genes**

It has been found that the genomes of the Trityps are highly syntenic, that is to say, they show conservation of gene order, with the *T. brucei* and *L. major* genomes containing 110 blocks of syntenicity spanning 19.9 and 30.7 Mb, respectively [7]. Many of these syntenicity blocks correspond to intact PGCs, which are transcribed by Pol II. In contrast, the majority of the tRNA clusters do not show syntenicity, but a few of them do show conservation (Fig. 5). Among the latter, the most remarkable example is a cluster of 13 Pol III genes that is highly syntenic; corresponding to chr23 in *L. major*, chr08-II from *T. brucei*, and the cluster located on contig Tc6288 from *T. cruzi* (Fig. 5A). Surprisingly, the order of the genes in this cluster is identical between *T. brucei* and *T. cruzi*, although the U1 snRNA, the 7SL RNA and the tRNA-Leu genes are located on different strands. Most of the 13 Pol III-transcribed genes are present in the *L. major* cluster, but their order is not identical to either of the other two clusters (Fig. 5A). Additionally, a 5S rRNA

gene replaced a 7SL RNA gene and a tRNA-Trp gene replaced one of the tRNA-Lys genes. Another Pol III-gene cluster that exhibits syntenicity is chr24-II in *L. major*, chr08-V in *T. brucei* and contig Tc6223 from *T. cruzi* (Fig. 5B). Here, we found tRNA genes for Ile, Leu and Gln that are syntenic among the three species. A second copy of a tRNA-Ile is conserved between *L. major* and *T. brucei*. One difference is that the *L. major* cluster contains an U6 snRNA gene that is replaced by a tRNA-Gln gene in *T. brucei*. Other syntenic tRNA clusters are chr33-I from *L. major*, chr10-III from *T. brucei* and contig Tc8001 from *T. cruzi* (Fig. 5C), and chr34-I from *L. major*, chr04-III from *T. brucei* and contig Tc4886 from *T. cruzi* (Fig. 5D). Several of the protein-coding genes that flank these four syntenic tRNA-gene clusters are also syntenic among Trityps (Fig. 5). As in Trityps, an overall weak syntenicity of Pol III-transcribed genes has been observed between two species of the oomycete *Phytophthora* [32], indicating a reduced importance of genome location of Pol III genes compared to protein-coding genes.

#### **Consensus sequences of promoter elements**

One of the distinctive features of most genes transcribed by Pol III is that their promoter sequences are internal, located within the transcribed region. In the case of tRNA genes, the promoter consists of two conserved elements: Boxes A and B [33]. While Box A is normally positioned close to the transcription start site, the location of Box B is variable, partly because some tRNAs have short introns within the coding region. In contrast, tRNA genes in prokaryotic cells contain promoter elements similar to those found in protein-coding genes: the start-point (usually a purine), the -10 sequence (the TATA Box) and the -35 sequence (the hexamer) [34]. Consensus sequences of trypanosomatid tRNA promoter elements (Fig. 6A) were determined by analyzing the sequences of all tRNA genes in *L. major*, *T. brucei* and *T. cruzi* and comparing them to the sequences of Boxes A and B from *S. cerevisiae* [9]. The tRNAs were divided into two classes, depending on the size of the variable loop (Fig. 3). Class I tRNAs have a short variable loop of 4 or 5 nucleotides, whereas class II tRNAs possess a long variable arm, with a double helical stem of 3 to 7 base pairs and a loop of 3 to 5 nucleotides (Fig. 3) [9]. In the Trityps, 43 genes belong to class II (all Leu and Ser tRNA genes, but excluding tRNA-Sec genes), and 215 genes are class I. Since we observed sequence differences between class I and class II tRNA genes, we analyzed them separately. Half of the bases from the consensus sequence of Box A (positions 1, 2, 4, 7, 10 and 11) are identical between class I and class II genes (and identical to the *S. cerevisiae* consensus sequence) (Fig. 6A). However, position 5 is different between both classes, since class I genes may have any nucleotide, whereas class II genes always have a C. Also, class I tRNA genes present C or T at position 6, while class II genes always have A or



**Figure 5**  
**Comparative order of Pol III-transcribed genes in Trityps.** Clusters that present some degree of synteny among Trityps are shown. In *L. major*, they correspond to loci chr23 (panel A), chr24-II (panel B), chr33-I (panel C) and chr34-I (panel D). The order of genes in clusters chr24-II and chr34-I was inverted compared to the maps shown in figure 1. The corresponding orthologous regions in *T. brucei* and *T. cruzi* are indicated. Orthologous genes are joined by grey lines. Figure is not to scale.

G. Regarding Box B, position 8 is different between both classes: an A is always present in class II genes, while class I genes may have any nucleotide. Around 20% of the class I tRNA genes in Trityps have an additional nucleotide in Box A, between positions 9 and 10 (marked with an asterisk in Fig. 6A). These tRNA genes are: eMet-CAT, Asn-GTT, Ile-AAT, Ile-TAT, Lys-CTT, Phe-GAG and Tyr-GTA. Regarding class II tRNAs, the Leu-TAA genes have a T between positions 9 and 10 of Box A (Fig. 6A).

