Molecular research and the control of Chagas disease vectors

FERNANDO ABAD-FRANCH$^{1,2}$ and FERNANDO A. MONTEIRO$^3$

$^1$Coordenação de Biodiversidade em Saúde, Entomologia Médica, Centro de Pesquisa Leônidas e Maria Deane Fiocruz-Amazônia, Rua Terezina 476, Adrianópolis, 69057-070 Manaus, AM, Brasil
$^2$Pathogen Molecular Biology Unit, Department of Infectious and Tropical Diseases London School of Hygiene and Tropical Medicine, Keppel St., London WC1E 7HT, United Kingdom
$^3$Departamento de Medicina Tropical, Instituto Oswaldo Cruz, Fiocruz, Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brasil

Manuscript received on March 3, 2005; accepted for publication on March 30, 2005; presented by LUCIA MENDONÇA PREVIATO

ABSTRACT

Chagas disease control initiatives are yielding promising results. Molecular research has helped successful programs by identifying and characterizing introduced vector populations and by defining intervention targets accurately. However, researchers and health officials are facing new challenges throughout Latin America. Native vectors persistently reinfest insecticide-treated households, and sylvatic triatomines maintain disease transmission in humid forest regions (including Amazonia) without colonizing human dwellings. In these scenarios, fine-scale vector studies are essential to define epidemiological risk patterns and clarify the involvement of little-known triatomine taxa in disease transmission. These eco-epidemiological investigations, as well as the planning and monitoring of control interventions, rely by necessity on accurate taxonomic judgments. The problems of cryptic speciation and phenotypic plasticity illustrate this need – and how molecular systematics can provide the fitting answers. Molecular data analyses also illuminate basic aspects of vector evolution and adaptive trends. Here we review the applications of molecular markers (concentrating on allozymes and DNA sequencing) to the study of triatomines. We analyze the suitability, strengths and weaknesses of the various techniques for taxonomic, systematic and evolutionary investigations at different levels (populations, species, and higher taxonomic categories).

Key words: molecular systematics, DNA, allozymes, Triatominae, Chagas disease.

CHAGAS DISEASE: BURDEN, CONTROL EFFORTS AND NEW CHALLENGES

Chagas disease (CD), caused by Trypanosoma cruzi and transmitted by triatomine bugs, is a major public health problem in Latin America. An overall prevalence of $\sim$ 18 million infections was estimated in the late 1980s, with 90 million people living at risk and 45000 deaths/year. CD was the most important parasitic disease in Latin America in terms of its impact on national economies and public health systems (WHO 1991, 2002, World Bank 1993, Miles et al. 2003).

In the absence of vaccines or adequate drugs for large-scale treatment, the reduction of disease burden critically depends on the control of transmission by triatomine vectors and infected blood transfusion. Several multinational initiatives have been launched...
with that purpose. Ten years of concerted action in the Southern Cone have resulted in the elimination of transmission by *Triatoma infestans* (the most widespread domestic vector) from vast areas of the region. Incidence dropped by an average of 94% in the area, and by 65% in Latin America (WHO 2002). Recent estimates show however that substantial efforts are still necessary: overall prevalence remains over 12 million, with 200000 new cases/year in 15 countries (Morel and Lazdins 2003).

The success of the Southern Cone initiative was eased by the fact that the main vector species, *T. infestans*, is entirely domestic throughout most of its range, making reinestation of insecticide-treated houses rare (Dias and Schofield 1999). Other initiatives (Andean Countries, Central America, Mexico, and Amazonia), launched after 1997, face distinct epidemiological situations. General trends are encouraging in that vector control programs are now active (and screening of blood donations is mandatory) in all endemic countries.

Recent studies have nonetheless revealed new eco-epidemiological scenarios posing further challenges to the control of Chagas disease. In areas were autochthonous vectors are synanthropic (e.g., *T. brasiliensis* in NE Brazil, *T. dimidiata* in Mesoamerica, or *R. prolixus* in Venezuela), longitudinal vigilance is required for the detection and spraying of reinfested dwellings (Ramsey and Schofield 2003). In the Amazon and in other humid forest regions, vector-borne transmission occurs without household colonization when sylvatic bugs invade houses or contaminate food-processing equipment. The design of adequate control strategies for those areas demands innovative approaches and will require substantial research efforts (Coura et al. 2002, Dias et al. 2001, Guhl and Schofield 2004).

In these newly recognized contexts, fine-scale studies on vector biology, ecology and behavior are essential to define epidemiological risk patterns and clarify the involvement of little-known triatomine taxa in disease transmission. Reliable taxonomic identification is obviously the keystone of any of these investigations.

**THE TAXONOMY AND SYSTEMATICS OF THE TRIATOMINAE: MORPHOLOGY-BASED APPROACHES**

Most triatomine species can be confidently identified by their external morphological-chromatic characters (Lent and Wygodzinsky 1979, Carcavallo et al. 1997a, 1999). Some groups are however problematic, with a few of them being essentially isomorphic; others present high levels of phenotypic plasticity (resulting in either convergence or divergence) that can confound taxonomic assignations (Lent and Wygodzinsky 1979, Dujardin et al. 1999b). Moreover, species descriptions based on one or a few type specimens overlook intraspecific variation, adding uncertainty in many cases.

Complementary qualitative techniques include the study of male genitalia (e.g., Jurberg 1996), cuticular structures (e.g., Carcavallo et al. 1997b), antennal sensilla (e.g., Catalá 1997) or eggshell structure (e.g., Barata 1998).

Quantitative phenotypic analyses regard the patterns of variability and similarity revealed by metric analyses as a reflection of underlying genetic relationships. Morphometric techniques based on discriminant analyses are used in the surveillance of reinestation of sprayed dwellings and in the study of phenotypic changes linked to microhabitat adaptations. Morphometric analyses may also be used to explore phylogenetic relationships among closely related entities. However, the partitioning of environment-related metric variance is problematic, obscuring the interpretation of results in terms of kinship (cf. Dujardin et al. 2002, Patterson 2002).

