Genetic subdivisions within *Trypanosoma cruzi* (Discrete Typing Units) and their relevance for molecular epidemiology and experimental evolution

Michel Tibayrenc*

Address: UR62 "Genetics of Infectious Diseases", UMR CNRS/IRD 9926, IRD Centre, BP 64501, 34394 Montpellier Cedex 5, France

Email: Michel.Tibayrenc* - Michel.Tibayrenc@mpl.ird.fr

* Corresponding author

Abstract

**Background:** This paper summarizes the main results obtained on *Trypanosoma cruzi* genetic diversity and population structure since this parasite became the theme of many genetic and molecular studies in the early seventies.

**Results:** *T. cruzi* exhibits a paradigmatic pattern of long-term, clonal evolution, which has structured its natural populations into several discrete genetic subdivisions or "Discrete Typing Units" (DTU). Rare hybridization events are nevertheless detectable in natural populations and have been recently obtained in the laboratory.

**Conclusions:** The DTUs and natural clones of *T. cruzi* constitute relevant units for molecular epidemiology and experimental evolution. Experimental mating opens the way to an in-depth knowledge of this parasite’s formal genetics.

Introduction

It is probable that *Trypanosoma cruzi*, the agent of Chagas disease, is the pathogenic microorganism for which intraspecific genetic diversity is the best known. Long-standing interest in this diversity has led many teams currently working on this parasite to follow and even to generate the recent technological progress in biochemical typing, molecular epidemiology* (Terms quoted * are explained in the glossary) and population genetics*. In the early seventies, obtaining knowledge of the population structure of pathogenic microorganisms was a major challenge, since their formal genetics was entirely speculative. The isoenzyme* era gave us the first insights to the genetic diversity of *T. cruzi* and other parasites. Pioneering studies by Miles et al. [1,2] showed clearly the existence of three main isoenzyme types, which were called "principal zymodemes**". Such zymodemes were taken as units of analysis for epidemiological surveys [3] and hypotheses on *T. cruzi* pathogenicity [4]. Numerical taxonomy showed their overall clustering [5]. However the biological nature and evolutionary origin of the zymodemes* remained entirely unknown. The specific contribution of our group has been to apply the concepts of population genetics to the study of *T. cruzi* biochemical and genetic polymorphism. The present paper describes the main results reached in this field by our group and other teams.

Methodology

The key-point in understanding the population structure of a pathogenic microorganism is its mating system. The classical view, that microbes reproduce clonally, has been upset by the population genetic era. Genetic exchange is very frequent in many microbial species. Whatever its precise cytological mechanism, horizontal gene transfer
Weinberg expectation is sometimes difficult. As an example, it requires that the ploidy of the organism is known. This approach when microbial pathogens are considered. First, Weinberg equilibrium* statistics. It is a demanding methodology. Allelic segregation is surveyed by the classical Hardy-Weinberg equilibrium test (Wahlund effect). Means to avoid such biases have been elaborated by others for bacteria [8] has been to look for the expected consequences of random allelic segregation and unilocus genotypes ephemeral [5].

Table 1: Criteria and tests of clonality (after 10)

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segregation</td>
<td>(within locus)</td>
</tr>
<tr>
<td>a</td>
<td>Fixed heterozygosity *</td>
</tr>
<tr>
<td>b</td>
<td>Absence of segregation genotypes *</td>
</tr>
<tr>
<td>c</td>
<td>Deviation from Hardy-Weinberg expectation</td>
</tr>
<tr>
<td>Recombination</td>
<td>(between loci)</td>
</tr>
<tr>
<td>d</td>
<td>Overrepresented, identical genotypes widespread *</td>
</tr>
<tr>
<td></td>
<td>(statistical tests d1 ** and d2 ** §)</td>
</tr>
<tr>
<td>e</td>
<td>Absence of recombinant genotypes ** §/P&gt;</td>
</tr>
<tr>
<td>f</td>
<td>Linkage disequilibrium ** §/P&gt;</td>
</tr>
<tr>
<td>g</td>
<td>Correlation between two independent sets of genetic markers ** §P</td>
</tr>
</tbody>
</table>

* = used qualitatively, without statistical calculations; ** = used with statistical calculations. # = Combinatorial analysis § = Monte Carlo simulation with 10^4 iterations ¶ = Mantel nonparametric test of correlation between genetic distances obtained from different markers

The "clones*" observed by a given set of genetic markers will prove to be genetically heterogeneous if a more discriminative marker is used. We have forged the term "clonet" to designate a set of stocks that appear genetically identical with a given set of markers in a clonal* species [11].

In complement with population genetics, classical phylogenetic analysis is useful to look for possible discrete genetic subdivisions in the species under study. In the case of microbial pathogens, many times, such subdivisions are observed, however they do not fulfill the rigorous criteria of phylodinetic analysis, since some level of horizontal gene transfer renders these subdivisions incompletely isolated from each other. The descriptive concept of "Discrete Typing Unit" (DTU) designates a set of stocks that are genetically more similar to each other than to any other stock, and are identifiable by common genetic, molecular, or immunological markers named "tags" [14].