A few exceptions to the consensus sequence were found among Trityps. For Box A from class I genes, these include the following: four tRNA-Val genes (LmjF09.VAL.01, LmjF09.VAL.02, Tb08\_tRNA\_Val\_1 and Tc00.1047053506459.249) have a C at position 2 (instead of A or G); all four tRNA-Val-TAC genes (LmjF23.VAL.01, Tb08\_tRNA\_Val\_1, Tc00.1047053504427.233 and Tc00.1047053508043.13) present an A at position 3 (instead of G); all six tRNA-Ala-AGC genes (LmjF17.TRNA

ALA.01, LmjF31.TRNAALA.01, Tb07\_tRNA\_Ala\_1, Tb07\_tRNA\_Ala\_2, Tc00.1047053510057.40 and Tc00.1047053508909.130) have an extra base (an A) between positions 8 and 9; and the six tRNA-Val-AAC genes from Trityps (LmjF21.TRNAVAL.01, LmjF34.TRNAVAL.01, Tb\_04\_tRNA\_Val\_1, Tb\_04\_tRNA\_Val\_2, Tc00.1047053506435.363 and Tc00.1047053504055.95) also have an additional base (a G) between positions 8 and 9.

Concerning class II genes, the exceptions to Box A consensus sequence are the four tRNA-Leu-TAA genes present in the Trityps genomes (LmjF24.TRNALEU.01, Tb08\_tRNA\_Leu\_2, Tc00.1047053510721.13 and Tc00.1047053511909.9), which have a T at position 2 (instead of G or A). In regard to Box B, the genes that do not have the consensus sequence are: LmjF09.TRNAHIS.01 and LmjF09.TRNAHIS.02 present a C at position 4; Tb07\_tRNA\_Ala\_3 has a T at position 5;

**A**

|                      | Box A |     |     |     |   |     |   |     |     |     |    |    |
|----------------------|-------|-----|-----|-----|---|-----|---|-----|-----|-----|----|----|
|                      | 1     | 2   | 3   | 4   | 5 | 6   | 7 | 8   | 9   | *   | 10 | 11 |
| Trityps class I      | T     | G/A | G   | C/T | N | C/T | A | G/A | C/T | C/T | G  | G  |
| Trityps class II     | T     | G/A | G/C | C/T | C | G/A | A | G   | T   | T   | G  | G  |
| <i>S. cerevisiae</i> | T     | G/A | G   | C/T | N | N   | A | G/A | H   | H   | G  | G  |

|                      | Box B |   |   |   |   |     |   |   |     |    |     |
|----------------------|-------|---|---|---|---|-----|---|---|-----|----|-----|
|                      | 1     | 2 | 3 | 4 | 5 | 6   | 7 | 8 | 9   | 10 | 11  |
| Trityps class I      | G/A   | G | T | T | C | G/A | A | N | H   | C  | C/T |
| Trityps class II     | G     | G | T | T | C | G/A | A | A | C/T | C  | C   |
| <i>S. cerevisiae</i> | N     | G | T | T | C | G/A | A | N | C/T | C  | N   |

**B**

|                        | tRNA-Sec box A |   |          |   |   |   |   |   |   |   |   |   |   |    |
|------------------------|----------------|---|----------|---|---|---|---|---|---|---|---|---|---|----|
| <i>H. sapiens</i>      | T              | G | <u>A</u> | T | C | C | T | C | A | G | T | G | G | 13 |
| <i>X. laevis</i>       | T              | G | <u>A</u> | C | C | C | T | C | A | G | T | G | G | 13 |
| <i>D. melanogaster</i> | T              | G | <u>A</u> | A | C | T | T | C | G | G | T | G | G | 13 |
| <i>C. elegans</i>      | T              | G | <u>A</u> | A | C | C | A | T | G | G | C | G | G | 13 |
| <i>P. falciparum</i>   | T              | G | <u>A</u> | G | T | T | A | G | C | A | T | G | G | 13 |
| Trityps                | T              | G | <u>A</u> | G | C | T | C | A | G | C | T | G | G | 13 |
|                        | *              | * | *        |   |   |   |   |   |   | * | * |   |   |    |

**Figure 6**

**Consensus sequences for Box A and Box B from tRNA genes in Trityps.** Sequences from class I and class II tRNA genes from Trityps are compared to the *S. cerevisiae* consensus sequences (panel A). In Box A, some tRNA genes contain an extra base between positions 9 and 10 (marked with an asterisk) (see text). Conserved positions are shown in bold type. Panel B shows a comparison of Box A from selenocysteine tRNA genes from the indicated species. Conserved nucleotides are indicated with an asterisk. An A in the third position (in bold and underlined) seems to be specific to Box A from Sec genes. H represents C, T or A.

Tc00.1047053511241.10 presents an A at position 10; and tRNA-Ala-AGC and tRNA-iMet genes have an A at position 3 (data not shown).

Analysis of the promoter sequences from tRNA-Sec genes in Trityps indicated that Box A contains an additional A between bases 2 and 3, compared with the consensus sequences (see Fig. 6B). This insertion was previously reported in tRNA-Sec genes from other organisms [35] (Fig. 6B). Regarding Box B, tRNA-Sec genes from Trityps present two changes compared to the highly conserved consensus sequence: a C at position 1 (instead of a G) and a G at position 11 (in lieu of C) (data not shown). In other species, the sequence of Box B from tRNA-Sec is identical

to the corresponding consensus sequence. In *Xenopus laevis*, transcription of tRNA-Sec genes is directed by three extragenic domains (a TATA Box, a proximal sequence element and an activator element) and Box B. Apparently, Box A is not part of the promoter [36,37]. Since both internal control elements from Sec genes in Trityps differ from the corresponding consensus sequences, it is possible that synthesis of tRNA-Sec is regulated only by external elements in these parasites. We are currently exploring this possibility.

