ELSEVIER

Infection, Genetics and Evolution 1 (2001) 143–150

Infection, Genetics and Evolution

www.elsevier.com/locate/meegid

Evolutionary rate differences in trypanosomes

Jamie Stevens [a,*], Andrew Rambaut [b]

[a] School of Biological Sciences, University of Exeter, EX4 4PS Exeter, UK
[b] Department of Zoology, University of Oxford, OX1 3PS Oxford, UK

Received 25 May 2001; received in revised form 4 September 2001; accepted 4 September 2001

Abstract

Ribosomal RNA-based studies of trypanosome phylogenies have highlighted considerable differences in genetic diversity within clades in the genus *Trypanosoma* and several-fold substitution rate differences between clades have been identified. While early 18S rRNA-based studies were hampered by highly variable substitution rates and long-branch attraction, it is apparent that genuine differences in evolution rates within localized clades do exist and questions remain regarding what rate or rates such clades are evolving at and why is the application of a single clock to trypanosome evolution so inappropriate? In this study, we explore rate heterogeneity in the commonly used 18S rRNA gene across genus *Trypanosoma*, using a maximum likelihood (ML) approach to explore local rate variations in clades of biological interest.
© 2001 Elsevier Science B.V. All rights reserved.

Keywords: *Trypanosoma brucei*; *Trypanosoma cruzi*; Evolution; Phylogenetics; Maximum likelihood; Small subunit ribosomal RNA; Biogeography

1. Introduction

Recent 18S ribosomal RNA-based phylogenetic studies of trypanosomes indicate considerable differences in genetic diversity within clades and, in particular, the apparently high rate of sequence evolution within *Trypanosoma brucei* and related tsetse-transmitted species (Lukes et al., 1997; Haag et al., 1998; Stevens et al., 1998, 1999). Significantly, these studies were the first to include sufficient taxa to allow meaningful sub-clades within the genus to be defined and explored. These were also the first 18S-based trees to confirm that the genus is monophyletic, in agreement with studies based on other genes (e.g. Alvarez et al., 1996; Adjé et al., 1998; Hannaert et al., 1998). The significance of this latter finding cannot be overstated, as it is only when the 'correct' tree topology has been defined, that meaningful comparisons and exploration of different sub-groupings within a tree can be undertaken.

Lukes et al. (1997) suggested that paraphyly of the genus *Trypanosoma*, as reported in a number of previous studies (e.g. Maslov et al., 1996), may have been the result of a high rate of nucleotide substitution and an associated high level of homoplasy in the 18S sequences of *T. brucei* and outgroup species. Later, a study by Haag et al. (1998) identified several-fold substitution rate differences between clades, with the rate in *T. vivax* being up to three times that in lineages leading to certain non-Salivarian taxa. Moreover, Haag et al. (1998) identified Salivarian trypanosomes as generally more distant from their outgroup species than were other taxa, while within the Salivaria, *T. vivax* was consistently more distant (by several percent sequence divergence) from the outgroup taxa than were other Salivarian species. Indeed, to reduce possible long-branch effects associated with such divergent taxa, Haag et al. (1998) ran their analyses excluding various rapidly evolving lineages, e.g. *T. vivax*. Similarly, and despite a paraphyletic conclusion, a study by Noyes (1998) again showed that *T. brucei* appears to have been pulled towards various outgroup taxa by a high, but unconnected in evolutionary terms, level of substitutions, i.e. due to homoplasy. Further to this, Noyes and Rambaut (1998), suggested the 'wrong' tree presented in Noyes (1998) may have been due to an increase in the rate of evolution in the Salivaria by as much as a factor of eight, a value confirmed by Stevens et al. (1999) in a broad-scale phylogenetic analysis of more than 50 trypanosomatid taxa.

Molecular phylogenetic studies have frequently relied on molecular clocks, calibrated by a variety of methods (e.g. Lake et al., 1988; Fernandes et al., 1993; Haag et al., 1998). However, given the controversy associated with such clocks (Sibley and Ahlquist, 1984; Wilson et al., 1987), Stevens et al. (1999) estimated a date of divergence of *T. brucei* and *T. cruzi* on the basis of biogeography and clade taxon composition (Wiley, 1988). In contrast to the majority of clock-based estimates, this approach yielded a phylogenetic

* Corresponding author. Tel.: +44-1392-263775; fax: +44-1392-263700.
E-mail address: j.r.stevens@exeter.ac.uk (J. Stevens).

1567-1348/01/$ – see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S1567-1348(01)00018-1

interpretation with biological relevance, which could be independently corroborated by reference to known biogeographical and geological events.

