Future Trypanosomatid Phylogenies: Refined Homologies, Supertrees and Networks

JR Stothard

Biomedical Parasitology Division, Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

There has been good progress in inferring the evolutionary relationships within trypanosomes from DNA data as until relatively recently, many relationships have remained rather speculative. Ongoing molecular studies have provided data that have adequately shown Trypanosoma *to be monophyletic and, rather surprisingly, that there are sharply contrasting levels of genetic variation within and between the major trypanosomatid groups. There are still, however, areas of research that could benefit from further development and resolution that broadly fall upon three questions.*

Are the current statements of evolutionary homology within ribosomal small sub-unit genes in need of refinement? Can the published phylograms be expanded upon to form 'supertrees' depicting further relationships? Does a bifurcating tree structure impose an untenable dogma upon trypanosomatid phylogeny where hybridisation or reticulate evolutionary steps have played a part? This article briefly addresses these three questions and, in so doing, hopes to stimulate further interest in the molecular evolution of the group.

Key words: gene trees - characterization - *Trypanosoma - Leishmania - Crithidia*

Until relatively recently the systematics of the kinetoplastid protozoa have remained comparatively stable (Vickerman 1994). Studies of molecular DNA variation, however, have begun to challenge the existing classifications and explore the generic and specific boundaries within this group. Owing to the lack of clear morphological characters, assessment of monophyletic groups has been difficult. As a consequence, molecular studies have been increasingly implemented to find new, reliable characters to relate groups or organisms. The provisional hope of these DNA studies is to find sequences that, on the one hand, allow differentiation of closely related species, e.g. intra-generic relationships, and, on the other, provide characters sufficiently conserved to infer deeper level relationships e.g. inter-generic.

Most central in these molecular studies has been sequence analysis of the nuclear ribosomal small subunit (18S). Since all cells contain ribosomal genes, comparisons between vastly diverse organisms are permitted (Hillis & Dixon 1991), with one or two exceptions, e.g. viruses, where rDNA is absent. Since there are certain domains within the 18S that are highly conserved, these regions act as convenient anchor sites for polymerase chain reaction amplification. Universal primers have allowed many new taxa to be quickly added to the growing 18S sequence database. For example, during the early 1990's only a handful of parasitic protozoa had been characterised. Within the last five years, 18S sequences of many kinetoplastids have now been determined; a quick survey of GENBANK (March 2000) shows that there are 76 entries reporting of either complete or partial coverage of the 18S and associated promoter regions.

REFINED HOMOLOGIES

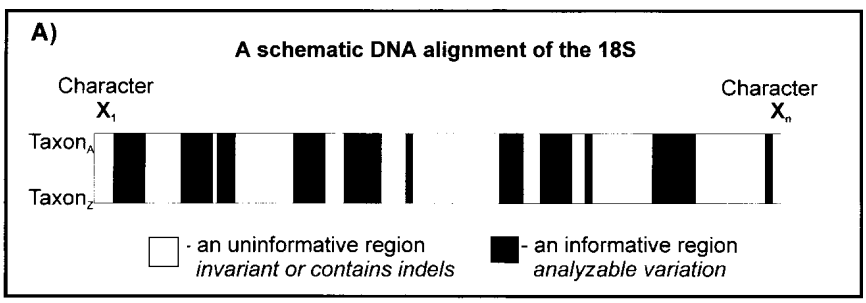
Whilst the 18S has been widely used, there are, however, some concerns about the more general use of this sequence in phylogenetic inference (Abouheif et al. 1998). Fernandes et al. (1993), Maslov et al. (1996) and Lukes et al. (1997) utilised variation within the 18S to infer phylogeny within the kinetoplatids leading to the suggestion that *Trypanosoma* was monophyletic. Noyes (1998) expanded upon the taxonomic coverage of the 18S data and stimulated much debate both in biogeographic implications (Stevens & Gibson 1998) and methodology of inference from these data (Maslov & Lukes 1998, Noyes & Rambaut 1998). Further taxon sampling led to the cladograms reported by Stevens and Gibson (1999) and Wright et al. (1999). Through the use of biogreographic vicariance, Stevens and Gibson (1999) were able to relate the evolutionary patterns with that of continental drift and dispersal of early hominids. The current 18S