**tRNA isodecoder genes**

Sequence analysis of isoacceptor tRNAs in several organisms indicated the presence of tRNA isodecoder genes

(tRNA genes having the same anticodon but different sequences elsewhere in the tRNA body) [10]. In eukaryotes, the number of isodecoder genes ranges from 10 (yeast) to 246 (chimp), while in bacterial genomes the number of isodecoders varies from 0 to 26 [10]. By comparing the sequences of isoacceptor tRNAs in the *Trypanosoma* spp., one isodecoder gene was found in *L. major* (tRNA-Ser-GCT), and one was found in *T. brucei* (tRNA-Ala-CGC) (Fig. 7 and Table 1). Since there are only two copies of the corresponding isoacceptor class in each case, we have arbitrarily designated LmtRNA-Ser.01 (LmjF17.TRNASER.01) and TbtRNA-Ala.01 (Tb07\_tRNA\_Ala\_3) as the isodecoders. Sequence identity between the isodecoder and the "majority member" is 98% in *L. major* and 97% in *T. brucei*. While in *L. major* the sequence difference locates near Box A, in *T. brucei* one of the two observed differences lies in one of the conserved bases of Box B (Fig. 7). In *T. cruzi*, six isodecoder genes were identified: TctRNA-Glu.01 (Tc00.1047053506435.336), TctRNA-Ala.01 (Tc00.1047053510057.40), TctRNA-Ala.03 (Tc00.1047053475029.40), TctRNA-Arg.01 (Tc00.1047053504427.243), TctRNA-His.01 (Tc00.1047053511241.10) and TctRNA-Ser.01 (Tc00.1047053510057.50) (Table 1 and Fig. 7). In four cases (TctRNA-Glu.01, TctRNA-Ala.03, TctRNA-Arg.01 and TctRNA-His.01), sequence differences were located to variable nucleotides from Box B (Fig. 7). As in *Trypanosoma* spp., sequence variations between human tRNA isodecoders have been located within internal control elements [10]. In such cases, changes were found in variable nucleotides from Boxes A and B. Thus, the occurrence of changes within internal control elements in tRNA isodecoder genes suggests that differential regulation of Pol III transcription is possible in *Trypanosoma* spp.; the fact that the highly conserved C at position 10 of Box B from TbtRNA-Ala.02 is changed to T in the corresponding tRNA isodecoder (TbtRNA-Ala.01) (Fig. 7) strongly supports this possibility. Sequence changes in isodecoders are not only restricted to internal control elements, but they might be present all along the tRNA body (Fig. 7) [10]. Therefore, the diversity of tRNA genes is much higher than originally thought. The functional meaning of such diversity has yet to be investigated.

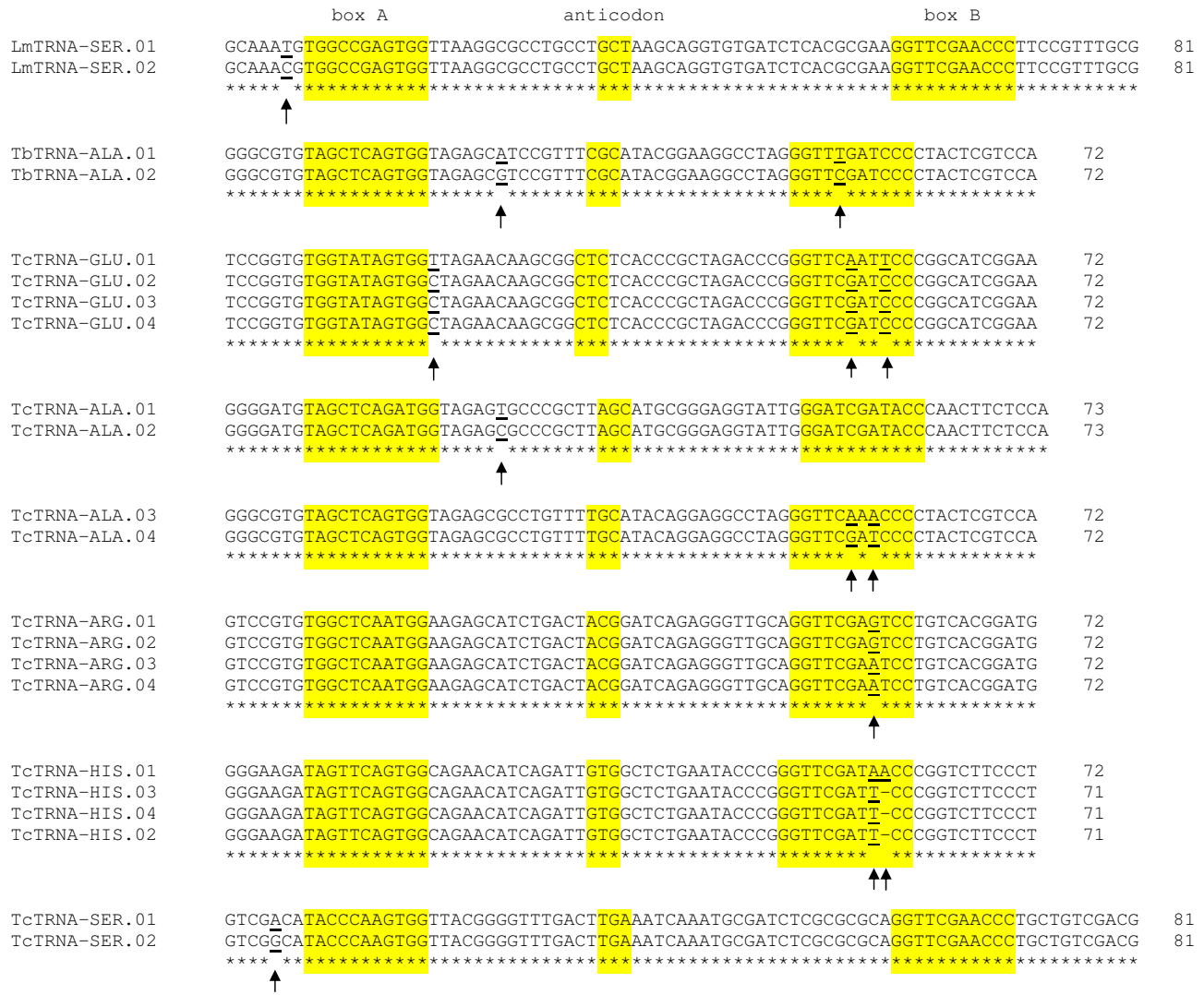
#### Signals for transcription termination