Main results in Trypanosoma cruzi

T. cruzi is aneuploid [9]. However such results do not falsify the line of evidence based on the analysis of linkage disequilibrium*, which remains valid. As a matter of fact, an impressive congruence of results corroborates the existence of strong linkage disequilibrium* in the agent of Chagas disease [15], even in sylvatic cycles and when each genetic subdivision is analyzed separately. A striking illustration of this linkage disequilibrium* is the existence of a highly significant correlation between independent genetic markers, including isoenzymes [16], Random Primed Amplified Polymorphic DNA* or RAPD [17] and microsatellites [18]. It seems that long-term clonal evolution in T. cruzi has been
predominant enough to lead to the individualization of several discrete genetic subdivisions or DTUs [14]. The number of observable DTUs within *T. cruzi* is a matter of debate. Most studies show the existence of two main subdivisions [13,19], referred to as *T. cruzi* I and II [20]. Multilocus markers reliably show a total of six DTUs, one corresponding to *T. cruzi* I, the others corresponding to subdivisions within *T. cruzi* II [16,17,21]. Classifications based on gene sequencing either support the division into 6 DTUs [22] or indicate a lesser number [23–25]. This illustrates the usefulness of the DTU concept. Two *T. cruzi* DTUs (2d and 2e; see figure 1) correspond actually to hybrid lineages stabilized by subsequent clonal propagation [9,25–27]. These lineages do not fulfill the strict criteria of cladistic analysis and actually, they are not clades, since they have two ancestors instead of one, which explains the inconsistency of gene sequence phylogenies. However, they correspond to reliable genetic subdivisions and are identifiable by an impressive set of tags.

Figure 1
Neighbour joining dendrograms based on the analysis of 20 RAPD* primers (left) and 22 isoenzymatic* loci (right) showing the genetic relationships between 49 *Trypanosoma cruzi* stocks and *T. cruzi marinkellei* stock M1117. The scale indicates the Jaccard distance [39] along the branches. Six gene tic clusters, or Discrete Typing Units (DTUs [14]), were distinguished and their names are given in the central column between the dendrograms. DTU 1 corresponds to the 1rst major lineage of *T. cruzi*, while the second major lineage is subdivided into DTUs 2a to e. Diagnostic RAPD fragments and isoenzymatic patterns, which were specifically observed in the stocks of a given cluster of the dendrograms (tags [14]), are indicated at the corresponding nodes (after [17]).
Molecular epidemiology

Currently, the 6 DTUs are the most reliable subdivisions of *T. cruzi*. They appear as robust units of analysis for molecular epidemiology studies. DTU 1 ("T. cruzi I" [20]) corresponds to all genotypes related to the formerly described "principal zymodeme I" [1,2]. It is a very broad and heterogeneous group. Its epidemiological and geographical specificity is low. It is found on the entire range of Chagas disease, from southern USA to Argentina. It can be found in domestic cycles in Andean countries as well as in Amazonian sylvatic cycles. It is very frequent in chronic cases of Chagas disease. Identifying therefore an isolate as DTU 1 has a very low predictive value on its expected properties. The case is different for the 5 DTUs that subdivide *T. cruzi* II [20]. Their epidemiological and geographical specificity is clearer, and identifying them is therefore informative. DTU2a and DTU2b correspond to stocks of "principal zymodemes III and II", respectively [1,2]. Interestingly, DTU 2b (zymodeme III), which was until now only known from Sylacnic cycles, has been recently recorded in chronic cases of Chagas disease in Ecuador [28]. This shows that our knowledge of the epidemiological implications of *T. cruzi* genetic variability still is incomplete. The epidemiological relevance of the 6 DTUs has been analyzed in details by Barnabé et al. [16]. It is relevant to identify also the clonets within each DTU for finer epidemiological studies. It is desirable for this purpose to perform a discriminative genetic characterization. Our group routinely uses 22 isoenzyme loci and 20 RAPD primers [16,17].