Fax: +44-207-942.5518. E-mail: r.stothard@nhm.ac.uk
Recived 13 April 2000
Accepted 15 May 2000

data also have some other, more implicit problems such as the designation of evolutionary homology. For example, in *T. cruzi* there would appear to be at least two divergent 18S types present within the genome (Stothard et al. 1998). The present alignment comparisons may therefore be between paralogous and not orthologous 18S copies and relationships may reflect a gene tree rather than a species tree (Stothard 2000). Hopefully, the present 'soft' polytomies may be resolvable in due refinement of homology.

Whilst the cladograms of Stevens and Gibson (1999) and Wright et al. (1999) have a degree of congruence, there are some interesting anomalies that require consideration. For example, Stevens and Gibson (1999) infer *Crithidia* to be a monophyletic sister taxon of *Leishmania*. This clear relationship was split, however, by Wright et al. (1999). *C. fasciculata* and *C. oncopelti* were more ambiguously placed as sister taxa of *Leptomonas* sp. and *Blastocrithidia culicis* respectively. The behaviour of *Crithidia* between the two studies is indicative of the complex effects of taxon sampling bias has upon cladograms. The changing cladogram topologies can be easily explained in consideration of the way in which 18S data are analysed and are explored briefly in this article. The 18S sequence data consist of regions that form secondary structures as well as regions that have little, if any, intrastrand folding (i.e. open loops). Regions can be highly variable, especially where sequences have undergone expansion and(or) contraction. As more divergent taxa are added to the alignment, certain previously informative characters are lost from later phylogenetic analysis through the inclusion of gaps or that characters change in their polarity.

For example, consider the following hypothetical alignment where there are X_1 to X_n characters taken from a collection of taxa A to Z. Within this alignment there are regions that are informative (i.e. synapomorphies) as well as those that are not (i.e. autapomorphies, invariant characters and insertions/deletions) (Fig. 1A). As further taxa are added to this alignment, and given that this prior alignment was satisfactory and does not have to be realigned, the evolutionary characters may, or may not, undergo changes of their informational content (Fig. 1B). Consider the addition of $taxon_E$ to a



B) **Additional taxa change character states with alignment**

	X_1	X_2	X_3	X_4	X_5
$Taxon_A$	G	A	A	G	A
$Taxon_B$	G	G	A	G	A
$Taxon_C$	G	G	G	G	G
$Taxon_D$	G	G	G	G	G
. . . . .					
$Taxon_E$	G	A	G	-	-
character change					
from	invariant *not informative*	variant *not informative*	variant *informative*	invariant *not informative*	variant *informative*
to	invariant *not informative*	variant *informative*	variant *informative polarity change*	variant *not informative*	variant *not informative*

Schematic representation of an alignment of 18S sequences collected from several taxa and the effect that the addition of a taxon can have upon DNA characters within the alignment. A: a hypothetical alignment of 18S data contains regions that are phylogenetically informative, as well as regions that are not; B: the addition of further taxa to an existing 18S alignment can affect the DNA characters in numerous ways e.g. characters may be lost, change in polarity or become informative.