A cluster of several T residues in the coding DNA strand acts as a signal to terminate Pol III transcription [33]. The cluster of Ts is usually located within the first 30 bases following the gene. In human and mice, Pol III needs four Ts to end transcription, and tRNA genes that have five or more Ts are very rare in these species. On the other hand, in the genomes of *S. pombe* and *S. cerevisiae* the majority of the tRNA genes have six and seven Ts, respectively [38,39]. Interestingly, they do not have any single gene whose termination signal is shorter than five Ts. For a par-

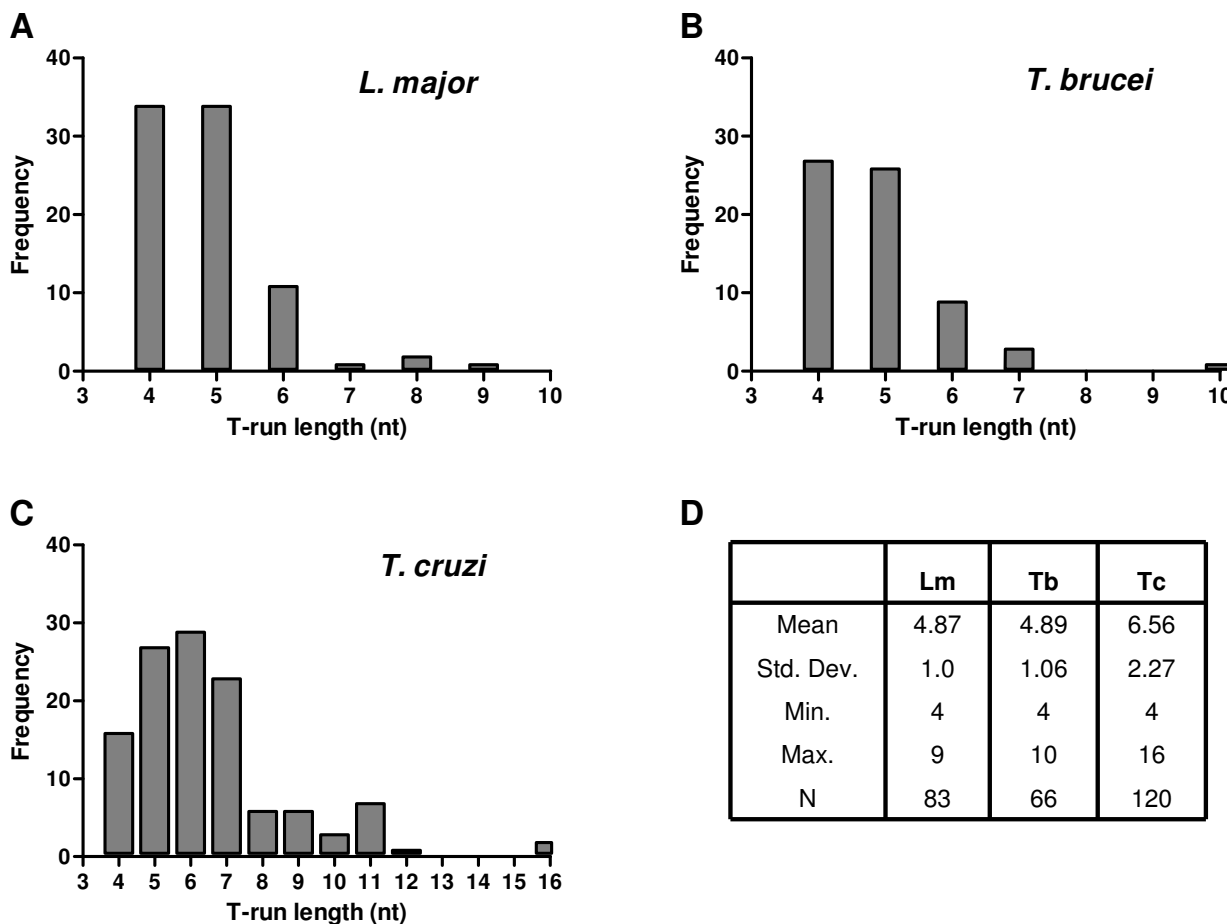
ticular species, termination efficiency tends to increase with the length of the T run. In *L. major*, it has been shown that transcription of the tRNA located on chromosome 3 terminates within a tract of four Ts [6]. To gain insight into Pol III termination signals in trypanosomatids, we decided to analyze the sequences downstream of all the tRNA genes. A cluster of Ts of variable length was found on every single tRNA gene in the *Trypanosoma* spp. (see Additional File 2); the distance between the end of the gene and the run of Ts varies from zero to seven bases. In *L. major*, the mean length of the run of Ts is 4.87 bases, with a minimum of four and a maximum of nine Ts (Fig. 8, panels A and D). Similar results were obtained in *T. brucei*, where the average T-run length is 4.89 bases (ranging from four to ten Ts) (Fig. 8, panels B and D). In the tRNA genes from *T. cruzi*, however, the stretches of Ts are longer, showing a mean length of 6.56, with two genes presenting a run of 16 consecutive T residues (Fig. 8C and 8D).

