Towards a framework for the evolutionary genomics of Kinetoplastids: what kind of data and how much?
Alberto MR Dávila*1 and Julius Lukeš2

Address: 1Lab. de Biologia Molecular de Tripanosomatídeos, Departamento de Bioquímica e Biologia Molecular, Instituto Oswaldo Cruz, Fiocruz. Av. Brasil 4365, Rio de Janeiro, RJ, Brasil. 21045-900 and 2Institute of Parasitology, Czech Academy of Sciences and Faculty of Biology, University of South Bohemia, Branišovská 31, 37005 Ceske Budejovice, Czech Republic

Email: Alberto MR Dávila* - davila@fiocruz.br; Julius Lukeš - jula@paru.cas.cz
* Corresponding author

Abstract
The current status of kinetoplastids phylogeny and evolution is discussed in view of the recent progresses on genomics. Some ideas on a potential framework for the evolutionary genomics of kinetoplastids are presented.

Editorial
In this journal, the concept of species are presented by Wendy Gibson [1] and Hooman Momen [2] while Michel Tibayrenc [3] reviews the evolution and variability of Trypanosoma cruzi. Veronique Hannnaert and collaborators [4] provide a comprehensive review on the evolution of energy metabolism in Kinetoplastida and Austin Hughes and Helen Piontkivska [5] discuss the controversial issue of the monophyletic versus paraphyletic status of the genus Trypanosoma. The latter authors concluded that the markers used to date failed to provide a consistent answer about the origin of trypanosomes. The arguments for the use of more protein-based sequences and a larger number of species are admirable. However, some important questions remain: which genes and how many are needed? Which other species should be selected and how many should be explored?

Generally, a limited number of genes available for a large number of taxa is sufficient for most phylogenetic studies. However, in order to resolve the relationships among closely related species and strains, and to better understand their origin, the availability of a set of evolutionary informative genes would certainly be useful. The ribosomal (r)RNA genes (18S, 5S, SL) still dominate the field, while only a few informative protein-coding genes have been described. The identification and choice of individual genes or gene families is sometimes difficult and less cost-effective when compared with high throughput sequencing initiatives [6]. This observation makes us to reflect about the need to discover new candidate genes for the study of kinetoplastid evolution. According to Podlipaev (2000, 2001) [7,8], from the 10 trypanosomatid genera [9], those (six) associated with insects have so far received little attention or have been "neglected" compared to species with an impact on human health. The same is true for all bodonids.

Still, Trypanosomatina flagellates are in the forefront among kinetoplastids, since several species (T. cruzi, T. brucei and Leishmania major) will have their genomes completely sequenced within the next few months. While the recent start of genome projects for L. braziliensis, L. infantum, T. vivax and T. congolense is very welcome, it is worth noticing that all the above mentioned species belong to only two (Trypanosoma and Leishmania) of the 20 kinetoplastid genera, showing the stark under-representation of the other genera in the sequencing initiative.
A casual search in the Entrez division of NCBI showed that although virtually omnipresent in nature, leptomonads and phytomonads have only dozens of entries, a mere trickle when compared with 139667 and 29777 entries for *Trypanosoma* and *Leishmania*, respectively (see Table 1). It is clear from the pattern of data available that efforts should be focused on obtaining more sequences from the "non-Leishmania" and "non-Trypanosoma" kinetoplastids. Centering on one "representative" or "model" species for each of the underrepresented genera *Bodo*, *Trypanoplasma* and *Cryptobia* would certainly be both worthwhile and rewarding.