previous alignment of $taxa_{AtoD}$, character X_1 remains unchanged since it is invariant. Character X_2, however, is significantly altered, an autapomorphy becomes a synapomorphy. Similarly, character X_3 has a potential change in character polarity. The character state 'G' could be considered a plesimorphy with 'A' as derived or *vice versa*. Only through an outgroup (re)definition could such polarity be resolved. More problematic, whilst characters X_4 and X_5 were present in the alignment, but only X_5 was informative, both characters are subsequently removed by the addition of $taxon_E$ i.e. by an insertion/deletion. Unless this deletion event can be recoded in some way [perhaps as a missing state(?)], many computer programs conducting maximum parsimony would now ignore character X_5 from analysis. It is therefore not surprising that as we add more and more divergent taxa to the same 18S alignment, there can be substantial change in its informational content. Resolution between closely related taxa could be lost and(or) poorly supported clades vanish. After the inclusion of further taxa, a better understanding, and more explicit analysis, of the behaviour of characters within an alignment would be highly desirable. Hopefully, this could be in the provision of a statistic that measures, in some way, character loss/change.

SUPERTREES

Despite the growth in number of phylogenetic studies, each reported study is often limited to no more than 50 included taxa. As a consequence taxon coverage between studies is often incomplete since only a few taxa may be common to each of the separate studies. One way in which to combine studies to generate a phylogeny depicting a wider taxon survey is to assemble each of the source trees into a supertree (Sanderson et al. 1998). A supertree allows inference of relationships not immediately apparent from the source trees. Generally, where the subjective and more informal methods of supertree construction fail, a variety of explicit methods now exist. For example, the relationships within source trees can be re-coded into a matrix representation and then analysed by parsimony e.g. by a semi-strict reduced consensus method (Wilkinson & Thorley 1998). In addition to the analysis of source trees derived from other DNA targets, supertrees might provide one potential way around the problem of 18S alignment. Perhaps the separate source trees from analysis of taxon sets of the 18S data could be combined, providing resolution at both broad and fine levels.

NETWORKS

Whilst the relationships between *Crithidia*, *Leishmania* and *Trypanosoma* can be adequately presented as a bifurcating tree, can such a tree depict the relationships within members of a given trypanosome clade? Since the monograph by Hoare (1972) trypanosomes were thought to reproduce exclusively by binary fission. Mainly through the use of genetic transformation technology, there has been a complete change in this perspective. Pioneering work, involving drug selective markers, showed the occurrence of non-obligatory sex and complete mating compatibility between strains of *T. brucei* s.l. (Gibson & Stevens 1999). There is also direct laboratory evidence for genetic exchange in *T. cruz*i (Stothard et al. 1999) and there is the occurrence of hybrid-like enzyme profiles for *Leishmania* e.g. amongst others putative hybrids between *L. V. panamensis* and *L. V. braziliensis* (Belli et al. 1994).

As a consequence are trypanosome trees full of reticulations – anastomoses (Maddison & Maddison 1993) – through hybridizations, sexual reproduction or genetic exchange? Methodologically this is very difficult to test exactly, given the current dogma of computer algorithms that impose strictly bifurcating structure. One potential way forward may be in the use of split decomposition and spectral analysis (Page & Holmes 1998). Such networks are useful for data exploration, competing splits that have almost equal support might indicate that a bifurcating tree is a poor representation of the data (Hendy et al. 1994).

Whilst there has been good progress in the use of DNA data to infer trypanosomatid phylogenies, there are still areas in need of further development. Better taxon coverage would be desirable, especially increasing the number of genetic loci analysed, and perhaps more emphasis upon data exploration rather than adopting the dogma of strictly bifurcating structure. The possibility of incongruence between gene trees and species trees should also not be ignored (Slowinski & Page 1999), neither should the caveat of what a species tree is really trying to represent (Maddison 1997).

ACKNOWLEDGEMENTS

Funding from The Wellcome Trust was gratefully received. Thanks are due to Jamie Stevens for his invitation to the "Workshop on Molecular Evolution of Trypanosomes".

REFERENCES

Abouheif ER, Zardoya A, Meyer A 1998. Limitations of metazoan 18S rRNA sequence data: implications for reconstructing a phylogeny of the animal kingdom and inferring the reality of the Cambrian explosion. *J Mol Evol 47*: 394-405.