Phylogenetic analyses of morphological characters have traditionally (and efficiently) made use of Hennigian cladistics. Within the Triatominae however, only two cladograms (on *Panstrongylus* and the Bolboderini, both lacking explicitness of character polarization) have ever been published (Lent and Wygodzinsky 1979).

Faced with the difficulties of (1) taxonomic assignation, at least within problematic groups, and (2) establishment of systematic/evolutionary relationships (which affect several epidemiologically
important taxa), researchers have turned to molecular approaches in an attempt to clarify the taxonomic status, systematic relationships, and evolution of the Triatominae.

MOLECULAR RESEARCH ON TRIATOMINES

Molecular analyses have revolutionized biological sciences; among the many applications, the definition of molecular markers for the distinction of biological entities and the assessment of their systematic and evolutionary relationships are particularly widespread.

Vector control programs are already benefiting from knowledge derived from molecular research. Major contributions have been the identification and characterization of introduced vector populations (e.g., *R. prolixus* in Central America or *T. infestans* in many Southern Cone countries; Dujardin et al. 1998a, Panzera et al. 2004) and the accurate definition of control intervention targets (e.g., *R. prolixus* as a valid entity separate from *R. robustus*, and chromatic forms of *T. brasiliensis* constituting distinct species; Monteiro et al. 2003, 2004). Crucial questions often lay within the range of fine-grained systematics (i.e., that of intraspecific variation), but the assessment of interspecific variability and systematic/evolutionary relationships is important in many cases as it may allow for broader extrapolations regarding epidemiologically relevant groups (Schaefer 2003).

Here we present an overview of the use of molecular markers in triatomine research. We also comment on the applicability, strengths and weaknesses of the main molecular techniques (Tables I and II), give suggestions for future research, and highlight important issues to be considered during research planning and experimental design (Table III).

We lay emphasis on the idea that correct identification of the bugs is crucial for a trustworthy understanding of vector biology. All ecological, behavioral, physiological, epidemiological or evolutionary investigations, as well as the planning and monitoring of control interventions, rely by necessity on accurate taxonomic judgments. At the same time, the analysis of molecular datasets illuminates more basic questions concerning the kinship, evolution, and adaptive trends of the vectors. We consider the use of different techniques for the analysis of genetic material (from caryotyping to DNA sequence analysis) and gene products (mainly allozymes) in the study of populations, species, and deeper phyletic clades within the Triatominae. We concentrate on those methods that have yielded a greater wealth of information, and especially on those that produced key knowledge for the design of rational disease control strategies.

Allozyme Electrophoresis

The analysis of the electrophoretic properties of allozymes (enzymes with identical function but distinct electrophoretic migration patterns encoded by different alleles of the same locus) has been extensively applied to the study of Triatominae. This approach has for decades represented the first choice to examine the molecular taxonomy and evolutionary relationships of many different organisms, including insects (Thorpe and Solé-Cava 1994, Loxdale and Lushai 1998, Dujardin et al. 2002). Intraspécific level analyses allow for (1) the assessment of the degree of genetic variability within a species or population (as the expected number of heterozygous individuals across loci – $H_e$), for (2) the detection of within-population departures from the Hardy-Weinberg [H-W] equilibrium (through the observation of the excess or deficit of heterozygotes – $F_{IS}$), and (3) for the estimation of levels of genetic structuring (measured as the differentiation between populations based upon the interpopulation component of the total genetic variation – $F_{ST}$). Finally, (4) rates of gene flow (i.e. migration) between putative populations can be estimated (Thorpe and Solé-Cava 1994, Frías and Dujardin 1996, Loxdale and Lushai 1998, Dujardin et al. 2002).

Allozymes are adequate markers for the study of intraspecific variation in most insect vector groups (Loxdale and Lushai 1998). However, triatomines...
TABLE I
Applicability of often-used molecular techniques to problems in systematics. Modified from Hillis et al. (1996).

<table>
<thead>
<tr>
<th></th>
<th>Allozymes</th>
<th>Cytogenetics</th>
<th>PCR-RFLP</th>
<th>RAPD</th>
<th>Microsatellites</th>
<th>DNA sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterozygosity</td>
<td>+</td>
<td>−</td>
<td>≈</td>
<td>−</td>
<td>+</td>
<td>+ ($)</td>
</tr>
<tr>
<td>Population subdivision</td>
<td>+</td>
<td>≈</td>
<td>≈</td>
<td>≈</td>
<td>+</td>
<td>+ ($)</td>
</tr>
<tr>
<td>Geographic variation</td>
<td>+</td>
<td>≈</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+ ($)</td>
</tr>
<tr>
<td>Species boundaries</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Phylogeny</td>
<td>≈</td>
<td>−</td>
<td>≈</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
</tbody>
</table>

Key: −, inappropriate use of technique; ≈, marginally appropriate or appropriate under limited circumstances; + ($), appropriate but expensive; +, appropriate and effective method.

in particular tend to present very low levels of allozyme variability (i.e. polymorphic loci), which poses certain limitations to the use of this marker for population level studies. Nevertheless, some studies have succeeded in finding enough variation to allow for inferences on population structure to be made.

*T. infestans* populations, for instance, are panmictic within Andean villages in Bolivia and Peru, but strongly structured (and conforming to a model of isolation by distance) among localities. Analysis of allelic frequencies allowed for the definition of the probable area of origin of *T. infestans* (Dujardin et al. 1998b). Further studies indicated that structuring within localities was also present, suggesting that the basic population unit is represented by individual households (Breniére et al. 1998). Using a similar approach, Noireau et al. (1999b) showed that the panmictic unit of Bolivian populations of *T. sordida* (allozymic group 1) is larger than that reported for *T. infestans*, with departures from H-W equilibrium detected only between populations located over ∼ 20km apart; this suggested a better dispersal capacity than previously thought for *T. sordida* (Noireau et al. 1999b).