The presence of a second stretch of Ts that acts as a potential "back up" termination signal is a common feature in tRNA genes from eukaryotes [38]. The second run of Ts is normally located within the first 30 bp downstream of the first one. In the case of *S. cerevisiae*, *S. pombe*, *H. sapiens* and *Mus musculus*, the percentages of tRNA genes that have a back up T-run are 44, 53, 31 and 33%, respectively [38]. Analysis of the sequences downstream of the T-runs in *T. cruzi* showed that 58 tRNA genes (48.3%) have a back up T-run, whose length is between 4 to 15 bases (Additional File 2). Therefore, the percentage of tRNA genes with a second run of Ts in *T. cruzi* is very similar to that found in *S. cerevisiae* and *S. pombe*. Interestingly, only 13 tRNA genes in *L. major* (15.6%) and 18 genes in *T. brucei* (27%) present a back up T-track (which is from 4 to 10 bases long) (see Additional File 2); these percentages are even smaller than those found in mammals. Thus, in *L. major* and *T. brucei* a single and short run of Ts seems to be sufficient to achieve proper transcription termination in the majority of the tRNA genes. On the other hand, *T. cruzi* seems to require longer T-runs, in addition to a second T stretch, to direct transcription termination. This indicates that the mechanism of Pol III transcription termination in *L. major* and *T. brucei* might be different from that one in *T. cruzi* and other eukaryotes. Sequences downstream and upstream of the run of Ts might contribute to the strength of the terminator, as observed in some tRNAs in *S. cerevisiae* [38].

Although most tRNA genes are clustered in trypanosomatids, the presence of runs of Ts located downstream of all the tRNA genes suggests that they are transcribed as monocistrons, which is common among eukaryotes. In prokaryotic cells genes encoding tRNAs are transcribed in either a monocistronic or polycistronic manner. In the latter case, an RNA containing several tRNA precursors in



**Figure 7**  
**Sequence comparison of tRNA isodecoder genes in the Trityps.** One isodecoder tRNA gene was found in *L. major* (LmTRNA-Ser) and *T. brucei* (TbTRNA-Ala), and six were found in *T. cruzi* (TcTRNA-Glu, -Ala.01, -Ala.03, -Arg, -His and -Ser). Bases that show variation are indicated in bold and underlined, and marked with an arrow. The position of internal control elements (Boxes A and B) and the anticodon are indicated. The genes included in this figure are the following (GeneDB names in parentheses): LmTRNA-SER.01 (LmjF17.TRNASER.01), LmTRNA-SER.02 (LmjF21.TRNASER.01), TbTRNA-ALA.01 (Tb07\_tRNA\_Ala\_3), TbTRNA-ALA.02 (Tb11\_tRNA\_Ala\_1), TcTRNA-GLU.01 (Tc00.1047053506435.336), TcTRNA-GLU.02 (Tc00.1047053504055.89), TcTRNA-GLU.03 (Tc00.1047053508999.180), TcTRNA-GLU.04 (Tc00.1047053510959.8), TcTRNA-ALA.01 (Tc00.1047053510057.40), TcTRNA-ALA.02 (Tc00.1047053508909.130), TcTRNA-ALA.03 (Tc00.1047053475029.40), TcTRNA-ALA.04 (gene located in contig 8001, not annotated in geneDB), TcTRNA-ARG.01 (Tc00.1047053504427.243), TcTRNA-ARG.02 (Tc00.1047053506619.59), TcTRNA-ARG.03 (Tc00.1047053508043.23), TcTRNA-ARG.04 (Tc00.1047053511191.29), TcTRNA-HIS.01 (Tc00.1047053511241.10), TcTRNA-HIS.03 (Tc00.1047053508087.5), TcTRNA-HIS.04 (Tc00.1047053508861.10), TcTRNA-HIS.02 (Tc00.1047053506663.10), TcTRNA-SER.01 (Tc00.1047053510057.50), TcTRNA-SER.02 (Tc00.1047053508909.120).



**Figure 8**  
**Length distribution of termination signals from tRNA genes in Trityps.** The size of the run of Ts in every tRNA gene from *L. major* (panel A), *T. brucei* (panel B) and *T. cruzi* (panel C) was plotted against frequency. Statistical data is shown in panel D.

tandem is processed to yield functional tRNAs [40]. In plants, dicistronic transcripts containing a tRNA and a snoRNA have been found [41]. Moreover, the presence of precursor RNA molecules that contain both a tRNA and a mRNA has been reported in *E. coli* [42].