Experimental evolution

The successful recombination experiments recently obtained by M. Miles' group [9] constitute a major step toward elucidating *T. cruzi* formal genetics and the evolutionary mechanisms that generated the observed genetic subdivisions of this parasite. The 6 *T. cruzi* DTUs provide us with a fine model for experimental evolution, making it possible to evaluate the impact of predominant clonal evolution on this parasite's relevant biomedical properties. Our laboratory has designed a standardized set of about 30 stocks representative of the 6 DTUs. Each stock has been laboratory-cloned with verification under the microscope. Many experimental parameters have been surveyed on this standardized sample by our group [29–35], including growth in in vitro culture, infectivity to cell cultures, pathogenicity in mice, transmissibility through triatominine bugs, and in vitro and in vivo drug sensitivity. All these parameters have been quantified and we have looked for a correlation between the biological differences on one hand, and genetic distances* among DTUs on the other hand. For all parameters tested, the correlation has been highly significant, suggesting that biological differences parallel phylogenetic divergence between *T. cruzi* DTUs. However, within each DTU, results are quite heterogeneous. For example, stocks pertaining to DTU 1 tend to be more pathogenic for mice than stocks from DTU 2b [32]. However, the values overlap, and the more pathogenic DTU 2b stocks are more pathogenic than the less pathogenic DTU 1 stocks. An interesting pattern has been observed in several cases: mixtures of clones present a different behavior from what would be a simple resultant of the behavior of each pure clone they are composed of. For example, a mixture of a very pathogenic and of a poorly pathogenic clone is more pathogenic than the more pathogenic pure clone [32]. The same has been observed for transmissibility through the insect vector [30]. This suggests a "clonal cooperation" or "clonal hitchhiking,"[36] that acts probably through biochemical messengers. It is possible that such mixtures of genotypes play an important role in the pathogenicity of Chagas disease. Other authors have noted that *T. cruzi* clonal genotypes seem to exhibit different organ tropisms [37].

Work is in progress in our group to identify the genetic mechanisms of biological differences between DTUs and of clonal cooperation through the analysis of gene expression and proteomic data.

Glossary of specialized terms

Clone, clonal, clonality: "Clonal" propagation does not amount to "mitotic" propagation: in population genetics, this term is used in all cases where the individuals of the progeny are genetically identical to one another and to the reproducing individual. Apart mitotic reproduction, this includes several cases of parthenogenesis as well as self-fertilization in haploid organisms. A clonal population structure can be therefore observed in animals exhibiting apparent meiosis, and even, mating.

Genetic distance: Various statistical parameters inferred from genetic data, estimating the genetic dissimilarities among individuals or populations. The most widely used are Nei's standard genetic distance [38] and Jaccard distance [39]. Although the statistics differ, many genetic distance indices start from an estimation of the percentage of band mismatch on electrophoresis gels.

Hardy-Weinberg: see segregation*

Isoenzyme: a set of electrophoretic variants of a given enzyme. Isoenzymes differ from each other only by their electrophoretic mobility. This last property is a reflection of the overall electric load of the protein, which itself is a resultant of the individual electric load of its aminoacids. The electrophoretic mobility of a given protein is therefore a reflection of its primary structure, and indirectly, of the sequence of the gene that encodes for it.

Linkage disequilibrium: nonrandom reassortment of genotypes occurring at different loci (see recombination*)
Molecular epidemiology: the various biochemical and molecular techniques used to type and subtype pathogens

Panmixia, panmictic: a situation in which gene exchanges occur randomly in the population under survey.

Population genetics: A set of statistics based on the analysis of genetic data aiming to give a snapshot of the population structure of a given organism, and the impact, on this population structure, of migration, genetic recombination and natural selection.

Random Amplification of Polymorphic DNA (RAPD): A method simultaneously proposed by Williams et al. [41] and Welsh & McClelland [42] to analyse genetic variability (other name: Arbitrarily-Primed Polymerase Chain Reaction = AP-PCR). While in the classical PCR method, the primers used are identified DNA sequences, the RAPD technique relies on primers which sequence is arbitrarily determined (usually 10-mer primers are used). Under low-stringency conditions, the PCR reaction generates fragments which polymorphism can be analyzed on either ethidium bromide-stained agarose gels [41], or polyacrylamide sequence gels with radiolabeling of the fragments [42].

Recombination, linkage disequilibrium: Free recombination makes that the expected probability of a given multilocus genotype is the product of the observed probabilities of the single genotypes it is composed of. For example, in a panmictic human population, if the observed frequency of the AB blood group is 0.5, and the observed frequency of the Rh (+) blood group is 0.5, the expected frequency of the individuals who are AB and Rh (+) is 0.5 x 0.5 = 0.25. Inhibition of recombination leads to linkage disequilibrium*, or nonrandom association among loci (the predictions of expected probabilities for multilocus genotypes are not found). For example, if the observed frequency of the individuals AB and Rh (+) was statistically higher than 0.25, this would show that the two loci are linked together (they are not transmitted independently). For example, if this frequency was 0.5, this would be the sign of a total linkage between AB and Rh (the two characters are transmitted as only one).

Segregation, Hardy-Weinberg equilibrium: in a panmictic* population of a diploid organism, let us consider a gene at which there are two possible alleles, a and b. The frequency of a is p, and the frequency of b is q = 1 - p. The Hardy-Weinberg law predicts that the frequency of each of the three possible genotypes, that is to say aa, ab and bb, will be p^2, 2pq and q^2, respectively. If the observed frequencies are statistically different from the expected ones, this is evidence that gene flow is restricted in the population under survey.

Zymodeme: a set of stocks that share the same isoenzyme* profile.

References

19. Souto RP, Fernandes O, Macedo AM, Campbell DA and Zingales B: DNA markers define two major phylogenetic lineages of