Thus, it appears that the majority of early 18S-based studies which found the genus *Trypanosoma* to be paraphyletic were the result of (i) highly variable substitution rates, both within the genus and between ingroup and outgroup taxa, (ii) inappropriate (simplistic) models of molecular evolution and (iii) a lack of suitable taxa to break up long branches. Nevertheless, from the phylogenies produced by various researchers, it is apparent that trends in evolutionary rate within localized clades exist and that related taxa do appear to be evolving at locally (in terms of phylogenetic proximity) similar rates. So, what rate(s) are such clades evolving at and why is the application of a single clock to trypanosome evolution so inappropriate? Surprisingly, while previous papers allude to large difference in substitution rates and branch lengths, to date, no formal analysis of these differences has been undertaken. Accordingly this paper sets out to characterize (i) the degree of heterogeneity of evolutionary rates operating amongst clades within genus *Trypanosoma*, (ii) the relative evolutionary rates of particular clades, and (iii) the date of divergence between the *Schizotrypanum* clade and an Australian kangaroo trypanosome, accounting for the rate differences amongst clades; the resolution and implications of this latter point are of considerable interest within the broader field of trypanosomatid evolution (Machado and Ayala, 2001; Stevens et al., 2001).

2. Materials and methods

2.1. Alignment of sequences

The sequence alignment of 70 taxa presented in Stevens et al. (2001), supplemented with three additional outgroup species (*Bodo saltans* AF208887, *Cruzella marina* AF208878, *Cryptobia helicis* AF208880; Dolezel et al., 2000) was used as the starting point for analyses; alignment available from http://evolve.zoo.ox.ac.uk/data/. For computational and statistical reasons, identical taxa and taxa that varied by ≤ 5 base pairs ($<0.25\%$ sequence divergence) were removed from the analysis leaving a reduced dataset of 44 taxa (see Table 1).

2.2. Constraints

Analyses were, thus, performed on a reduced number of taxa ($n = 44$), a factor which has been implicated as one of the main reasons for recovering incorrect tree topologies (Lukes et al., 1997) in previous—unlike this study—fundamentally systematic studies. Accordingly, constrained phylogenetic analyses were undertaken to ensure that taxa, which previously grouped together in ‘genuine’ well-supported clades (as defined in Stevens et al., 1999, 2001), remained together in the reduced taxa rate analyses.

Table 1
Taxon, 18S rRNA sequence and clade complement details[a]

Species	Acc. no.	Clade code
T. cobitis	AJ009143	1
T. sp. [K&A—*Piscicola geometra*]	AJ009167	1
T. boissoni	U39580	1
T. triglae	U39584	1
T. binneyi	AJ132351	1
T. rotatorium	AJ009161	1
T. therezieni	AJ223571	1
T. mega	AJ009157	1
T. b. gambiense	AJ009141	2
T. congolense—Savannah	U22315	2
T. congolense—Kilifi	U22317	2
T. congolense—Riverine-forest	AJ009145	2
T. congolense—Kilifi	AJ009144	2
T. simiae—Tsavo*	U22318	2
T. simiae	AJ009162	2
T. godfreyi	AJ009155	2
T. vivax	U22316	2
T. cruzi	M31432	3
T. cruzi marinkellei	AJ009150	3
T. rangeli	AJ009160	3
T. dionisii	AJ009151	3
Trypanosoma sp. [*Rousettus aegyptiacus*]	AJ012418	3
T. vespertilionis	AJ009166	3
T. conorhini	AJ012411	3
Trypanosoma sp. [H25—*M. giganteus*]	AJ009168	3
T. lewisi	AJ009156	NC
T. avium	AJ009140	NC
T. *bennetti*	AJ223562	NC
T. grayi	AJ005278	NC
T. varani	AJ005279	NC
T. scelopori	U67182	NC
T. theileri	AJ009163	NC
T. cyclops	AJ131958	NC
Endotrypanum monterogeii	X53911	4
Leishmania amazonensis	X53912	4
Leishmania guyanensis	X53913	4
Crithidia oncopelti	L29264	4
Trypanosoma sp. (H26—*V. ursinus*)	AJ009169	+
T. pestanai	AJ009159	+
Cryptobia helicis	AF208880	+
Cruzella marina	AF208878	+
Bodo caudatus	X53910	+
Bodo saltans	AF208887	+
Trypanoplasma borreli	L14840	+

[a] Clade code: numbers 1–4 signify the clade to which certain taxa were constrained, based on the results of previous large-scale analyses (Stevens et al., 1999, 2001); NC: not constrained to a specific clade; +: not included in rate analyses.
* See Gibson et al. (2001).

Overall, 29 of the 44 taxa were constrained together in four clades (Table 1).

