Advances in PARASITOLOGY

VOL 36

Edited by

J.R. BAKER

Royal Society of Tropical Medicine and Hygiene, London, England

R. MULLER

International Institute of Parasitology, St Albans, England

and

D. ROLLINSON

The Natural History Museum, London, England

VOLUME 36

ACADEMIC PRESS

Harcourt Brace & Company, Publishers
London San Diego New York Boston
Sydney Tokyo Toronto

CONTENTS

CONTRIBUTORS TO VOLUME 36 v
PREFACE vii

Rare, New and Emerging Helminth Zoonoses

J.D. Smyth

1. Introduction 1
2. Trematode Zoonoses 2
3. Cestode Zoonoses 10
4. Nematode Zoonoses 12
5. Conclusions 34
References 35

Population Genetics of Parasitic Protozoa and Other Microorganisms

M. Tibayrenc

1. Introduction 48
2. What is the Problem under Study? 50
3. Techniques for the Study of Population Genetics of Microorganisms 51
4. A Paradigm of the Clonal Model: *Trypanosoma cruzi* . . . 64
5. Other Parasitic Protozoa 71
6. General Conclusion Concerning Parasitic Protozoa 81
7. Extending the Clonal Model: Pathogenic Yeasts 82
8. The Population Genetics of Bacteria 83
9. Emerging Debates 85
10. Two Main Kinds of Population Structure 93
11. The Relevance of Time and Space for Population Genetics and Strain Typing of Microorganisms 97

12. Population Genetics and the Notion of Species in Microorganisms . . . 101
13. Concluding Remarks . . . 102
Acknowledgements . . . 103
References . . . 103
Appendix: Glossary . . . 112

The Biology of Fish Haemogregarines

A.J. Davies

1. Introduction . . . 118
2. Life Cycles . . . 123
3. Structure and Development . . . 142
4. Seasonality . . . 164
5. Pathology . . . 167
6. Organisms that have been Confused with Fish Haemogregarines . . . 174
7. Conclusion . . . 182
Acknowledgements . . . 185
References . . . 192
Appendix . . . 201
Note Added in Proof . . . 203

The Taxonomy and Biology of Philophthalmid Eyeflukes

P.M. Nollen and I. Kanev

1. Introduction . . . 206
2. The Genus *Philophthalmus* . . . 207
3. Eyefluke Disease . . . 218
4. Adult Stage . . . 220
5. Egg Stage . . . 242
6. Miracidium . . . 244
7. Redia . . . 247
8. Cercaria . . . 253
9. Metacercaria . . . 257
10. Conclusions . . . 260
Acknowledgements . . . 261
References . . . 261

Human Lice and Their Management

I.F. Burgess

1. Introduction . . . 272
2. Biology . . . 272
3. Population Structure . . . 279
4. Pathology . . . 280
5. Clinical Aspects . . . 283
6. Transmission and Epidemiology . . . 287
7. Treatment and Control . . . 297
Acknowledgements . . . 321
References . . . 321

Ticks and Lyme Disease

C.E. Bennett

1. Introduction . . . 344
2. The Discovery: History . . . 345
3. Seasonality . . . 346
4. Lyme Disease in the USA . . . 346
5. Tick Life Cycles . . . 347
6. Spirochaete Life Cycles . . . 348
7. Incubation Period . . . 349
8. Pathology . . . 349
9. Genetic Predisposition to Severe Pathology . . . 354
10. Pathogenesis . . . 355
11. Treatment . . . 356
12. Prognosis . . . 357
13. *In Vitro* Culture . . . 357
14. Experimental Use of Ticks in Xenodiagnosis and in Giving Live Infection . . . 358
15. The Genome . . . 358
16. Strain Variation . . . 359
17. Serodiagnosis . . . 362
18. Examples of International Research Outside the USA . . . 366
19. Infected Ticks . . . 371
20. Tick Host Potential . . . 373
21. Animals Implicated as Reservoirs of Lyme Disease . . . 374
22. Incompetent/Non-susceptible (Though Often Antibody Positive) . . . 376