Moreover, low levels of allozyme diversity in Triatominae have been regarded as a possible indication of higher vulnerability to chemical control as insecticide resistance would be less likely to emerge in genetically depauperate insecticide-treated populations (Scholfield et al. 1995, Guhl and Scholfield 1996, Schofield and Dujardin 1997, Dujardin et al. 1998a, 2002, Monteiro et al. 2001).

In taxonomy (or alpha systematics, i.e. species identification), allozyme electrophoresis is used for the distinction of cryptic species and the determination of the correct status of dubious populations. Detection of reproductive isolation between populations of two (or more) cryptic species is particularly easy when populations are sympatric. This is because the rationale ‘if the populations belong to the same gene pool they must be exchanging genes’ can be used as the null hypothesis. Thus, if loci are scored that present different alleles fixed for both populations (i.e., they are diagnostic), the null hypothesis is rejected in favor of the alternative hypothesis of the existence of more than one taxon occurring in that area (for more details see Thorpe and Solé-Cava 1994).

This reasoning cannot be applied to allopatric populations, among which gene exchange might be absent simply due to geographical separation. These cases require the comparison of pairwise genetic distance values between the putative populations with those from the literature. This approach also works well because, provided enough individuals and loci are surveyed, genetic distances are usually much higher between populations of closely related species than between conspecific populations.

Cut-off values of genetic distance (*D*, Nei 1972) for species distinction have been proposed as
### TABLE II
Strengths and weaknesses of major molecular approaches applied to vector systematics.

<table>
<thead>
<tr>
<th>Method</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allozymes</td>
<td>• established marker</td>
<td>• samples must be fresh or frozen</td>
</tr>
<tr>
<td></td>
<td>• relatively inexpensive</td>
<td>• some gels can be difficult to genetically score</td>
</tr>
<tr>
<td></td>
<td>• co-dominant marker</td>
<td>• requires considerable practice</td>
</tr>
<tr>
<td></td>
<td>• provides information on several loci (and individuals) simultaneously</td>
<td></td>
</tr>
<tr>
<td>Cytogenetics</td>
<td>• relatively inexpensive and simple technique</td>
<td>• often too conserved for comparison of closely related taxa</td>
</tr>
<tr>
<td></td>
<td>• may be informative on recent range expansion processes</td>
<td>• live specimens required</td>
</tr>
<tr>
<td>PCR-RFLP</td>
<td>• established marker</td>
<td>• might reveal little polymorphism</td>
</tr>
<tr>
<td></td>
<td>• relatively inexpensive</td>
<td>• partial digestion can lead to problems</td>
</tr>
<tr>
<td></td>
<td>• co-dominant marker</td>
<td>• gives information on one locus at a time</td>
</tr>
<tr>
<td>RAPD</td>
<td>• inexpensive</td>
<td>• often presents reproducibility problems</td>
</tr>
<tr>
<td></td>
<td>• fast execution</td>
<td>• dominant marker</td>
</tr>
<tr>
<td></td>
<td>• requires small amount of molecular equipment and reagents</td>
<td>• interpretation is subjective</td>
</tr>
<tr>
<td></td>
<td>• provides information on several loci (and individuals) simultaneously</td>
<td>• little guarantee of homology among co-migrating bands</td>
</tr>
<tr>
<td>Microsatellites</td>
<td>• highly polymorphic (much more than allozymes)</td>
<td>• development of primers can be time- (and money-) consuming</td>
</tr>
<tr>
<td></td>
<td>• co-dominant markers</td>
<td>• requires access to sequencer</td>
</tr>
<tr>
<td></td>
<td>• provides information on several loci (and individuals) simultaneously</td>
<td></td>
</tr>
<tr>
<td>DNA sequencing</td>
<td>• highly informative</td>
<td>• labor-intensive</td>
</tr>
<tr>
<td></td>
<td>• interpretation is objective</td>
<td>• reagents are expensive</td>
</tr>
<tr>
<td></td>
<td>• allows for studies on any taxonomic level</td>
<td>• must have access to a sequencer</td>
</tr>
<tr>
<td></td>
<td>• allows for the comparison of results among different labs through DNA databanks</td>
<td>• gives information on a single locus</td>
</tr>
</tbody>
</table>

$D \approx 0.16$ (Thorpe and Solé-Cava 1994). In medically important insects, $D$ values $> 0.1$ are considered to indicate specific rank (Noireau et al. 1998). In triatomines, mean $D$ values of $0.504 \pm 0.341$ have been scored in 30 interspecific comparisons, while conspecific populations (142 comparisons) were separated by mean $D = 0.013 \pm 0.009$ (Dujardin et al. 2002).
TABLE III
Main areas of study, suggestions for future research, and requirements (and cautions) to be considered during research planning and experimental design.