2.3. Phylogenetic analysis

Preliminary starting trees for the maximum likelihood (ML) rate analysis were obtained by parsimony, which produced 14 most parsimonious trees (MPTs) of length 1624, each of which showed the same basic topology and clade

Table 2
Tree likelihood statistics[a]

Model	ln-likelihood	fps	2 × LR	*P*-value
1 rate model	−7349.54	36	29.66	5.15×10^{-8}
2 rate model (Salivarian/the rest)	−7334.71	37	89.06	3.83×10^{-21}
3 rate model[(a)] (Salivarian/aquatic/the rest)	−7290.18	38	84.88	3.17×10^{-20}
3 rate model[(b)] (Salivarian/*T. vivax*/the rest)	−7292.27	38	20.50	5.96×10^{-6}
4 rate model (Salivarian/aquatic/*T. vivax*/the rest):	−7279.93	39	135.00	1.28×10^{-14}
Non-clock model	−7212.41	71		

[a] fps = number of free parameters in model—see Section 2.5; LR: likelihood ratio. Likelihood ratio values for both 3 rate models are calculated with reference to the 2 rate model.

complement with respect to four major clades: Salivarian, aquatic, *Schizotrypanum* and others, *Leishmania*. Following the construction of preliminary starting trees, five outgroup taxa which had been included to provide a robust estimate of the root, but which were superfluous to the analysis of rates of molecular evolution in *Trypanosoma* were removed, leaving *Endotrypanum monterogeii*, *Leishmania amazonensis*, *L. guyanensis* and *Crithidia oncopelti* as outgroup taxa. Two isolated and unsupported, long-branch taxa (*Trypanosoma* sp. [H26—*Vombatus ursinus*] and *T. pestanai*—see Stevens et al., 1999) were also removed, leaving 33 ingroup taxa; the capacity of isolated, rapidly evolving taxa to artefactually affect evolutionary rate analysis is well documented (Haag et al., 1998). For the remaining 37 taxa, the maximum likelihood tree was obtained with a heuristic search in PAUP* version 4b4a (Swofford, 2001), using the general time-reversible model of substitution with gamma-distributed rate heterogeneity between sites (Yang, 1994); the general time-reversible model with rate heterogeneity was chosen as it has the most parameters and would, therefore, give the largest confidence intervals in subsequent analyses.

2.4. Estimation of rate heterogeneity between lineages

Analysis of the variation in the rate of molecular evolution across genus *Trypanosoma* was performed using the program RHINO (Rambaut, 2001; http://evolve.zoo.ox.ac.uk). This program allows the estimation of relative rates of substitution between specified lineages of a tree ("local clocks") within a maximum likelihood framework in the manner described by Rambaut and Bromham (1998) and Yoder and Yang (2000). By specifying a priori clades that potentially have differences in rate, a hierarchy of competing models of rate heterogeneity was tested. Analysis of the main clades (Salivarian, aquatic) was performed both with and without *T. vivax*, allowing a general Salivarian rate, uninfluenced by the unusually high rate observed in the *T. vivax* lineage, to be calculated. For all these analyses, the phylogeny and substitution model used was that estimated by the maximum likelihood tree search, above.

First, a standard ML analysis was performed in which all branch lengths were free to vary independently (71 independent rates, free parameters (fps) = 71).[1] This is the unconstrained model which makes no molecular clock assumptions (non-clock model, Table 2). In successive analyses, taxa were constrained together in groups assigned specific rate parameters (Fig. 1) on the basis of a priori biological knowledge (Stevens et al., 1999, 2001):

- *Rate model*: All 37 taxa were constrained to evolve at a single rate, fps = 36.
- *Rate model (Salivarian/others)*: The nine taxa within the Salivarian clade (Fig. 1) were constrained to evolve at one rate, while the remaining 28 taxa were assigned a separate rate, fps = 37.
- *Rate model[(a)] (Salivarian/aquatic/others)*: The nine taxa within the Salivarian clade (Fig. 1) were assigned one rate, the eight taxa within the aquatic clade (Fig. 1) were assigned another rate, the remaining 20 taxa were assigned a third rate, fps = 38.
- *Rate model[(b)] (Salivarian/T. vivax/others)*: The eight taxa within the Salivarian clade (excluding *T. vivax*) were assigned one rate, the clade containing the single taxon—*T. vivax*—was assigned another rate, the remaining 28 taxa were assigned the third rate, fps = 38.
- *Rate model (Salivarian/aquatic/T. vivax/others)*: The eight taxa within the Salivarian clade (excluding *T. vivax*) were assigned a rate, the clade containing the single taxon—*T. vivax*—was assigned a second rate, the eight taxa within the aquatic clade (Fig. 1) were assigned a third rate, the remaining 20 taxa were assigned a fourth rate, fps = 39.

2.5. Testing the fit of the multi-rate models

Each of the models above can be compared using the likelihood ratio test. When one model is a specific case of another (nested), twice the difference in log-likelihood for the two models is expected to be χ^2 distributed with degrees of freedom equal to the difference in the number of free parameters. The models outlined above are nested, because the models with fewer rates are specific cases of those with

[1] The difference in free parameters between models provides the degrees of freedom when testing the significance of the difference in fit (χ^2) between models (see Section 2.5).