23. Spirochaetes per Tick 376
24. How Ticks are Infected 377
25. Monitoring the Cycles 378
26. Complex Modelling 379
27. Risk Assessment 379
28. Spatial Assessment 380
29. Prevention 381
30. Vaccination 382
References 383

Index 407

Rare, New and Emerging Helminth Zoonoses

J.D. Smyth

3 Braid Mount View, Edinburgh EH10 6JL, UK

1. Introduction 1
2. Trematode Zoonoses 2
2.1 Cercarial invasion 2
2.2 Mesocercarial invasion 3
2.3 Adult infection: eurytremiasis 9
3. Cestode Zoonoses 10
3.1 Species involved 10
4. Nematode Zoonoses 12
4.1 Ancylostomiasis 12
4.2 Angiostrongyliasis 15
4.3 Dirofilariasis 21
4.4 Cerebrospinal nematodiasis 26
4.5 Oesophagostomiasis 28
4.6 Brugian filariasis 33
5. Conclusions 34
References 35

1. INTRODUCTION

A zoonosis has been defined by WHO (1959, 1979) as: "Those diseases and infections [the agents of] which are naturally transmitted between [other] vertebrate animals and man". This definition separates a zoonosis from those infections, such as malaria, transmitted by invertebrate vectors, e.g. mosquitoes. At least 150 zoonoses have been recognized world-wide (Bell *et al.*, 1988), the best known of which are listed in the Annex in WHO

ADVANCES IN PARASITOLOGY VOL 36
ISBN 0-12-031736-2

Population Genetics of Parasitic Protozoa and other Microorganisms

M. Tibayrenc

UMR CNRS/ORSTOM 9926, Génétique moléculaire des Parasites et des Vecteurs, ORSTOM, Centre de Montpellier, 911 avenue Agropolis, BP 5045, 34032 Montpellier Cedex 01, France

1. Introduction 48
2. What is the Problem under Study? 50
3. Techniques for the Study of Population Genetics of Microorganisms 51
   3.1. Technical tools 51
   3.2. Concepts and statistics 53
   3.3. Possible biological obstacles to gene flow 60
   3.4. Possible biases 61
4. A Paradigm of the Clonal Model: *Trypanosoma cruzi* 64
   4.1. Circumstantial evidence for clonal propagation of *T. cruzi* 64
   4.2. Impact of clonal evolution on the biological properties of *T. cruzi* 70
5. Other Parasitic Protozoa 71
   5.1. *Trypanosoma brucei sensu lato* 71
   5.2. *Leishmania* spp. 73
   5.3. *Giardia duodenalis* 75
   5.4. *Plasmodium falciparum* 76
   5.5. *Toxoplasma gondii* 79
   5.6. Other species of parasitic protozoa 80
6. General Conclusion Concerning Parasitic Protozoa 81
7. Extending the Clonal Model: Pathogenic Yeasts 82
   7.1. *Candida albicans* 82
   7.2. *Cryptococcus neoformans* 83
8. The Population Genetics of Bacteria 83
9. Emerging Debates 85
   9.1. Are zymodemes and electrophoretic types reliable genotype markers or merely plastic phenotypes? 85
   9.2. Opportunistic infections in persons infected with HIV: a new model for microbial population genetics 86

ADVANCES IN PARASITOLOGY VOL 36
ISBN 0-12-031736-2

9.3. Does linkage disequilibrium equate with clonality? 87
10. Two Main Kinds of Population Structure 93
10.1. Non-structured species 93
10.2. Structured species 93
10.3. Possible additional categories 94
11. *The Relevance of Time and Space for Population Genetics and Strain Typing of Microorganisms* 97
11.1. Four different levels of analysis 98
11.2. Two different categories of genetic marker 99
11.3. Setting the molecular "clock" 100
12. Population Genetics and the Notion of Species in Microorganisms 101
12.1. Non-clonal microorganisms 101
12.2. Basically clonal microorganisms 101
12.3. Clonets and major clones 102
13. Concluding Remarks 102
Acknowledgements 103
References 103
Appendix: Glossary 112