Belli AA, Miles MA, Kelly JM 1994. A putative *Leishmania panamensis/Leishmania braziliensis* hybrid is a causative agent of human cutaneous leishmania-

sis in Nicaragua. *Parasitology 109*: 435-442.
Fernandes AP, Nelson K, Beverley SM 1993. Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: perspectives on the age and origins of parasitism. *Proc Natl Acad Sci USA 90*: 11608-11612.
Gibson WC, Stevens J 1999. Genetic exchange in the Trypanosmatidae. *Adv Parasitol 43*: 2-46.
Hendy MD, Penny D, Steel MA 1994. A discrete Fourier analysis for evolutionary trees. *Proc Natl Acad Sci USA 91*: 3339-3343.
Hillis DM, Dixon MT 1991. Ribosomal DNA: molecular evolution and phylogenetic evidence. *Quart Rev Biol 66*: 411-453.
Hoare CA 1972. *The Trypanosomes of Mammals – A Zoological Monograph*, Blackwell Scientific Publications, Oxford, 749 pp.
Lukes J, Jirku M, Dolezel D, Kralova I, Hollar L, Maslov DA 1997. Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic. *J Mol Evol 44*: 521-527.
Maddison WP 1997. Gene trees in species trees. *Syst Biol 46*: 523-526.
Maddison WP, Maddison DR 1993. MacClade: Analysis of Phylogeny and Character Evolution, (Computer program for Macintosh), Sinauer Associates, Sunderland, Massachusetts.
Maslov DA, Lukes J, Jirku M, Simpson L 1996. Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: Implications for the evolution of parasitism in the trypanosomatid protozoa. *Mol Biochem Parasitol 75*: 197-205.
Maslov DA, Lukes J 1998. Searching for a tree that can be trusted. *Parasitol Today 14*: 334.
Noyes HA 1998. Can trypanosome trees be trusted? *Parasitol Today 14*: 49-50.
Noyes HÁ, Rambaut A 1998. A key to understanding *Trypanosoma* trees-reply. *Parasitol Today 14*: 335.
Page RDM, Homes EC 1998. *Molecular Evolution: A Phylogenetic Approach*, Blackwell Science Ltd, Oxford, 346 pp.
Sanderon MJ, Purvis A, Henze C 1998. Phylogenetic supertrees: assembling the trees of life. *TREE 13*: 105-109.
Slowinski JB, Page RDM 1999. How should species phylogenies be inferred from sequence data. *Syst Biol 48*: 814-825.
Stevens JR, Gibson WC 1998. A key to understanding *Trypanosoma* trees. *Parasitol Today 14*: 334-335.
Stevens JR, Gibson WC 1999. The molecular evolution of trypanosomes. *Parasitol Today 15*: 432-437.
Stothard JR 2000. Trypanosome trees and homologies. *Parasitol Today 16*: 173.
Stothard JR, Frame IA, Carrasco HJ, Miles MA 1998. Temperature gradient gel electrophoresis (TGGE) analysis of riboprints from *Trypanosoma cruzi*. *Parasitology 117*: 249-253.
Stothard JR, Frame IA, Miles MA 1999. Genetic diversity and genetic exchange in *Trypanosoma cruzi*: dual drug-resistant "progeny" from episomal transformants. *Mem Inst Oswaldo Cruz 94*: 189-193.
Vickerman K 1994. The evolutionary expansion of the trypanosmatid flagellates. *Int J Parasitol 24*: 1317-1331.
Wilkinson M, Thorley JL 1998. Reduced supertrees. *TREE 13*: 283.
Wright ADG, Li S, Feng S, Martin DS, Lynn DH 1999. Phylogenetic position of the kinetoplastids, *Cryptobia bullocki*, *Cryptobia catostomi*, and *Cryptobia salmositica* and monophyly of the genus *Trypanosoma* inferred from small subunit ribosomal RNA sequences. *Mol Biochem Parasitol 99*: 69-76.