**Conclusion**

In comparison to most eukaryotic organisms, Trityps present a low number of tRNA genes. A total of 46 isoacceptor types were identified, which are able to read the 61 codons that specify the canonical amino acids, in addition to Sec. Trypanosomatids use the A1- or G1-sparing strategy as a decoding mode, by allowing flexible base pairing between G1 or A1 of the anticodon and C3 or U3 in the codon. Most tRNA genes in Trityps are organized into clusters (from 2–10 genes) that may contain other Pol III genes. Some of the clusters show a remarkable conservation of gene order among Trityps. The distribution of

tRNA genes in the genomes of *L. major* and *T. brucei* does not seem to be totally random. We also found that 14 of the 39 convergent strand-switch regions present in the *L. major* genome are separated by at least one tRNA gene, which raise the possibility that other tRNA genes (in addition to the one present on chromosome 3) are involved in termination of Pol II transcription of convergent PGCs in this parasite. A run of Ts of variable length was found downstream of all the 269 tRNA genes present in the Trityps. In *T. cruzi* the clusters of Ts are larger than in *L. major* and *T. brucei* (an average of 6 Ts versus 5 Ts, respectively); moreover, the presence of a back up T run is more common in *T. cruzi* than in the other two Trityps. Analysis of the internal promoter elements allowed us to establish consensus sequences for Boxes A and B of class I and class II tRNA genes. Interestingly, special characteristics were found in Boxes A and B from tRNA-Sec genes in Trityps, which suggests that the mechanisms that regulate their

transcription might be different from those of other tRNA genes. Lastly, we have identified several tRNA isodecoder genes in the *Trityps*, especially in *T. cruzi*. The fact that in some cases the sequence differences occur within the internal promoter elements suggests the possibility of differential expression of tRNA genes in these organisms.

## Methods

All tRNA genes annotated in the *L. major*, *T. brucei* and *T. cruzi* genome databases <http://www.genedb.org> (versions 2.1) were analyzed with the tRNAscan-SE program <http://lowelab.ucsc.edu/tRNAscan-SE/> [43] to verify the presence and identity of the tRNAs. In addition to the examination of all the features associated to the typical tRNA cloverleaf structure, we also analyzed the presence of internal promoter elements (Boxes A and B) and T-tracts at the 3' end of the tRNA gene (which should be present in tRNA genes, but not necessarily in pseudogenes). Sequence comparisons among putative tRNA isoacceptors were performed using the ClustalW2 program <http://www.ebi.ac.uk/Tools/clustalw2/index.html>. Information for the genomic and synteny maps was obtained from the GeneDB databases. BLAST searches were performed in these databases to locate the tRNA-Sec genes [13], as well as the sRNA76 [12]. Codon usage data was obtained from the Sanger site <http://www.sanger.ac.uk/> for *L. major*, and from the Kazusa web page <http://www.kazusa.or.jp/codon/> for *T. brucei*. Codon usage data for *T. cruzi* was calculated by analyzing coding sequences (obtained from the Sanger site) in the codon usage page from SMS [http://www.bioinformatics.org/sms2/codon\\_usage.html](http://www.bioinformatics.org/sms2/codon_usage.html). The Spearman correlation analysis and the descriptive statistical analysis of T-run data were performed with the GraphPad Prism5 program <http://www.graphpad.com>.

## Authors' contributions

NPM performed data collection, synteny analysis, sequence comparisons, location of tRNA-Sec genes and the statistical analysis of the data. She also carried out the analysis of the transcription terminations signals and helped draft the manuscript. LEFM analyzed the tRNA genes with the tRNAscan-SE program and helped draw the genome maps. EEFA helped collect data from the *T. brucei* databases, and analyzed the sequences to obtain the consensus sequences for boxes A and B. RGMC was responsible for data collection from the *T. cruzi* database and participated in the codon usage analysis. RHR helped draft the manuscript and participated in the statistical analysis of the data. PJM participated in the writing and editing of the manuscript. SMC conceived the study, analyzed data and wrote the manuscript. All authors read and approved the manuscript.

## Additional material

### Additional file 1

**Table S1. List of tRNA genes in the *Trityps*.** In *L. major* and *T. brucei*, genes are shown by chromosomal location. Other Pol III-transcribed genes that are associated to tRNA genes are also listed. In *T. brucei*, the names of the genes correspond to the temporary systematic name. For those genes that already have a permanent systematic name in the GeneDB database, it is indicated between parentheses. tRNA genes in *T. cruzi* are organized by amino acid.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-232-S1.xls>]

### Additional file 2

**Table S2. tRNA-gene transcription termination signals in *Trityps*.** The sequences shown start immediately downstream of the end of the tRNA genes. Clusters of Ts are shown in bold type. Those genes that present a second (back up) T-run of 4 or more bases are highlighted in yellow.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-232-S2.xls>]

## Acknowledgements

We thank Ana M Cevallos-Gaos for fruitful discussions. This work was supported by grants 1 R01 TW007255-01 from The Fogarty International Center of NIH, 47543 from CONACyT and IN203606-3 from PAPIIT to S. Martínez-Calvillo. N.E. Padilla-Mejía (Posgrado en Ciencias Biológicas, UNAM) was recipient of a fellowship from CONACyT.