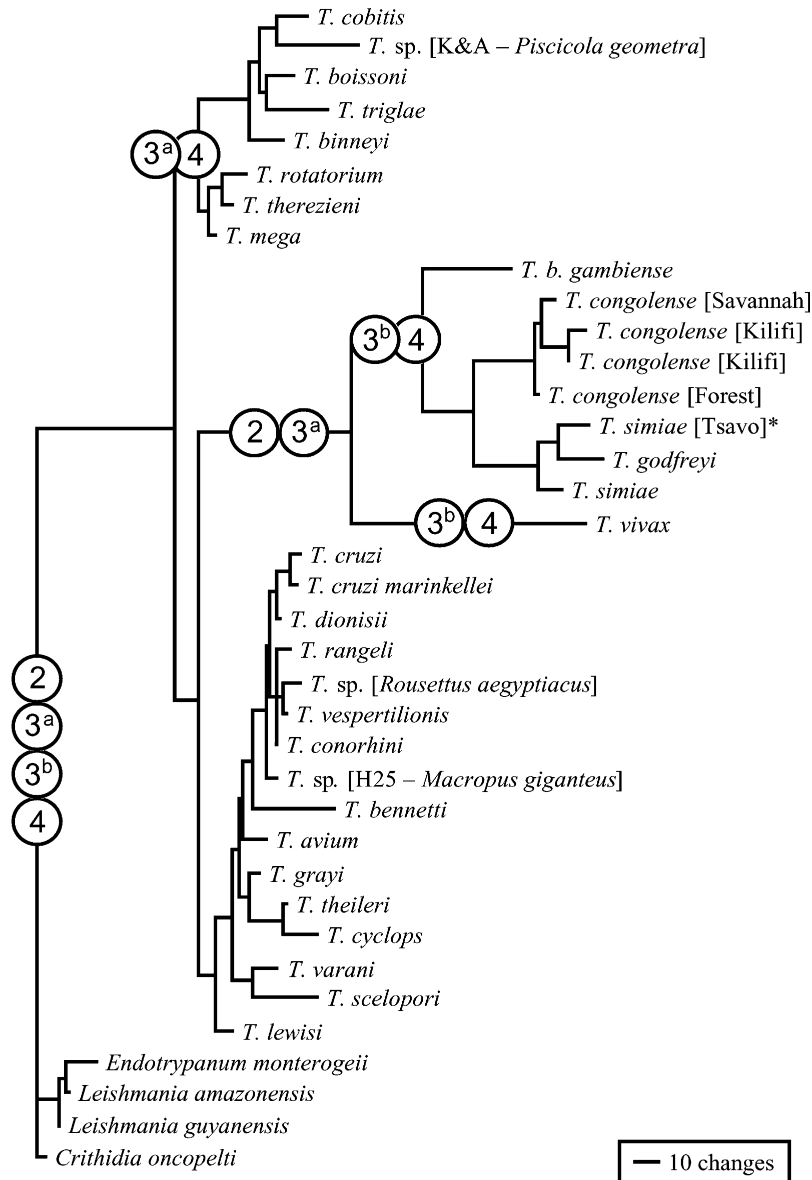

Fig. 1. Assignment of taxa to independent rate parameter groups (clades) for each of the four multi-rate models tested. Clades within the constrained phylogenetic tree were identified on the basis of a priori knowledge (see text). Numbers given in circles (2, $3^{(a)}$, $3^{(b)}$, 4) indicate the multi-rate model for which a particular clades has been assigned an independent rate parameter. For unidentified *Trypanosoma* species, the isolate identity and host are given, e.g. *Trypanosoma* sp. [H25—*M. giganteus*] = isolate H25, Eastern gray kangaroo. *T. simiae* [Tsavo]*, see Gibson et al. (2001).

more rates when the rates are equal. Thus, starting with the 1 rate model, we can test whether the 2 rate model fits the data significantly better by comparing $2[\ell(2 \text{ rate model}) - \ell(1 \text{ rate model})]$ against the appropriate χ^2 critical value with 1 d.f. (where ℓ(model) is the log-likelihood of a particular model).

2.6. Dating the kangaroo trypanosome divergence

For each of the single or multi-rate models described above, it is possible to obtain an estimate of the date of a particular node in the tree by reference to a calibration node of known date. Such a calibration allows the estimation of the overall rate of substitution for the entire tree (for the multi-rate models, the rates for the specified clades will be relative to this overall rate). Confidence intervals for the date of divergence of a particular node can be obtain using standard maximum likelihood procedures (e.g. Rambaut and Bromham, 1998). Of course, our primary interest was in characterizing rRNA rate variation in *Trypanosoma* sp. However, the estimates of divergence dates are contingent on the model (and the calibration being correct) and highlight the error that can occur if rate variation is not accounted for in such analyses.