<table>
<thead>
<tr>
<th>Field of study</th>
<th>Subjects</th>
<th>Examples</th>
<th>Requirements and cautions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population genetics</td>
<td>Dispersion-migration and gene flow; speciation; hybridization</td>
<td>Populations of main vectors in their areas of origin – mainly assessment of gene flow between sylvatic and domestic populations (e.g., <em>Rhodnius prolixus</em> in Venezuela)</td>
<td>Adequate sampling, with use of field-collected specimens; use of fast evolving markers; richness of associated ecological data</td>
</tr>
<tr>
<td>Taxonomy</td>
<td>Sibling species and species complexes; synonyms; rapid species/populations diagnosis</td>
<td>Sibling species within <em>R. robustus</em> or <em>Triatoma dimidiata</em>; chromatic variants in <em>Panstrongylus</em> and <em>Triatoma</em></td>
<td>Wide sampling to account for intraspecific variation; use more than one marker; obtain controversial forms, or taxa, in sympatry (whenever possible)</td>
</tr>
<tr>
<td>Phylogenetic relationships</td>
<td>Monophyly/non-monophyly of the subfamily and of other major groups (tribes, genera); phyleogeography – historical biogeography; relationships among members of problematic groups</td>
<td>Address relationships between members of: (1) Little-known tribes (Cavernicolini, Bolboderini, Alberproseniini); (2) <em>Triatoma</em> of North (vs.) South America; (3) <em>Panstrongylus-Triatoma</em> (and <em>Meccus</em>); (4) Old World species; (5) trans-Andean <em>Rhodnius</em> species; (6) little-known species groups (e.g., dispar complex, <em>Mepraia</em>, <em>Eratyrus</em>); (7) groups of species including the main vectors of Chagas disease; (8) geographical forms of widely distributed species (e.g., <em>P. geniculatus</em>, <em>P. rafotuberculatus</em>)</td>
<td>Selection of markers and of adequate analytical methods; taxon sampling, including the designation of suitable outgroups; area sampling in the case of widely distributed species</td>
</tr>
<tr>
<td>Ecology and adaptive trends</td>
<td>Ecological/evolutionary adaptations; phenotypic plasticity</td>
<td>Adaptive consequences of domestication in primary and secondary vectors; phenotypic-behavioral changes along ecological gradients</td>
<td>Combine molecular data with rich ecological datasets and quantitative phenotypic assessment</td>
</tr>
</tbody>
</table>
In general, when clear differences (diagnostic loci in sympatry or large genetic distances in allopatry) were detected between populations, the usual outcome was that more than one taxon was involved, either because of misidentification (e.g., R. colombiensis wrongly identified as R. prolixus) or cryptic speciation (as with T. sordida) (Moreno et al. 1999, Noireau et al. 1998, respectively). The characterization of phenotypically diverse populations of T. brasiliensis seemed to represent an exception in that fixed allozyme differences were found between conspecific populations (Costa et al. 1997). The alternative view that several distinct taxa were involved was later supported by DNA sequence analysis (Monteiro et al. 2004).

Several examples illustrate the use of allozymes in the assessment of triatomine taxonomy. The absence of fixed allozyme differences between R. prolixus (a major disease vector) and R. robustus (sylvatic and of comparatively minor medical importance) from Venezuela led to the suggestion that they were a single taxon (Dujardin et al. 1991, Harry et al. 1992a, b, Harry 1993, Solano et al. 1996). An allozyme comparison of R. prolixus and R. robustus individuals whose identity was previously confirmed by mitochondrial DNA again revealed no differences. This was interpreted as an indication of recent divergence rather than conspecificity (Monteiro et al. 2002). In similar cases, allozymes revealed negligible differentiation between closely related species within the phyllosoma and oliverai complexes (Flores et al. 2001, Noireau et al. 2002).

T. sordida, T. garciabesi, T. guasayana and T. patagonica are very similar but display distinct synanthropic trends; their overlapping ranges and high intraspecific anatomical variability make specific determination complicated (Gorla et al. 1993). Allozyme electrophoresis confirmed sordida, guasayana and patagonica as valid taxa, and led to the revalidation of T. garciabesi and the discovery of two cryptic species within T. sordida (García et al. 1995, Panzera et al. 1997, Jurberg et al. 1998, Noireau et al. 1998). The specific identity of T. petrochii was likewise confirmed using allozymes (Monteiro et al. 1998).

Finally, allozyme-based identification of field-collected specimens (including nymphs) has been instrumental for meticulous eco-epidemiological investigations. After defining allozyme markers for various species and populations, Noireau and colleagues thoroughly investigated the ecotopes, behavior, and epidemiological significance of several Bolivian triatomines (Noireau et al. 1998, 1999a, 2000a, b, Noireau and Dujardin 2001). A major discovery was that sylvatic foci of T. infestans are much more widespread than previously thought, extending from the Andean highlands well into the Chaco (Noireau et al. 2005).

The evolutionary interpretation of allozyme data has been done based on the relatively straightforward relationships between zymograms and genes. Phylogenetic relationships may be explored by comparing measures of genetic distance derived from allele frequencies (e.g., Nei 1972). Cladistic analyses rely on the ability to identify primitive (plesiomorphic) and derived (apomorphic) character states in the dataset by comparing the ingroup with a suitable sister taxon (or outgroup); only characters that are shared and derived (synapomorphic) are used for phylogenetic inference purposes (Avise 1994, Thorpe and Solé-Cava 1994). Based on these premises, the phylogenetic relationships of several triatomine groups have been explored using allozymes. These studies usually involved assessment of genetic distances among a few, closely related species (e.g. García et al. 1995, Pereira et al. 1996b, Solano et al. 1996, Panzera et al. 1997, Flores et al. 2001), but more complete investigations have been conducted with members of the tribe Rhodniini. Using phenetic techniques, three main clusters were identified within the genus Rhodnius: one basal clade (pallescens (ecuadoriensis-colombiensis)) and two sister groups (prolixus (nasutus-neglectus)) and (stali (pictipes-brethesi)) (Chávez et al. 1999). Most of these relationships were confirmed by a cladistic analysis, but a group (brethesi (pictipes-stali)) appeared in a basalmost position in the cladogram; the position of Psammolestes core-
odes within the tribe could not be satisfactorily re-
solved (Dujardin et al. 1999a). Recently the phy-
logeny of the Rhodniini was studied using 12 en-
zyme loci. Distance analysis produced a dendro-
gram showing paraphyly of Rhodnius, with Psam-
molestes tertius as a sister group to the prolixus
group within the cluster (domesticus (Ps. tertius
(nasutus-neglectus) (prolixus-robustus))). A second
main cluster was comprised of (pictipes (brethesi
(pallescens-ecuadoriensis))) (Monteiro et al. 2002).