## References

- Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchetti G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarlioni L, Clayton C, Coulson RM, Cronin A, Cruz AK, Davies RM, De GJ, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N, Frasch AC, Fraser A, Fuchs M, Gabel C, Goble A, Goffeau A, Harris D, Hertz-Fowler C, Hilbert H, Horn D, Huang Y, Klages S, Knights A, Kube M, Larke N, Litvin L, Lord A, Louie T, Marra M, Masuy D, Matthews K, Michaeli S, Mottram JC, Muller-Auer S, Munden H, Nelson S, Norbertczak H, Oliver K, O'neil S, Pentony M, Pohl TM, Price C, Purnelle B, Quail MA, Rabinowitsch E, Reinhardt R, Rieger M, Rinta J, Robben J, Robertson L, Ruiz JC, Rutter S, Saunders D, Schafer M, Schein J, Schwartz DC, Seeger K, Seyler A, Sharp S, Shin H, Sivam D, Squares R, Squares S, Tosato V, Vogt C, Volckaert G, Wambutt R, Warren T, Wedler H, Woodward J, Zhou S, Zimmermann W, Smith DF, Blackwell JM, Stuart KD, Barrell B, Myler PJ: **The genome of the kinetoplastid parasite, *Leishmania major*.** *Science* 2005, **309**:436-442.
- Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UC, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DM, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabinowitsch E, Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CM, Tait A, Tivey AR, Van AS, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman



- J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El-Sayed NM: **The genome of the African trypanosome *Trypanosoma brucei***. *Science* 2005, **309**:416-422.
3. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberg S, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de JP, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van AS, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B: **The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease**. *Science* 2005, **309**:409-415.
  4. Campbell DA, Thomas S, Sturm NR: **Transcription in kinetoplastid protozoa: why be normal?** *Microbes Infect* 2003, **5**:1231-1240.
  5. Martinez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ: **Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region**. *Mol Cell* 2003, **11**:1291-1299.
  6. Martinez-Calvillo S, Nguyen D, Stuart K, Myler PJ: **Transcription initiation and termination on *Leishmania major* chromosome 3**. *Eukaryot Cell* 2004, **3**:506-517.
  7. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renaud H, Worthey EA, Hertz-Fowler C, Ghedin E, Peacock C, Bartholomeu DC, Haas BJ, Tran AN, Wortman JR, Alsmark UC, Angiuoli S, Anupama A, Badger J, Bringaud F, Cadag E, Carlton JM, Cerqueira GC, Creasy T, Delcher AL, Djikeng A, Embley TM, Hauser C, Ivans AC, Kummerfeld SK, Pereira-Leal JB, Nilsson D, Peterson J, Salzberg SL, Shallow J, Silva JC, Sundaram J, Westenberg S, White O, Melville SE, Donelson JE, Andersson B, Stuart KD, Hall N: **Comparative genomics of trypanosomatid parasitic protozoa**. *Science* 2005, **309**:404-409.
  8. Geiduschek EP, Kassavetis GA: **The RNA polymerase III transcription apparatus**. *J Mol Biol* 2001, **310**:1-26.
  9. Marck C, Grosjean H: **tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features**. *RNA* 2002, **8**:1189-1232.
  10. Goodenbour JM, Pan T: **Diversity of tRNA genes in eukaryotes**. *Nucleic Acids Res* 2006, **34**:6137-6146.
  11. Paule MR, White RJ: **Survey and summary: transcription by RNA polymerases I and III**. *Nucleic Acids Res* 2000, **28**:1283-1298.
  12. Beja O, Ullu E, Michaeli S: **Identification of a tRNA-like molecule that copurifies with the 7SL RNA of *Trypanosoma brucei***. *Mol Biochem Parasitol* 1993, **57**:223-229.
  13. Cassago A, Rodrigues EM, Prieto EL, Gaston KW, Alfonso JD, Iribar MP, Berry MJ, Cruz AK, Thiemann OH: **Identification of *Leishmania* selenoproteins and SECIS element**. *Mol Biochem Parasitol* 2006, **149**:128-134.
  14. Allmang C, Krol A: **Selenoprotein synthesis: UGA does not end the story**. *Biochimie* 2006, **88**:1561-1571.
  15. Lobanov AV, Gromer S, Salinas G, Gladyshev VN: **Selenium metabolism in *Trypanosoma*: characterization of selenoproteomes and identification of a Kinetoplastida-specific selenoprotein**. *Nucleic Acids Res* 2006, **34**:4012-4024.
  16. Tan TH, Pach R, Crausaz A, Ivans A, Schneider A: **tRNAs in *Trypanosoma brucei*: genomic organization, expression, and mitochondrial import**. *Mol Cell Biol* 2002, **22**:3707-3717.
  17. Horn D: **Codon usage suggests that translational selection has a major impact on protein expression in trypanosomatids**. *BMC Genomics* 2008, **9**:2.
  18. Drabkin HJ, Estrella M, Rajbhandary UL: **Initiator-elongator discrimination in vertebrate tRNAs for protein synthesis**. *Mol Cell Biol* 1998, **18**:1459-1466.
  19. Astrom SU, von Pawel-Rammingen U, Bystrom AS: **The yeast initiator tRNA<sup>Met</sup> can act as an elongator tRNA(Met) in vivo**. *J Mol Biol* 1993, **233**:43-58.
  20. Mandal N, Mangroo D, Dalluge JJ, McCloskey JA, Rajbhandary UL: **Role of the three consecutive G:C base pairs conserved in the anticodon stem of initiator tRNAs in initiation of protein synthesis in *Escherichia coli***. *RNA* 1996, **2**:473-482.
  21. Percudani R, Pavesi A, Ottonello S: **Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae***. *J Mol Biol* 1997, **268**:322-330.
  22. Rocha EP: **Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization**. *Genome Res* 2004, **14**:2279-2286.
  23. Duret L: **tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes**. *Trends Genet* 2000, **16**:287-289.
  24. Hasan G, Turner MJ, Cordingley JS: **Ribosomal RNA genes of *Trypanosoma brucei*: mapping the regions specifying the six small ribosomal RNAs**. *Gene* 1984, **27**:75-86.
  25. Hernandez-Rivas R, Martinez-Calvillo S, Romero M, Hernandez R: ***Trypanosoma cruzi* 5S rRNA genes: molecular cloning, structure and chromosomal organization**. *FEMS Microbiol Lett* 1992, **71**:63-67.
  26. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Showlken R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendt MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubinfeld M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la BM, Dedhia N, Blocker H, Hornischer K, Nordseik G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korff I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de JP, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ: **Initial sequencing and analysis of the human genome**. *Nature* 2001, **409**:860-921.
  27. Thompson M, Haeusler RA, Good PD, Engelke DR: **Nucleolar clustering of dispersed tRNA genes**. *Science* 2003, **302**:1399-1401.
  28. Takahashi K, Murakami S, Chikashige Y, Niwa O, Yanagida M: **A large number of tRNA genes are symmetrically located in fission yeast centromeres**. *J Mol Biol* 1991, **218**:13-17.
  29. Scott KC, White CV, Willard HF: **An RNA polymerase III-dependent heterochromatin barrier at fission yeast centromere I**. *PLoS ONE* 2007, **2**:e1099.
  30. Obado SO, Bot C, Nilsson D, Andersson B, Kelly JM: **Repetitive DNA is associated with centromeric domains in *Trypanosoma brucei* but not *Trypanosoma cruzi***. *Genome Biol* 2007, **8**:R37.