Table 3

Clade rates
2 rates (Salivarian/the rest)
Rate for rest of tree: 2.64×10^{-10}[a]
Relative rate of Salivarian clade: 4.22×10^{-10}
3 rates[(a)] (Salivarian/aquatic/the rest)
Rate for rest of tree: 1.96×10^{-10}
Relative rate of Salivarian clade: 5.36×10^{-10}
Relative rate of aquatic clade: 3.05×10^{-10}
3 rates[(b)] (Salivarian/*T. vivax*/the rest)
Rate for rest of tree: 2.55×10^{-10}
Relative rate of Salivarian clade: 3.64×10^{-10}
Relative rate of *T. vivax*: 7.72×10^{-10}
4 rates (Salivarian/aquatic/*T. vivax*/the rest)
Rate for rest of tree: 1.96×10^{-10}
Relative rate of Salivarian clade: 4.57×10^{-10}
Relative rate of aquatic clade: 3.07×10^{-10}
Relative rate of *T. vivax*: 9.67×10^{-10}

[a] Rate of base substitutions per site per year.

In order to date the split of the kangaroo trypanosome (*Trypanosoma* sp. [H25—*Macropus giganteus*], AJ009168) from the South American and bat trypanosomes with which it clusters, evolutionary rates within trees were calibrated by fixing the date of divergence of the Salivarian clade (Fig. 1) from all other taxa at 100 million years. This date was selected based on geological (Smith et al., 1994), biogeographical (Cox and Moore, 2000) and parasitological data (Stevens et al., 1999).

3. Results

Rate analysis using the program RHINO yielded a series of likelihood scores for each set of rate parameters imposed on the starting tree (Table 2). The addition of each successive rate to a clade resulted in a significant improvement (increase) in ln-likelihood value as demonstrated by successive likelihood ratio tests (Table 2). This result indicates that for the taxa within genus *Trypanosoma* their 18S rRNA genes are evolving at least four significantly different rates.

Analysis of specific clade evolution rates (Table 3) shows that for the 18S rRNA sequences studied, Salivarian trypanosomes are evolving at more than four times the rate of other taxa (2 rate analysis, Table 3). Analysis of the data constrained to three rates shows aquatic and Salivarian trypanosomes to be evolving three and five times faster than other taxa, respectively (3 rate analysis), while a second three rate analysis shows *T. vivax* to be evolving at more than twice (7.7:3.6) the rate of other Salivarian trypanosomes (3[(b)] rate analyses, Table 3). Indeed, *T. vivax*, which is evolving at approximately 7–10 times (3[(b)] and 4 rate analyses, Table 3) the rate of non-Salivarian trypanosomes, appears to be the single, most rapidly evolving taxon within the entire range of trypanosomatid taxa studied.

Finally, using a calibration date for the divergence of the Salivarian clade of 100 million years, we were able to estimate the timing of the divergence of the kangaroo trypanosome from the *Schizotrypanum* clade. The values estimated (with 95% confidence intervals around the divergence date) for each rate model are given in Table 4.

4. Discussion

The concept of a molecular clock was first proposed by Zuckerkandl and Pauling (1965). Since then the exact nature of the workings of such clocks have remained controversial (e.g. Sibley and Ahlquist, 1984; Wilson et al., 1987; Martin and Palumbi, 1993). It is apparent that, if they do exist, they are at best only stochastically constant (Fitch, 1976), and that different types of DNA sequence undoubtedly evolve at significantly different rates. Nevertheless, within given taxonomic groups and defined categories of genetic marker, the concept of a molecular clock can provide a useful tool for understanding phylogenetic relationships. For trypanosomatids, as for many organisms, the assumption of a single molecular clock for all taxa is too simplistic; nevertheless, comparisons of phylogenetic branch lengths do suggest the possible existence of clade-specific clocks, each of which is ‘ticking’ at a distinct rate.

Maximum likelihood-based rate analyses of a tree containing 37 representative *Trypanosoma* and appropriate outgroup taxa (Table 2) indicates that within genus *Trypanosoma* at least four significantly different evolutionary rates are operating on the 18S rRNA gene, each of which is associated with a particular phylogenetic clade (as defined in Fig. 1). Of course, clades identified in the current study were selected on the basis of a priori biological characteristics (see Stevens et al., 2001) and with regard to current computational constraints. In future, the testing of all

Table 4
Divergence estimates for the Australian kangaroo trypanosome[a]

Model	Divergence date (million year)	95% CI (million year)
1 rate model	9.5	–
2 rate model (Salivarian/the rest)	19.7	(12.6–30.8)
3 rate model[(a)] (Salivarian/aquatic/the rest)	25.0	(15.7–39.5)
3 rate model[(b)] (Salivarian/*T. vivax*/the rest)	20.4	(13.1–31.6)
4 rate model (Salivarian/aquatic/*T. vivax*/the rest)	25.5	(16.1–40.0)

[a] Confidence intervals could not be calculated for the divergence estimate produced with the single rate model. Due to computational constraints, divergence estimate and confidence intervals were not calculated for the Non-clock model.

possible combinations of clades and taxa to determine the exact number of significantly different rates by which a given gene is evolving should be possible. For the moment, we have demonstrated that the use of a single molecular clock is inappropriate to describe the rate of evolution of trypanosomes, even within such biologically distinct entities as the Salivarian (African tsetse-transmitted) trypanosomes.