Table 1: Survey of kinetoplastid sequences available in the Genbank.

| Kinetoplastida          | Nucleotides | Proteins |
|-------------------------|-------------|----------|
| *Trypanosomatina*       |             |          |
| *Trypanosoma*           | 139667      | 3695     |
| *Leishmania*            | 29777       | 2946     |
| *Crithidia*             | 163         | 206      |
| *Herpetomonas*          | 137         | 15       |
| *Phytomonas*            | 112         | 12       |
| *Leptomonas*            | 55          | 23       |
| *Blastocrithidia*       | 12          | 7        |
| *Wallaceina*            | 10          | 3        |
| *Endotrypanum*          | 9           | 5        |
| *Bodonina*              |             |          |
| *Bodo*                  | 64          | 12       |
| *Trypanoplasma*         | 22          | 24       |
| *Cryptobia*             | 14          | 5        |
| *Dimastigella*          | 9           | 1        |
| *Rhynchomonas*          | 6           | 1        |
| *Perkinsiella*          | 6           | 0        |
| *Rhynchobodo*           | 3           | 0        |
| *Cruella*               | 3           | 0        |
| *Ichthyobodo*           | 2           | 0        |
| *Parabodo*              | 2           | 0        |
| *Procyptobia*           | 1           | 0        |

An important issue towards the exploration of the transcriptome of kinetoplastids outside Trypanosomatina is the choice of "model species". These "models" would be used to get insights into the evolution of bizarre structures and processes so far mostly described in *Trypanosoma* and less so in *Leishmania*. One proposal could be the study of the evolution of VSGs, as well as complex proteins involved in RNA editing or trans-splicing. A comparative approach has recently proved to be useful for our understanding of how the complex kinetoplast DNA network may have evolved [12]. The "models" should be members of major kinetoplastid clades in the current phylogenetic trees [13] that grow in reasonable densities in simple culture media, preferably axenic. *Bodo saltans* may be a good model, because it branches off on the border between bodonids and trypanosomatids, and is an ecologically important free-living species of potentially world-wide distribution. Its sole disadvantage is the growth with a feeder bacteria. As far as we know, the only bodonid that grows in an axenic medium is *Trypanoplasma borreli*, a parasite of marine and freshwater fish, another potential candidate. The final proposed candidate is *Diplonema papillatum*, since it grows axenically in simple media to high densities. Moreover, it would be an excellent outgroup for kinetoplastids, since it represents the most related group and likely shares with them numerous peculiarities.

Inferring from the recently unveiled diversity of extremely small eukaryotes [14], one would predict that a great diversity of kinetoplastids is still to be discovered. Perhaps an enthusiastic overestimate, some discuss the existence of a great trypanosomatid biodiversity in insects alone [15]. Recent surveys have found new species in ocean depths [16,17] or unexpected niches [18]. One of the approaches used for the discovery of these new eukaryote taxa is culture-independent-survey by PCR (cI-PCR), which has also helped to double the number of prokaryotic
phyla [19]. This approach would be particularly interesting in a search for difficult-to-cultivate Kinetoplastida. For such a survey, a number of known conserved markers/genes would be needed for the design of PCR-based assays and, unfortunately, that is not possible with the limited data currently available. There may be some requirements for choosing the markers since according to Charlebois and collaborators [20] protein-coding regions are usually less universal, more difficult to PCR-amplify and often shorter and less information-rich than the rRNA genes. Possible discrepancies between the rRNA- and protein-based trees can be avoided by choosing informative and unbiased genes and by constructing trees on the basis of concatenated alignments. Whole-genome trees are becoming popular given the availability of a large number of bacterial genomes in the databases, however, for the near future, it is not going to be an option for non-pathogenic kinetoplastids and in general for eukaryotes because of the size of their genomes. New technologies for sequencing large genomes and cheaper costs per base could, in the future, facilitate that approach even for protozoans.

It seems that with the available genomics approaches, there are now more chances to map the real diversity of kinetoplastids in nature, and explore their fascinating biology, evolution and “roots”.

Acknowledgements
We are very grateful to Wendy Gibson, Sergei Podlipaev and Hooman Momen for their comments on the manuscript, however, all mistakes and wrong interpretations remain ours.

