Cell PRESS

How clonal are *Trypanosoma* and *Leishmania*?

Michel Tibayrenc[1*] and Francisco J. Ayala[2]

[1] Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle (MIVEGEC), [Institut de Recherche pour le Développement (IRD) 224–Centre National de la Recherche Scientifique 5290–Unité Mixte 1/2], IRD Center, Boite Postale 64501, 34394 Montpellier Cedex 5, France
[2] Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA

The clonal theory of parasitic protozoa has been recently challenged by researchers stating that recombination in Kinetoplastida is much more frequent than previously believed, or that selfing and homogamy should be distinguished from 'strict' clonality. These researchers and many others show that the concept of clonality proposed by us is not correctly understood. A recapitulation of the clonal theory will thus be addressed herein. Comparisons with various other pathogens evidence general features among them and enhance our understanding of *Trypanosoma* and *Leishmania* population genetics. The relevance is considerable not only for our knowledge of the basic biology of these organisms but also for applied research: molecular epidemiology (strain-typing), clinical research, vaccine and drug design, and experimental evolution.

An indispensable recall

The clonal theory states that, in natural populations of many pathogen species, recombination is not frequent enough to break the pattern of preponderant clonal evolution (PCE). The only and unambiguous definition of PCE is therefore: drastically restrained genetic recombination. We have shown [1] that this definition is shared by most authors working on pathogens. The theory clearly states two important points: first, it does not refer to any precise cytological mechanism, but instead to the genetical consequences of clonality (see Glossary). Hence, it includes selfing or strong homogamy (which lead to restricted recombination) within the PCE concept, contrary to some statements [2]. Second, it does not claim that recombination is totally absent or plays a minor role [3]. Conversely, it affirms that recombination most probably obtains in all pathogens, and could play a major role, but only on an evolutionary scale, not in each generation. Another important point [4] is that restrained recombination is due to inbuilt genetic properties of the organisms and is not caused by preponderant natural selection or incidental physical obstacles (the Wahlund effect). Lastly, the model considers each species as a whole. It is therefore not 'challenged' by analyses of lesser subdivisions considered separately, instead of considering the whole species [5–8].

Generalization and development of the clonal model

Based on an extensive examination of data concerning bacteria (48 species), fungi and yeasts (nine species), parasitic protozoa (21 species), and viruses (11 species or categories), we have proposed [1] to replace subjective assertions such as 'widespread' genetic exchange [8] or 'gross' incongruences (between phylogenetic trees) [5] with a clear definition of PCE relying on a firmly settled cursor. According to this definition, the main features of PCE (which are linked to one another) include: (i) strong linkage disequilibrium (LD), or nonrandom association of genotypes occuring at different loci, leading to widespread, identical or nearly so, multilocus genotypes (MLGs); and (ii) clear phylogenetic signal leading to widespread genetic subdivisions ('near-clades') that are stable in space and time, even on an evolutionary scale. We also have discussed the relevance of distinguishing selfing/homogamy from 'strict' clonality [1].

LD is the logical consequence of restrained recombination and purifying selection

LD has been criticized [2,9,10] for its supposed lack of statistical power and the fact that LD population genetics analyses are biased by a lack of resolution of the markers considered, which, in fact, should bias the result towards the null hypothesis of panmixia [1]. LD lack of resolution is exaggerated. As with all statistics, its resolution depends upon the richness of the information available. If many loci are considered, it is very powerful [11]. Moreover, by comparison with segregation tests, LD has two considerable advantages: (i) it can be performed irrespective of the ploidy of the organism, and even without knowing it [1,11], a crucial point when *Trypanosoma* and *Leishmania* are considered (see below). (ii) LD is precisely the statistics able to detect restrained recombination, the central parameter on which our definition of PCE is based, and the most important point to consider for molecular epidemiology (follow-up of stable MLGs) [1].

Strong phylogenetic signal leading to widespread clustering ('near-clading')

This consequence of PCE is less intuitive than LD. However, it is a crucial PCE parameter whose relevance for molecular epidemiology and taxonomy is considerable.

Corresponding author: Tibayrenc, M. (Michel.Tibayrenc@ird.fr)
[*] *Present address:* Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA.
Keywords: Kinetoplastida; clonality; population genetics; molecular epidemiology; linkage disequilibrium; preponderant clonal evolution; selfing; homogamy.