Analysis of relative clade rates shows the Salivarian trypanosomes to be the most rapidly evolving group of taxa analyzed (Table 3). However, as shown in the $3^{(b)}$ and 4 rate analyses (Table 3), not only is *T. vivax* 18S rRNA evolving at approximately 7–10 times the rate of non-Salivarian trypanosomes, but it is also evolving significantly faster than all other Salivarian trypanosome sequences. Indeed, it appears to be the single most rapidly evolving taxon within the entire range of trypanosomatid taxa studied; such a result is in accordance with the findings of Haag et al. (1998) and concurs with the apparently unusual evolutionary history (Gardiner, 1989) and molecular profile (Dickin and Gibson, 1989) of *T. vivax*. Reasons for such rapid evolution in this single taxon are unclear and may be biological, artefactual or a combination of both. For example, as reiterated by Stothard (2000), while the assumption of sequence heterogeneity between rRNA genes is widely accepted because concerted evolution is thought to act in all eukaryotes, there is a growing list of examples, including several *Trypanosoma* species, where complete homogeneity is not always the case, e.g. Stothard et al. (1998). Accordingly, we cannot rule out the possibility that the single *T. vivax* 18S rRNA sequence employed in these (and numerous other) phylogenetic analyses is either a pseudogene, or another form of divergent copy which is not being homogenized by concerted evolution. Additionally, *T. vivax* appears unique in having an 18S rRNA G + C content of 55.4%, a figure some 3% higher than any other trypanosomatid included in the study of Haag et al. (1998); as indicated by Hasegawa and Hashimoto (1993), large differences in G+C content can adversely affect phylogenetic reconstruction, although recent parsimony-based simulation studies by Conant and Lewis (2001) suggest that convergence in nucleotide composition alone is insufficient to cause any commonly used methods to fail.

The estimates of divergence date should be considered with caution because even the 4 rate model was a significantly worse fit than the unconstrained model. However, the analysis serves to highlight influence of clade-specific rate heterogeneity on estimates of divergence dates. There is a relatively minor influence of a single rapidly evolving taxon on overall estimates of clade rates and divergence dates. For example, going from the 2 rate model to the $3^{(b)}$ model gives very little difference in the estimated date of divergence of the kangaroo trypanosome (19.7 million years versus 20.4 million years), while going from the $3^{(a)}$ model to the 4 rate model also makes very little difference to the estimated divergence date (25.0 million years versus 25.5 million years); it is the addition of a complete clade rate that makes the biggest difference in estimated divergence dates (i.e. going from 1 rate to 2 rates with the addition of the Salivarian clade or going from the 2 rate model to $3^{(a)}$ model, or from the $3^{(b)}$ model to the 4 rate model with the addition of the aquatic clade).

Unfortunately, all published molecular-based phylogenetic analyses of *T. vivax* to date (Lukes et al., 1997; Haag et al., 1998; Stevens et al., 1999) are based on 18S and 28S large subunit ribosomal RNA sequences (EMBL Acc. no. U22316; Urakawa and Majiwa, unpublished results) from a single isolate, the much studied rodent-adapted Y486 stock, which was originally isolated in Nigeria from a Zebu cow (*Bos indicus*) in 1973. In all analyses, however, a similar pattern of results is observed. The *T. vivax* lineage shows a greatly elevated substitution rate within its rRNA sequences, which prompted Haag et al. (1998) to exclude *T. vivax* from their final phylogenetic analyses. *T. vivax* is the first taxon to diverge from the monophyletic clade of Salivarian trypanosomes in phylogenetic analyses (Lukes et al., 1997; Stevens et al., 1999). The well-supported early separation of the *T. vivax* lineage is consistent with the view that *T. vivax* represents the most ancient of the Salivarian trypanosomes (Haag et al., 1998) and that *T. vivax* represents an early stage of adaptation to transmission by tsetse flies (Lukes et al., 1997), both hypotheses having originally been proposed by Hoare (1972). Clearly, phylogenetic analysis of freshly isolated strains of *T. vivax* is required.