1. INTRODUCTION

In the last 20 years, since the pioneering work by Kilgour and Godfrey (1973), a considerable amount of effort has been spent on exploring parasite diversity with biochemical or molecular tools. Although valuable results have been obtained in terms of epidemiology and basic science, the general outcome has been somewhat disappointing. Many basic questions remain unanswered. For example, is the pig a reservoir of human trypanosomiasis in Africa? Are clinical forms of Chagas disease specifically associated with particular strains of *Trypanosoma cruzi*? This field of research is at a standstill, and many epidemiologists are getting disenchanted with, and distrustful of, genetic studies.

It seems to me that three facts have hampered this line of research. First, the rather narrow application of strain typing has been overemphasized to the detriment of broader and richer approaches. Second, as a direct consequence of this, many workers have relied on an empirical, descriptive interpretation of the results only. Third, this field of research has been highly compartmentalized.

(i) Strain typing overemphasis. Although strain typing is very useful in epidemiological investigations, studies dealing with genetic variability of parasitic protozoa and other microorganisms are much more rewarding, not only in basic science (evolution and population genetics), but also in

applied research (virulence, resistance to drugs, immunological properties, etc.).

(ii) Empiricism. Overemphasis of strain typing is mainly responsible for this deficiency. But even within this application, empiricism remains a problem. There is a tendency merely to continue describing zymodemes*, schizodemes*, rapdemes*, karyodemes*, etc. The merely descriptive stage must be followed by construction of a model and clear, falsifiable working hypotheses must be stated. A theoretical study must, therefore, precede and accompany any benchwork dealing with genetic variability of microorganisms, and priority must be given to basic research.

(iii) Compartmentalization. Consider a few practical situations: a malariologist worried by the spread of chloroquine resistant malaria; a veterinarian in Africa surveying cattle disease caused by trypanosomes; an agronomist trying to elucidate an epidemic of heart-rot in coconut trees; a clinician noticing that a normally harmless yeast becomes lethal in patients with acquired immune deficiency syndrome (AIDS); a doctor finding that antibiotics no longer work to cure tuberculosis; a supermarket owner losing money because his stocks of Roquefort cheese or *rillettes* caused listeriosis in his customers. All these situations have in common that they are related to the genetic variability of the microorganisms involved. Although the same questions are asked, the answers are sought separately. With few exceptions, each microbe has its own group of researchers, with their own methods of analysis. "Leishmaniacs" use techniques that are different from those used by malariologists, and both are generally poorly aware of work dealing with genetic variability in bacteria, etc.

Since the problems are closely similar from one microbe to another, I have long advocated a common approach, with standardized techniques and statistics, in order to study comparatively the population genetics of microorganisms (Tibayrenc *et al.*, 1990, 1991a; Tibayrenc and Ayala, 1991). Such a common approach would both save effort and money and allow informative comparison which would reveal the "common denominators", the general laws governing microbe population diversity and evolution, as well as the peculiarities of each category of microorganisms.

The present paper is another attempt to reach this goal. Although it focuses mainly on medically important parasitic protozoa, some fungal organisms are considered, and extensive comparisons are made with population genetics of bacteria.

* Terms marked with an asterisk are defined in the Appendix.

2. WHAT IS THE PROBLEM UNDER STUDY?

In the present text, "sex" is used in a very broad sense and refers to any kind of genetic exchange. Bacterial conjugation will therefore be called "sex" here. The problem emphasized by population genetics, rather than sex itself, is the "downstream" impact of sex on the diversity of natural populations. Indeed, broad-sense sex is one of the main features that condition population structure and evolution.