DNA-BASED METHODS

The use of DNA approaches for the study of tri-
atomine bugs has a short, albeit rather dynamic, his-
tory. Molecular studies have provided new insights
into the systematics and evolutionary trends of the
subfamily, with surveys ranging from inter-tribe to
intra-specific levels. Both nuclear and mitochon-
drial gene sequence polymorphisms have been sur-
veyed. Many of the results have been used to refine
and improve disease control strategies throughout
Latin America, and will likely prove crucial for the
study of secondary vectors as well. Several DNA-
base techniques have been hitherto applied to tri-
atomines (Beard and Lyman 1999, Monteiro et al.
2001). These include randomly amplified polymor-
phic DNA (RAPD) (Carlier et al. 1996, Dujardin et
al. 1998a, García et al. 1998, Noireau et al. 2000a,
Borges et al. 2000, 2005, Jaramillo et al. 2001,
Pacheco et al. 2003), species-specific length vari-
ation of rDNA amplicons (Jaramillo et al. 2001),
single strand conformational polymorphism analy-
sis (Stothard et al. 1998), sequencing of selected
nuclear (Monteiro et al. 2000, Bargues et al. 2000,
2002, Marcilla et al. 2000, 2001, 2002) and mito-
chondrial gene fragments (García and Powell 1998,
Stothard et al. 1998, Lyman et al. 1999, Monteiro
2003, Gaunt and Miles 2002, Hypša et al. 2002,
Sainz et al. 2004), and characterization of poly-
morphic microsatellite markers (Harry et al. 1998,

DNA Sequencing

Sequencing of selected genomic fragments allows
for the direct assessment of DNA polymorphisms,
providing researchers with the ultimate information
for phylogenetic inference and evaluation of kinship
among organisms and populations. Different seg-
ments of the mitochondrial (mt) and nuclear genome
evolve at different rates. Fast-evolving regions are
suitable for the study of closely related organisms,
whereas conserved sections are more appropriate
for comparisons among more diverged taxa (Li and
Holmes 1998).

Mitochondrial genes

Mitochondria are cytoplasmic organelles of eukary-
otic cells; they are involved in key physiological and
pathologic processes (Avise 1994, Page and Holmes
1998, Saccone et al. 2000). In metazoans, each
mitochondrion has a single DNA molecule, 15-20
kilobases in length and containing genes coding for
2 rRNAs, 22 tRNAs, and 13 mRNAs responsible for
the synthesis of cell respiration proteins. This same
gene composition is found in the only mitochon-
drial genome of a triatomine sequenced to date (that
of Triatoma dimidiata). When compared to those
of other insects, the 17019 bp-long T. dimidata mt
genome has a lower adenine + thymine composition
bias (69.5% A+T) (Dotson and Beard 2001).

The single, circular mtDNA molecule, present
in multiple copies (~10^2-10^4 per cell), is mater-

an Acad Bras Cienc (2005) 77 (3)
a constant rate of $\sim 2.3\%$ pairwise sequence divergence per million years (my) in various arthropod taxa (Brower 1994). A global estimate for animal mtDNA is of $\sim 2\%$ divergence/my (Avise 1994).

Mitochondrial genes are widely used in population genetics and in evolutionary studies at different levels of divergence. Nucleotide sequence polymorphisms of rapidly evolving mtDNA genes are suitable for the assessment of relationships among recently diverged taxa (Avise 1994, Simon et al. 1994, Page and Holmes 1998). At higher levels of divergence, variable sites are prone to accumulate multiple substitutions (especially transitional hits at third codon positions), and thus become saturated. Saturation erases phylogenetic signal and is a major cause of homoplasy. Further cautions regarding the interpretation of mtDNA results in the study of closely related taxa include the possibility of introgression in hybrid zones (generally detectable by using complementary data from nuclear loci) and the retention of ancestral polymorphisms (or incomplete lineage sorting). On the other hand, direct sequencing is easier than it is with nuclear gene fragments, and sequence alignment is usually straightforward.

In 1999, FA Monteiro et al. reported the first survey of mtDNA sequence polymorphisms (cytochrome $b$ [cyt $b$] gene) among populations of a triatomine vector species, $T. infestans$ (Monteiro et al. 1999). Results revealed that $T. melanosoma$ and the sylvatic dark morphs of $T. infestans$ are phenotypic variants of $T. infestans$. Small but consistent differences were found between Bolivian populations and those from Argentina and Brazil. Within Bolivia, all Andean bugs had identical haplotypes, whereas dark morphs from the Chaco lowlands had slightly divergent sequences. These dark morphs differ from Andean $T. infestans$ in their phenotypes and in ecological features (they are melanic and occupy hollow trees in the Chaco). Allozymes, cytogenetics, interbreeding patterns, RAPDs, and mtDNA all supported the idea of a single species (Noireau et al. 1997, 2000a, Monteiro et al. 1999). Using two mtDNA fragments combined (12S+16S rDNA, totaling 878bp), higher levels of sequence diversity were found among geographic populations of $T. infestans$ from Argentina. Results showed a pattern of restricted gene flow, and suggested local population recovery from survivors after insecticide spraying. Heteroplasmy (individuals presenting more than one mt haplotype) was revealed in 15% of the specimens analyzed (García et al. 2003).

A recent mt cyt $b$ study involving all known chromatic variants of $T. brasiliensis$ showed that the $juazeiro$ and $melanica$ forms most likely deserve the status of species. $T. b. brasiliensis$ and $T. b. macro-melasoma$ (very different from $juazeiro$ and $melanica$) seem to represent distinct evolutionary lineages. However, because they differ only slightly from each other, a sound taxonomic decision will only be possible when further information is analyzed (Monteiro et al. 2004).

MtDNA sequence analysis revealed that $T. dimidiata$ populations from Yucatan are strongly divergent (perhaps a sibling species), and that the natural dispersal of the species probably followed a north-to-south route from southern Mexico through Central America, reaching Colombia. Ecuadorian populations however are probably recent derivatives from Honduran-Guatemalan conspecifics, supporting the hypothesis of an artificial introduction of the species to Ecuador and Peru (Marcilla et al. 2001, Abad-Franch 2003, Harris and Beard, personal communication).