Finally, having assessed the extent of evolutionary rate differences within the phylogenetic tree, we shall focus on the divergence date between the Australian kangaroo trypanosome and the *Schizotrypanum* species; the timing of evolutionary divergence events within subgenus *Schizotrypanum* have recently been the subject of considerable debate (e.g. Stevens et al., 1999, 2000, 2001; Gaunt and Miles, 2000; Machado and Ayala, 2001). Using a calibration date of 100 million years for the divergence of the African Salivarian clade, the program estimated divergence dates at 20–25 million years (excluding the single rate model estimate of <10 million years), with 95% confidence values ranging between 12 and 40 million years (Table 4). However, geological (Smith et al., 1994) and biogeographical (Cox and Moore, 2000) data indicate that Australia split away finally from Antarctica around 45 million years, this presumably being the last time that these taxa could have shared a common ancestor (or host/vector) and when all interaction between South American and Australian fauna effectively ceased. While our confidence limits (12.6–40 million years depending on model) approach this value (45 million years), our primary estimates are only around half of that which might be expected by direct comparison with geological/biogeographical evidence. While such a discrepancy may indicate a bias in our methodology and/or calibration dates, such an underestimate may be an indication of evolutionary saturation and the limitations of the 18S rRNA gene over this time scale in trypanosomatids (Stothard, 2000). A recent study by Machado and Ayala (2001) using

protein gene sequences (trypanothione reductase (TR), cytochrome oxidase subunit II/NADH dehydrogenase subunit 1 (COII–ND1)) estimates the age of the most recent common ancestor of *T. cruzi* and bat trypanosomes,[2] which equates to the node immediately below the kangaroo trypanosome (*Trypanosoma* sp. [H25—*M. giganteus*])/*T. cruzi* node in our tree (Fig. 1), at 6.2–10.1 million years (TR) and 13.1–20.1 million years (COII–ND1), using *T. brucei* as the outgroup. Additionally, such an apparent underestimate of divergence time may reflect a delay due to the close association of these parasites with their vector species, although the antiquity of the host–parasite relationship between triatomine bugs and other New World species of *Trypanosoma* remains much in debate (Schofield, 2000; Gaunt and Miles, 2000).

Overall, the differences in evolutionary rates between trypanosome clades demonstrated in this paper are in keeping with results from previous studies (e.g. Maslov et al., 1996; Haag et al., 1998) and, while there are further rate differences within the clades to examine (even the 4 rate model is rejected when compared to the non-clock model), for the first time we have been able to characterize patterns of evolutionary rate heterogeneity between trypanosomatids sequences. Perhaps most importantly, this study provides a formal demonstration of the significant limitations of using single rate molecular clocks when extrapolating evolutionary events from phylogenetic trees.

Acknowledgements

The financial support of The Wellcome Trust is gratefully acknowledged. J.S. is funded by a Wellcome Trust Biodiversity Fellowship (grant no. 050808/Z/97/Z); A.R. is supported by Wellcome Trust grant no. 50275. Thanks to Wendy Gibson and Harry Noyes for helpful comments on the manuscript.

References

Adjé, C.A., Opperdoes, F.R., Michels, P.A.M., 1998. Molecular analysis of phosphoglycerate kinase in *Trypanoplasma borreli* and the evolution of this enzyme in Kinetoplastida. Gene 217, 91–99.

Alvarez, F., Cortinas, M.N., Musto, H., 1996. The analysis of protein coding genes suggests monophyly of *Trypanosoma*. Mol. Phylogen. E 5, 333–343.

Conant, G.C., Lewis, P.O., 2001. Effects of nucleotide composition bias on the success of the parsimony criterion in phylogenetic inference. Mol. Biol. E 18, 1024–1033.

Cox, C.B., Moore, P.D., 2000. Biogeography: An Ecological and Evolutionary Approach, 6th Edition. Blackwell, Oxford.

Dickin, S.K., Gibson, W.C., 1989. Hybridisation with a repetitive DNA probe reveals the presence of small chromosomes in *Trypanosoma vivax*. Mol. Biochem. Parasitol. 33, 135–142.

[2] Including *T. vespertilionis.*

Dolezel, D., Jirku, M., Maslov, D.A., Lukes, J., 2000. Phylogeny of the bodonid flagellates (Kinetoplastida) based on small-subunit rRNA gene sequences. Int. J. Syst. Evol. Microbiol. 50, 1943–1951.

Fernandes, A.P., Nelson, K., Beverley, S.M., 1993. Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: perspectives on the age and origins of parasitism. Proc. Natl. Acad. Sci. U.S.A. 90, 11608–11612.

Fitch, W.M., 1976. Molecular evolutionary clocks. In: Ayala, F.J. (Ed.), Molecular Evolution, Sinauer, Sunderland, MA, pp. 160–178.

Gardiner, P.R., 1989. Recent studies of the biology of *Trypanosoma vivax*. Adv. Parasitol. 28, 229–317.

Gaunt, M., Miles, M., 2000. The ecotopes and evolution of triatomine bugs (Triatominae) and their associated trypanosomes. Mem. Instit. Oswaldo Cruz 95, 557–565.