The present paper hence has a rather different goal than a previous review (Baker, 1989), which focused on the sexual processes of parasitic protozoa rather than on their consequences for population diversity.

Whatever species is considered, population genetics can extend basic knowledge of microbial evolution, which is its explicit goal. But, apart

Table 1 The applied and basic aspects of population genetics of microorganisms.

Main applications of population genetics and evolutionary studies in microbiology	Translation in terms of basic science
Epidemiological tracking	
Checking for the stability of microbial genotypes over space and time	Structure and dynamics of microbial populations
Short-term level: nosocomial epidemiology[a]	Impact of genetic recombination on population structure; evolutionary role of sex
Long-term level: broad-scale epidemiology[b]	
Taxonomy	
Must be first based upon phylogeny[c]	Molecular phylogeny; evolutionary role of sex
Exploring the relationships between genetic diversity and the commonly accepted taxonomical nomenclature	
Looking for hidden genetic subdivisions within presently identified species	
Studies downstream from genetics	
Impact of genetic diversity and phylogenetic divergence on the relevant properties of microorganisms	
Virulence, resistance to drugs, immunological patterns, susceptibility to potential vaccines	Adaptative significance of microbial genetic diversity
Vector and host specificity	Vector/host/parasite co-evolution

[a] Time and space scales: days–months, hospital based.
[b] Time and space scales: months–years, village based.
[c] Time and space scales: millions of years, country- or continent-wide, up to the whole geographical range of the species.

from basic research, population genetics can provide valuable insights into three main applications (Table 1).

(i) Epidemiological tracking. When characterizing microbe genotypes to study epidemic spread, only population genetics is able to evaluate rigorously the stability of these genotypes over space and time. What is the use of strain characterization, if microbe genotypes have no stability because sex regularly re-forms their genetic make-up? The risk of this occurrence can never be ruled out, and it is especially high in some species (see below).

(ii) Taxonomy in a broad sense. Not only is population genetics useful in defining and delimiting currently described taxa, but one of its major applications is in the search for hidden genetic subdivisions within species.

(iii) Evaluation of the impact of the genetic diversity of microbes on their biological properties that are of practical importance (virulence, resistance to drugs, immunological diversity, etc.).

These three lines of research are closely linked to one another since, for example, it is vain to search for links between given microbe genotypes and virulence if these genotypes are unstable (unless the very genes* that govern virulence are studied), or if hidden, stable genetic subdivisions exist within a species, causing it to exhibit a range of distinct biological properties.

In this review, emphasis is placed on those aspects of population genetics which are more specifically relevant to microbiology. More general information about population genetics and phylogenetic methods can be obtained from textbooks in which the use of molecular markers is especially emphasized (Ayala and Kiger, 1984; Richardson *et al.*, 1986; Pasteur *et al.*, 1987; Hartl and Clark, 1989; Hillis and Moritz, 1990; Avise, 1994).

3. TECHNIQUES FOR THE STUDY OF POPULATION GENETICS OF MICROORGANISMS

3.1. Technical Tools

3.1.1. *Isoenzymes**

Isoenzyme analysis remains the "gold standard" for population genetics, especially in the case of microbes (see Figure 5), for three reasons. (i) Isoenzymes represent a universal marker, as they can be used, from a technical point of view, for any organism (see Section 9). (ii) Isoenzymes have been widely used for many years to study many different organisms;

thus it is possible to make informative comparisons between organisms whose formal genetics are well known (humans and fruitflies) and other organisms whose formal genetics are still obscure (many microorganisms). (iii) The mendelian inheritance and evolutionary behaviour of isoenzyme markers are well known.