Similarly, several populations of $R. prolixus$ (Venezuela, Colombia, Honduras, Guatemala, and Brazil) were found to have virtually identical mt cyt $b$ haplotypes, suggesting recent, artificial dispersal of synanthropic forms (Monteiro et al. 2000, 2003). This conclusion is also supported by metric and RAPD analyses (Dujardin et al. 1998a). On the other hand, high levels of structuring were found among various populations of $R. robustus$, suggesting that the taxon encompasses a complex of at least four sibling species (Monteiro et al. 2003).

In a study of mt cyt $b$ sequence data from five populations of $R. ecuadoriensis$ (four from Ecuador and one from Peru), Abad-Franch et al. (2004) found that Peruvian bugs present a markedly di-
divergent haplotype, while all Ecuadorian sequences (nine haplotypes) were similar. The results suggested that the two clades represent discrete phylogenetic groups (or even incipient species), and indicate that control programs can target them independently.

MtDNA sequence data have also been used to solve taxonomic questions. The specific status of several problematic taxa was confirmed (as with *R. prolixus*, robustus, nasutus, and neglectus), while previously recognized species were found to be just chromatic variants (e.g., *T. infestans* and *T. melanosoma*, see above). In several cases, the analyses revealed that a single putative species was in fact composed of more than one sibling taxon (e.g., *T. dimidiata*, *T. brasiiliensis*, *R. robustus*, or *R. ecuadoriensis*, see above). In a recent mt cyt *b* analysis, *R. pallescens* was found to be a very variable species, with four moderately divergent haplotypes detected in a small sample. The closely related *R. colombiensis* presented a haplotype that fell within the range of the intraspecific variation of *pallescens*, suggesting either conspecificity (*colombiensis* perhaps representing a race or subspecies) or very recent divergence (Abad-Franch et al. 2003). In a different case, sequencing of a single specimen (identified as *R. robustus* and belonging to a population frequently collected from palm trees in the Ecuadorian Amazon) revealed a cyt *b* haplotype not found in any other species of *Rhodnius*. Phylogenetic analyses show that this specimen surprisingly represents the most basal taxon of the *robustus* lineage (including *R. prolixus*, neglectus, the four sibling *robustus*, and *Psammolestes*) (Abad-Franch 2003). If further studies confirm these preliminary results, a new triatomine species would have been discovered after the results of DNA analysis.

The use of mtDNA for assessing systematic and phylogenetic questions among triatomines began in the mid 90s. An early study showed that the three main genera (*Rhodnius*, *Triatoma* and *Panstrongylus*) are well separated. Two main clades (*Rhodniini* and *Triatomini*) were evident, but genetic distances were greater between *R. pictipes* and *R. prolixus* than between *Panstrongylus* and *Triatoma* (Stothard et al. 1998).

The phylogeny of 11 species of the *infestans* complex plus four others has been explored using mt 12S and 16S rRNA and cytochrome oxidase I (COI) genes (García and Powell 1998, García et al. 2001). The close relatedness of *T. infestans* and *T. platensis* was confirmed, and evidence suggesting mtDNA introgression between these species was found. *T. circunnacula* and *T. rubrovaria* appeared within the *infestans* complex, while *P. megistus* clustered (with low bootstrap support values) with North American species. *T. sordida* was close to *T. matogrossensis* and very different from *T. guasayana* (García et al. 2001).

The mt 16S rRNA and cyt *b* genes have been used by Lyman et al. (1999). The Rhodniini were well separated from the Triatomini as indicated by a deep divergence between the two tribes. The Triatomini species could be further separated in a Central and North American species group and a South American species group. The placement of *P. megistus* and *Dipetalogaster maxima* was uncertain, but both were nested well within *Triatoma*, an indication of the paraphyletic nature of this genus.

Monteiro et al. (2000) assessed the relationships among several species of Rhodniini, including *Rhodnius* and *Psammolestes*. Results confirmed the paraphyly of *Rhodnius*, with *Psammolestes* appearing closer to the *prolixus* clade than this was to the *pictipes* clade. Parsimony analysis of 1429bp (including fragments of the mt 16S rRNA, mt cyt *b*, and the nuclear D2 variable region of the 28S rRNA gene) yielded two main clades: (*brethesi-pictipes* (*ecuadoriensis-pallescens*)) and (*Ps. tertius* (*neivai* (*domesticus* (*nasutus* (*neglectus-prolixus*) (*robustus-prolixus*)*)))).

Hypša et al. (2002) combined different DNA gene fragments (mostly mitochondrial 16S fragments) to construct a phylogeny including 57 species of Triatominae in nine genera, including Old World *Linshcosteus*. The deep split between Triatomini and Rhodniini was again confirmed. Most of the current generic subdivisions were not recognized as monophyletic clades. Thus, *Panstrongylus*, *Dipeta-
logaster, Mepraia, and Linshcosteus all appeared as sister taxa of different “Triatoma” species. The paraphyly of Rhodnius with respect to Psammolestes was also confirmed. Several taxonomic and systematic rearrangements were proposed, but some important questions (such as the apparent non-monophyly of Panstrongylus) were not considered (Hypša et al. 2002).

However, the most important finding of this work, albeit not so well supported (Stevens and Schofield 2003), was the indication that the Triatominae are monophyletic. The observation that the Old World genus Linshcosteus clusters within the Triatomini (with *T. rubrofasciata* as its sister taxon) clearly rules out the possibility of a separate and independent origin for Linshcosteus. An elucidative discussion on the issue of a mono- or polyphyletic origin of the Triatominae has recently been published by Schaefer (2003).