Gibson, W.C., Stevens, J.R., Mwendia, C.M.T., Makumi, J.N., Ngotho, J.M., Ndung'u, J.M., 2001. Unravelling the phylogenetic relationships of African trypanosomes of suids. Parasitology 122, 625–631.

Haag, J., O'hUigin, C., Overath, P., 1998. The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria. Mol. Biochem. Parasitol. 91, 37–49.

Hannaert, V., Opperdoes, F.R., Michels, P.A.M., 1998. Comparison and evolutionary analysis of the glycosomal glyceraldehyde-3-phosphate dehydrogenase from different Kinetoplastida. J. Mol. E 47, 728–738.

Hasegawa, M., Hashimoto, T., 1993. Ribosomal RNA trees misleading? Nature 361, 23.

Hoare, C.A., 1972. The Trypanosomes of Mammals: A Zoological Monograph. Blackwell, Oxford, p. 749.

Lake, J.A., De La Cruz, V.F., Ferreira, P.C.G., Morel, C., Simpson, L., 1988. Evolution of parasitism: kinetoplastid protozoan history reconstructed from mitochondrial rRNA gene sequences. Proc. Natl. Acad. Sci. U.S.A. 85, 4779–4783.

Lukes, J., Jirku, M., Dolezel, D., Kral'ova, I., Hollar, L., Maslov, D.A., 1997. Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic. J. Mol. E 44, 521–527.

Machado, C.A., Ayala, F.J., 2001. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. Proc. Natl. Acad. Sci. U.S.A. 98, 7396–7401.

Martin, A.P., Palumbi, S.R., 1993. Body size, metabolic rate, generation time and the molecular clock. Proc. Natl. Acad. Sci. U.S.A. 90, 4087–4091.

Maslov, D.A., 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.

Noyes, H., 1998. Can *Trypanosoma* trees be trusted? Parasitol. Today 14, 49–50.

Noyes, H.A., Rambaut, A., 1998. Reply. Parasitol. Today 14, 335.

Rambaut, A., 2001. RHINO: a program to estimate relative rates of substitution between specified lineages of a phylogenetic tree within a maximum likelihood framework, http://evolve.zoo.ox.ac.uk.

Rambaut, A., Bromham, L., 1998. Estimating divergence dates from molecular sequences. Mol. Biol. Evol. 15, 442–448.

Schofield, C.J., 2000. *Trypanosoma cruzi*—the vector–parasite paradox. Mem. Instit. Oswaldo Cruz 95, 535–544.

Sibley, C.G., Ahlquist, J.E., 1984. The phylogeny of the hominoid primates as indicated by DNA–DNA hybridization. J. Mol. Evol. 20, 22–25.

Smith, A.G., Smith, D.G., Funnell, B.M., 1994. Atlas of Mesozoic and Cenozoic Coastlines. Cambridge University Press, Cambridge.

Stevens, J., Noyes, H., Gibson, W., 1998. The evolution of trypanosomes infecting humans and primates. Mem. Instit. Oswaldo Cruz 93, 669–676.

Stevens, J.R., Noyes, H.A., Dover, G.A., Gibson, W.C., 1999. *Trypanosoma brucei* and *T. cruzi*. Parasitology 118, 107–116.

Stevens, J., Gibson, W., Noyes, H., 2000. Trypanosome evolution under the microscope. Parasitol. Today 16, 270–271.

Stevens, J.R., Noyes, H.A., Schofield, C.J., Gibson, W., 2001. The molecular evolution of Trypanosomatidae. Adv. Parasitol. 48, 1–56.

Stothard, J.R., 2000. Trypanosome trees and homologies. Parasitol. Today 16, 173.

Stothard, J.R., Frame, I.A., Carrasco, H.J., Miles, M.A., 1998. Temperature gradient gel electrophoresis (TGGE) analysis of riboprints from *Trypanosoma cruzi*. Parasitology 117, 249–253.

Swofford, D.L., 2001. PAUP*: Phylogenetic Analysis Using Parsimony (and other methods*), Version 4, Sinauer Associates, Sunderland, MA.

Wiley, E.O., 1988. Vicariance biogeography. Ann. Rev. Ecol. System. 19, 513–542.

Wilson, A.C., Ochman, H., Prager, E.M., 1987. Molecular time scale for evolution. Trends Genet. 3, 241–247.

Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. E 39, 306–314.

Yoder, A.D., Yang, Z., 2000. Estimation of primate speciation dates using local molecular clocks. Mol. Biol. E 17, 1081–1090.

Zuckerkandl, E., Pauling, L., 1965. Evolutionary divergence and convergence in proteins. In: Bryson, V., Vogel, H.J. (Eds.), Evolving Genes and Proteins, Academic Press, New York, pp. 97–166.