1471-4922/$ – see front matter
 http://dx.doi.org/10.1016/j.pt.2013.03.007

Glossary

Amplified fragment length polymorphism (AFLP): selective amplification of genomic restriction fragments (obtained by restriction fragment length polymorphism) by PCR using randomly selected primers (see random amplified polymorphic DNA).
Clonality: absent or severely restrained genetic recombination.
Copy-number variation (CNV): variable number of copies of one or more segments of the DNA in a given organism. CNVs correspond to relatively large regions of the genome that have been deleted or duplicated. The impact of CNV on phenotypic variability is thought to be important.
Diploidy: two copies of each chromosome in a given organism.
G test: correlation between genetic distances calculated from different genetic markers [11]. When positive, it manifests a striking case of LD.
Heterozygote deficit/excess: in a panmictic population of a diploid species, allele frequencies are given by the Hardy–Weinberg formula (Hardy–Weinberg equilibrium). Heterozygote excess or deficit shows that allele segregation does not occur at random.
Homogamy: tendency to mate with partners that are genetically identical, or nearly so. It may happen even if different genotypes of the pathogen are present in the host.
Homoplasy: possession in common by distinct phylogenetic lineages of similar or identical characters that do not originate from common ancestry. The origin of homoplasic characters include the following: (i) convergence (possession of similar characters derived from different ancestral traits owing to convergent evolutionary pressure; for example, wings of birds and wings of bats); (ii) parallelism (possession of identical traits derived from a same ancestral character, but generated independently in different phylogenetic lineages); and (ii) reversion (restoration of an ancestral character from a derived character).
Linkage disequilibrium (LD): nonrandom association of genotypes occurring at different loci. LD is observed in sexual populations between genes that are close to each other on the same chromosome, or in predominantly clonal populations due to restrained recombination, or in the case of a Wahlund effect (see this term).
Microsatellite: a short DNA sequence, usually 1 to 4 bp in length, which is consecutively repeated along a DNA molecule. There is variation from individual to individual and among different populations in the number of repeats. Numbers of repeats define microsatellite alleles. There are hundreds of DNA regions that contain microsatellites. Microsatellites are fast-evolving markers, with a high resolution level, but with a high rate of homoplasy.
Montpellier 1 (MON 1): in the multilocus enzyme electrophoresis (MLEE) typing system standardized by a reference research group of Montpellier (France), the name of a *Leishmania infantum* MLEE genotype that is frequently recorded in the Old and New Worlds.
Mosaic aneuploidy: the number of copies of a given chromosome varies between chromosomes, or between different species, or between different strains of a given species [24].
Muller's ratchet: the process by which the genomes of an asexual species accumulate more and more deleterious mutations.
Multilocus enzyme electrophoresis (MLEE): isoenzyme analysis based on a broad range of enzyme systems. Each enzyme system corresponds to one or several genetic loci.
Multilocus genotype (MLG): the composite genotype made up of all individual genotypes recorded at each locus under survey in a given individual.
Multilocus sequence typing (MLST): a standardized typing procedure that relies on the sequencing of fractions 450 bp in length, and is typically carried out for each of seven housekeeping genes. Only two types of alleles are considered: identical and nonidentical.
Near-clade: occasional bouts of genetic exchange explain that the genetic clusters observed in a predominantly clonal population cannot be equated to real clades (between which genetic exchange is completely absent). We have coined the term 'near-clades' to designate such genetic clusters, which are observed in many pathogen species [1].
Panmixia: a population is panmictic when genetic exchanges occur at random within it.
Phylogenetic signal: the observable consequences with appropriate phylogenetic tools of a situation where a given collection of organisms is composed of distinct entities among which genetic exchange is either absent or severely restricted.
Purifying selection = purging selection = directional selection: one, or a few, multilocus genotypes are positively selected to the detriment of other MLGs which tend to be under-represented or eliminated.
Random amplified polymorphic DNA (RAPD): in the classical PCR method, the primers used are known DNA sequences, whereas the RAPD technique relies on primers whose sequences are arbitrarily determined. RAPD primers are generally 10 nt in length (see AFLP).
Restrained recombination: genetic recombination among loci does not occur at random. In predominant clonal evolution (PCE), restrained recombination is strong.
Selfing = self-fertilization: fusion of sex cells produced by the same individual. In the case of parasitic protozoa, if only one genotype is available in the host, it will mate with itself.
Single-nucleotide polymorphism (SNP): one-letter polymorphisms in the DNA sequence. SNPs contribute to differences among individuals and populations. They are widely used as high-resolution population markers.
Tree incongruence: when phylogenetic trees derived from different loci or different genetic markers are very different for the same set of organisms.
Type II error: in a statistical test, when it is impossible to reject the null hypothesis, not because it is verified, but because the test is not powerful enough.
Wahlund effect: if two or more subpopulations have different allele frequencies, then the overall heterozygosity is lowered, even if each subpopulation is in Hardy–Weinberg equilibrium. More broadly: all population genetic consequences of plotting together different subpopulations that have different allele and genotype frequencies.
Widespread clustering: when a given species is subdivided through its entire range into discrete genetic clusters that are stable in space and time.

Many pathogen species [1] show genetic clusters that are stable in space and time, and can be observed in many different places at scales of years and decades. We have called them 'near-clades', rather than simply clades, because when pathogens are concerned some genetic recombination always interferes. 'Clade' is therefore improper, although frequently (and misleadingly) used in the case of pathogens (Box 1). Consequently [1], the search for near-clades should not be based on strict cladistic demands, but rather on a flexible phylogenetic approach relying on the congruence principle [12], which states that evidence increases as more data are considered. For example, near-clading becomes more and more apparent when more loci are analyzed [13], when various clustering or phylogenetic approaches, based on different working hypotheses, give convergent results [14], when genetic distances based on different genetic markers are strongly correlated ('G test') [1,11], or when near-clades are confirmed by different markers [13]. If, on the contrary, indication for near-clading does not increase or vanishes when more data are considered, this would be evidence that the phylogenetic signal is weak and that recombination plays a preponderant role. This proposal is true, however, only if suitable data are considered. For example, if markers with inappropriate resolution are used, or if sampling is scarce, the congruence principle may not apply, even if the phylogenetic signal is present.

Debate within the debate: selfing and homogamy versus 'strict' (mitotic) clonality

Whether PCE should include selfing and homogamy or not is a matter of definition. Evidencing selfing is evolutionarily relevant. The main evolutionary interest of selfing is DNA repair, which is not allowed by mitotic clonality. However, we have proposed [1,4,11] definitely to include selfing and homogamy in the PCE definition. The main reasons are that: (i) most authors working on pathogens do not distinguish selfing or homogamy from clonality [1]; (ii) according to our definition of PCE, the main evolutionary consequence of selfing and homogamy is very similar to 'strict' clonality (strongly restricted recombination); (iii) the authors infer selfing and homogamy from the presence of a heterozygote deficit, and separately evaluate the probability of other possible causes [2], although they are not exclusive of each other – these other causes include homoplasy (of high impact when microsatellites are concerned [15,16]), isolation by distance or time (the Wahlund