In a recent work, Gaunt and Miles (2002) used mt COI sequences and amino acid data to calibrate a mitochondrial molecular clock for various insect orders, including Hemiptera. Their results suggest that the evolutionary split of the ancestors of extant Triatomini (represented by South American species of *Triatoma*, two *Panstrongylus*, and *Eratyrus mucronatus*) and Rhodniini took place over 93 mya, roughly coinciding with the beginning of the separation of South American and African landmasses during the Cretaceous. The paraphyletic nature of *Triatoma* was apparent in that both *Eratyrus* and *Panstrongylus* nested again within representatives of that genus.

**Nuclear genes**

Nuclear genomic fragments are generally more conserved than mitochondrial genes. They are therefore better suited for analyses of diversity and relationships above the species level. However, ribosomal RNA spacers such as the internal transcribed spacers (ITs) can be informative for population analyses (Hillis et al. 1996). These markers have also been used in molecular taxonomic and evolutionary studies on Triatominae. The main concerns with the use of ITS regions in systematics refer to problematic sequence alignment and intragenomic variability.

Apart from the combined analysis of mt gene fragments with the D2 variable region of the 28S rRNA gene mentioned above, the second internal transcribed spacer (ITS-2) of the nuclear rDNA has recently been tested as a molecular marker for populations, species, and phylogenetic relationships in various Triatominae. These comparisons mainly involved Mesoamerican *Triatoma* species belonging to the phyllosoma complex, some other South American *Triatoma*, and several species of *Panstrongylus*; various populations of *T. infestans* and a few Rhodniini have also been analyzed (Marcilla et al. 2000, 2001, 2002).

Phylogenetic analyses of ITS-2 sequence data corroborated important findings based on mtDNA by revealing two major clades within the Triatomini, one comprised of Central and North American species (including *T. dimidiata*, three species of the phyllosoma complex, *T. barberi*, and *D. maxima*) and the second encompassing South American species (*T. infestans*, *T. sordida*, *T. brasiliensis*, and *P. megistus*). They also confirmed the paraphyly of Rhodnius by showing that *P. tertius* is more closely related to *R. prolixus* than the latter is to *R. stali* (Marcilla et al. 2001). This analysis was later extended to include several other species of *Panstrongylus*; the topology of the inferred trees varied slightly, with the main difference referring to the tendency of all *Panstrongylus* species to cluster together with the Mesoamerican Triatomini, rather than with the South American species (Marcilla et al. 2002). These studies provided evidence suggesting that some taxonomic rearrangements might be necessary within the phyllosoma complex and the genus *Panstrongylus*. For instance, *T. picturata* and *T. longipennis* had identical sequences, and the degree of variability within *T. dimidiata* was larger than that recorded for the rest of phyllosoma species. *P. lignarius* (considered sylvatic) and *P. herreri* (a major disease vector in northern Peru) presented identical ITS-2 sequences in spite of wide geographical range.

An Acad Bras Cienc (2005) 77 (3)
sampling. They have been recently synonymized by Galvão et al. (2003), a procedure that might also be appropriate for \textit{P. chinai} and \textit{P. howardi}, which are probably chromatic variants of the same species (F Abad-Franch and MD Bargues, unpublished data). The unexpected position of \textit{P. rufotuberculatus} as a sister taxon to the \textit{dimidiata-phyllosoma} clade, with \textit{T. barberi} – \textit{D. maxima} occupying the immediate external branch, was interpreted as a strong indication of the non-monophyly of \textit{Panstrongylus} (Marcilla et al. 2002).

The 100\% sequence identity between Ecuadorian and Honduran specimens of \textit{T. dimidiata} suggested that the former are recent derivatives of a Mesoamerican population probably introduced by people into western Ecuador in the recent past (Marcilla et al. 2001). Similar conclusions regarding human intervention in the passive dispersal of synanthropic bug populations were drawn from the analysis of ITS-2 sequence variation among \textit{T. infestans} geographic populations (Marcilla et al. 2000). The 18S subunit of the nuclear rDNA is much more conserved than ITS-2 or mt protein-coding genes; in triatomines, the substitution rate seems to approach 1.8\% sequence divergence per 100 million years, up to 55 times slower than ITS-2 (Bargues et al. 2000). These substitution rate estimates were used to calculate the time to common ancestry among various triatomine taxa; under the mentioned hypothesis of an 18S rDNA molecular clock, divergence between the ancestors of the Triatomini and Rhodniini would give an estimate of \(~48.9-64.4\) million years ago (mya) (Palaeocene-Eocene) (Bargues et al. 2000). Estimates of divergence between Meso- and South American species of Triatomini fell between 22.8 and 31.9 mya (18S) and between 19.5 to 34.1 mya (ITS-2), long before the Panama Isthmus linked North and South America in the middle Pliocene (\(~3\) mya, although biogeographical evidence shows that some biotic interchange could have occurred from the late Miocene-early Pliocene, i.e. \(~6-7\) mya); these results were interpreted as strong evidence suggesting independent evolutionary origins of both groups, whose common ancestors would have diverged during the Oligocene-Miocene (Bargues et al. 2000; see also Cox and Moore 2000).

**RANDOMLY AMPLIFIED POLYMORPHIC DNA (RAPD)**

Amplification of polymorphic DNA with random primers has been used to investigate genetic variability in several medically important insects. In triatomines, diversity among \textit{T. infestans} and \textit{T. sor-dida} populations was assessed by Carlier et al. (1996). The banding patterns clearly separated both species and discriminated sylvatic and domestic populations of \textit{T. infestans} (Carlier et al. 1996). In a further study on Bolivian \textit{T. infestans}, RAPD profiles separated two main clusters (Andes and Chaco) and, within each clade, sylvatic and domestic populations (Noireau et al. 2000a). García et al. (1998) found diagnostic RAPD profiles for the pairs \textit{R. prolixus-robustus} and \textit{R. neglectus-nasutus}. Honduran and Colombian populations of \textit{R. prolixus} had population-diagnostic RAPD patterns. Reduced variability among Central American bugs suggested recent human-mediated dispersal of a South American population, prompting the idea that local eradication of the species could be contemplated in Central America (Dujardin et al. 1998a).