References
1. Gibson W: Species concepts for trypanosomais: from morphological to molecular definitions? Kinetoplastid Biology and Disease 2003, 2(10).
2. Momen H: Concepts of species in trypanosomatids. Kinetoplastid Biology and Disease 2003, 2(13).
3. Ticayrenc M: Genetic subdivisions within Trypanosoma cruzi (Discrete Typing Units) and their relevance for molecular epidemiology and experimental evolution. Kinetoplastid Biology and Disease 2003, 2(12).
4. Hanaert V, Bringaud F, Opprøddes FR and Michels PAM: Evolution of energy metabolism and its compartmentation in Kinetoplastida. Kinetoplastid Biology and Disease 2003, 2(11).
5. Hughes AL and Piontkivska H: Molecular Phylogenetics of Trypanosomatidae: Contrasting Results from 18S rRNA and Protein Phylogenies. Kinetoplastid Biology and Disease 2003, 2(15).
6. Davila AMR, Majwa FAO, Grissard EC, Aksoy S and Melville SE: Comparative genomics to uncover the secrets of tsetse and livestock-infective trypanosomases. Trends Parasitol 2003, 19(10):436-439.
7. Podlipaev S: The more insect trypanosomatids under study—the more diverse Trypanosomatidae appears. Int J Parasitol 2001, 31(5-6):648-52.
8. Podlipaev S: Insect trypanosomatids: the need to know more. Mem Inst Oswaldo Cruz 2000, 95(4):517-22.
9. Vickerman K: Order Kinetoplastida Honigberg, 1963. In: An Illustrated Guide to the Protozoa Volume 2. Second edition. Edited by: Lee JL, Leedale GF, Bradbury P. Society of Protozoologists, USA; 2000:1159-1180.
10. El-Sayed NM and Donelson JE: A survey of the Trypanosoma brucei rhodesiense genome using shotgun sequencing. Mol Biochem Parasitol 1997, 84(2):167-78.
11. Porcel BM, Tran AN, Tammi M, Nyarady Z, Rydaker M, Urmenyi TP, Rondinelli E, Pettersson U, Andersson B and Aslund L: Gene survey of the pathogenic protozoan Trypanosoma cruzi. Genome Res 2000, 10(8):1103-7.
12. Lukej J, Guilbride DL, Vojtiska J, Zilová A, Benne R and Englund PT: Kinetoplast DNA network: evolution of an improbable structure. Eukaryot Cell 2002, 1(4):495-502.
13. Simpson AGB, Lukej J and Roger AJ: The evolutionary history of kinetoplastids and their kinetoplasts. Mol Biol Evol 2002, 19:2071-2083.
14. Massana R, Guillou L, Diez B and Pedros-Alio C: Unveiling the organisms behind novel eukaryotic ribosomal DNA sequences from the ocean. Appl Environ Microbiol 2002, 68(9):4554-8.
15. Stevens JR: One million insects – a lot of parasites? Trends Parasitol 2001, 17:119-120.
16. Atkins MS, Teske AP and Anderson OR: A survey of flagellate diversity at four deep-sea hydrothermal vents in the Eastern Pacific Ocean using structural and molecular approaches. J Eukaryot Microbiol 2000, 47(4):460-11.
17. Lopez-Garcia P, Philippe H, Gail F and Moreira D: Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. Proc Natl Acad Sci U S A 2003, 100(2):697-702.
18. Dyková I, Flasa I, Lom J and Lukej J: Perkinsiella amoeboae-like endosymbionts of Neoparamoeba spp, relatives of the kinetoplastid Ichthyobodo. Europ J Protistol 2003, 39:37-52.
19. DeLong EF and Pace NR: Environmental diversity of bacteria and archaea. Syst Biol 2001, 50(4):470-8.
20. Charlebois RL, Beiko RG and Ragan MA: Microbial phylogenomics: Branching out. Nature 421(6920):217. 2003 Jan 16

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