31. Thomas S, Green A, Sturm NR, Campbell DA, Myler PJ: **Histone acetylations mark origins of polycistronic transcription in *Leishmania major*.** *BMC Genomics* 2009, **10**:152.
32. Tripathy S, Tyler BM: **The repertoire of transfer RNA genes is tuned to codon usage bias in the genomes of *Phytophthora sojae* and *Phytophthora ramorum*.** *Mol Plant Microbe Interact* 2006, **19**:1322-1328.
33. Geiduschek EP, Tocchini-Valentini GP: **Transcription by RNA polymerase III.** *Annu Rev Biochem* 1988, **57**:873-914.
34. Brun YV, Sanfacon H, Breton R, Lapointe J: **Closely spaced and divergent promoters for an aminoacyl-tRNA synthetase gene and a tRNA operon in *Escherichia coli*. Transcriptional and post-transcriptional regulation of *glxX*, *valU* and *alaW*.** *J Mol Biol* 1990, **214**:845-864.
35. Lee BJ, Kang SG, Hatfield D: **Transcription of *Xenopus* selenocysteine tRNA<sup>Ser</sup> (formerly designated opal suppressor phosphoserine tRNA) gene is directed by multiple 5'-extragenic regulatory elements.** *J Biol Chem* 1989, **264**:9696-9702.
36. Carbon P, Krol A: **Transcription of the *Xenopus laevis* selenocysteine tRNA(Ser)<sup>Sec</sup> gene: a system that combines an internal B box and upstream elements also found in U6 snRNA genes.** *EMBO J* 1991, **10**:599-606.
37. Park JM, Yang ES, Hatfield DL, Lee BJ: **Analysis of the selenocysteine tRNA[Ser]<sup>Sec</sup> gene transcription in vitro using *Xenopus* oocyte extracts.** *Biochem Biophys Res Commun* 1996, **226**:231-236.
38. Braglia P, Percudani R, Dieci G: **Sequence context effects on oligo(dT) termination signal recognition by *Saccharomyces cerevisiae* RNA polymerase III.** *J Biol Chem* 2005, **280**:19551-19562.
39. Matsuzaki H, Kassavetis GA, Geiduschek EP: **Analysis of RNA chain elongation and termination by *Saccharomyces cerevisiae* RNA polymerase III.** *J Mol Biol* 1994, **235**:1173-1192.
40. Vold BS, Green CJ, Narasimhan N, Strem M, Hansen JN: **Transcriptional analysis of *Bacillus subtilis* rRNA-tRNA operons. II. Unique properties of an operon containing a minor 5 S rRNA gene.** *J Biol Chem* 1988, **263**:14485-14490.
41. Kruszka K, Barneche F, Guyot R, Ailhaas J, Meneau I, Schiffer S, Marchfelder A, Echeverria M: **Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z.** *EMBO J* 2003, **22**:621-632.
42. van Delft JH, Schmidt DS, Bosch L: **The tRNA-tufB operon transcription termination and processing upstream from tufB.** *J Mol Biol* 1987, **197**:647-657.
43. Lowe TM, Eddy SR: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic Acids Res* 1997, **25**:955-964.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