Although the resolution of RAPD analysis is much higher than that of allozyme electrophoresis in detecting intraspecific variability, the use of random primers requires extreme caution in order to minimize contamination and problems with reproducibility (Black 1993). Another serious problem with RAPDs is that, because primers anneal randomly, there is little guarantee of homology between co-migrating bands. Even specific conditions used for the preparation of samples and PCR may also affect the results, reducing reliability and reproducibility. Finally, because RAPD markers are dominant, there is no way to determine whether a single band on the gel corresponds to a homozygous or heterozygous individual. This precludes the use of such markers for population genetic studies because conformance to H-W expectations cannot be tested (Black 1993, Loxdale and Lushai 1998).
Microsatellites

Microsatellites are series of short repetitive motifs [e.g., (GT)$_n$ or (AT)$_n$] within nuclear DNA; they are highly polymorphic, neutral, and exhibit Mendelian inheritance and codominance (Jarne and Lagoda 1996). Microsatellites have become the tools of choice for population genetics when other markers do not show an adequate degree of polymorphism. Harry et al. (1998) studied sylvatic populations of *R. pallescens* from *Attalea* palm trees; six out of the 10 microsatellites evaluated were present in frequencies not different from those expected under H-W equilibrium, suggesting panmixia among palm populations in the area. Amplicons were also obtained using template DNA from *R. ecuadoriensis* and *R. prolixus* (Harry et al. 1998). More recently, Anderson et al. (2002) identified and characterized eight microsatellite loci from populations of *T. dimidiata* from Mexico, Guatemala and Honduras, and García et al. (2004) have characterized 10 loci for *T. infestans*.

Cytogenetics

Triatomines have holocentric chromosomes, usually with a diploid complement of 2n=20A+XX/XY (20 autosomes plus XX ♀ and XY ♂). Caryotypes with 18A and 22A are rare. Males of North and Central American species of *Triatoma* usually present an X$_1$X$_2$Y sex mechanism, while South American species typically present XY. Both *Panstrongylus* and *Eratyrus* also present X$_1$X$_2$Y. All Rhodniini have a 2n=20A+XX/XY diploid complement. Special cytogenetic techniques, such as C-banding and detailed analysis of heterochromatic regions, or the study of the meiotic behavior of male chromosomes, have been used to complement morphological characterization of Triatominae. Applications of these methods range from identification of similar species and detection of intraspecific variability to the investigation of evolutionary relationships at different levels of divergence (reviewed by Dujardin et al. 2002).

CONCLUSIONS

Different molecular approaches, combining sound sampling strategies with an intelligent choice of markers, can be used to address diverse questions on the biogeography, behavior, taxonomy, evolution, and population structure of triatomines. We have presented an overview of what has already been done in these fields. We also aimed at stimulating further work by proposing new avenues for molecular research, underlining, at the same time, what requirements and cautions apply to the design of the studies. It is by producing high-quality results that molecular research will keep providing the basis for more effective schemes of disease vector control and surveillance.

ACKNOWLEDGMENTS

Supported by the UNDP/World Bank/WHO TDR Special Program and the ECLAT Network.

RESUMO

Programas destinados ao controle da doença de Chagas vêm apresentando resultados positivos. Estudos moleculares têm auxiliado programas de controle exitosos através da identificação e caracterização de populações de vetores introduzidas, além da definição precisa das espécies a serem combatidas. Contudo, pesquisadores e autoridades da área de saúde estão enfrentando novos desafios, no âmbito da América Latina. Vetores nativos estão continuamente re-infestando habitações previamente tratadas com inseticidas, e triatomíneos silvestres estão mantendo ciclos de transmissão da doença em regiões de floresta tropical úmida (incluindo a Amazônia), sem colonizar habitações humanas. Nessas situações, estudos detalhados dos vetores são essenciais na definição de padrões de risco epidemiológico e no esclarecimento do envolvimento de espécies de triatomíneos pouco conhecidas, na transmissão da doença. Investigações eco-epidemiológicas dessa natureza, assim como o planejamento e monitoramento de intervenções de controle, dependem fortemente de identificações taxonômicas precisas. Problemas decorrentes de especiação críptica e da plasticidade fenotípica, ilustram essa necessidade – e de como a sistemática molecular pode contribuir na geração das respostas necessárias. A análise de dados moleculares...
também auxilia no entendimento de aspectos básicos da evolução e tendências adaptativas dos vetores. Neste artigo, fazemos uma revisão da aplicação de marcadores moleculares (concentrando em isoenzimas e sequenciamento de ADN) no estudo de triatomíneos. Analisamos também a aplicabilidade, vantagens e desvantagens dos métodos mais utilizados, nas investigações em diferentes níveis sistemáticos (populações, espécies e categorias taxonômicas mais elevadas).

**Palavras-chave:** sistemática molecular, ADN, isoenzimas, Triatominae, doença de Chagas.

**REFERENCES**


Beard CB and Lyman DF. 1999. DNA-based approaches for inter- and intraspecific analysis of Chagas disease vectors, p. 74-77 in (Schofield CJ and Ponce C (Eds)), Proceedings of the second International Workshop on population genetics and control of Triatominae, Tegucigalpa, Honduras, INDER, Mexico City, Mexico.


Frías D and Dujardin JP. 1996. Detección de la variación genética a nivel poblacional a través de electroforesis de isoenzimas, p. 65-71 in (Schofield CJ, Dujardin JP and Jurberg J (Eds)), Proceedings of the International Workshop on population genetics and control of Triatominae, Santo Domingo de los Colorados, Ecuador, INDRE, Mexico City, Mexico.


An Acad Bras Cienc (2005) 77 (3)


MOLECULAR RESEARCH ON TRIATOMINES