3.1.2. *Random Amplification of Polymorphic Deoxyribonucleic Acid (RAPD)**

This new, fashionable marker is presently widely used in population genetics. It seems to me to have the same status, and the same hopes, as isoenzymes studies in the 1960s. RAPD has a double interest. (i) Since each primer generates a specific kind of variability, and the number of different primers that can be used is virtually unlimited, the discriminative level of the method itself is potentially unlimited. (ii) RAPD can be used for any organism. A promising peculiarity of RAPD variability, at least for parasitic protozoa (Tibayrenc *et al.*, 1993), is that many RAPD fragments convey valuable phylogenetic information, and appear to be specific to given phylogenetic subdivisions: species, intraspecific subdivisions, or individual genotypes (synapomorphic characters). These specific RAPD fragments can thus be used conveniently to design specific probes and diagnostic tools for use in the polymerase chain reaction (PCR).

The method at present has two drawbacks, which may be reduced in the future. (i) The technique is "touchy", and some people are disappointed in its lack of reproducibility. Nevertheless, in my experience, reproducibility is fair provided that the experimental conditions (especially the brands of *Taq* polymerase and thermocyclers) are strictly controlled. (ii) The mendelian inheritance of RAPD variability is difficult to elucidate for those organisms with which mating experiments are difficult or impossible (most microorganisms). This last drawback does not, however, prevent the use of population genetic statistics (see below).

3.1.3. *Restriction Fragment Length Polymorphism (RFLP)**

When used simply by cutting deoxyribonucleic acid (DNA) with restriction enzymes and reading the band profiles on agarose gels, the RFLP technique generally gives poorly informative, smeared patterns with many bands. A notable exception is the schizodeme technique (Morel *et al.*, 1980), an RFLP technique applied to purified kinetoplast DNA* which gives highly discriminative patterns with discrete bands. Two drawbacks of schizodeme analysis are that it explores the variability of an extranuclear genome and it is limited to the study of Kinetoplastida. Sometimes the term "schizodeme" is extended to include the result of any kind of RFLP analysis.

When RFLP is performed in conjunction with Southern hybridization and probes, its resolution is generally much better.

3.1.4. *Pulse Field Gel Electrophoresis (PFGE)**

Although promising, the PFGE technique gives results which are presently difficult to interpret in terms of population genetics, and have been used rather for empirical typing or gene mapping on chromosomes. The evolutionary behaviour of the variability recorded is still obscure. Nevertheless, some attempts have been made to interpret PFGE data in population genetics terms by Bastien *et al.* (1992) and Dujardin *et al.* (1993). The latter study suggested that PFGE polymorphism is adaptive, and is driven by environmental pressures.

PFGE studies on bacteria represent a different approach. In this case, the single bacterial chromosome is cut into large fragments by low-frequency cutting restriction endonucleases, and hence the fragments separated by PFGE do not represent chromosomes; this is actually a special kind of RFLP technique.

3.2. Concepts and Statistics

The techniques involved in microbe population genetics (described in Section 3.1) are standard. However, the theoretical basis of the study is far less codified than in population genetics of humans, mice, fruitflies, etc. It is a nascent and rather controversial field.

3.2.1. *General Principles*

As stated above, the main goal of microbe population genetics is to see whether natural populations of microorganisms are subdivided into discrete genetic lines between which gene flow is either restricted or absent. The question might be addressed by mating experiments in the laboratory; but such experiments, even when successful (Jenni *et al.*, 1986; Walliker *et al.*, 1987), show only that the potentiality for gene exchange is still present in the organism under study, and reveal nothing about the frequency and actual impact of these phenomena in natural microbe populations.

An indirect approach based on the study of natural populations is hence preferred. A null hypothesis is proposed, in which the population under study is considered as panmictic*, because this is the only situation for which statistical expectations are well codified. The null hypothesis is

evaluated by various statistical tests (see below). If the results are incompatible with panmixia*, the null hypothesis will be rejected; this is circumstantial evidence that gene flow is restricted in the population under survey, for whatever reason. The working hypotheses explored in this approach deal with biological obstacles to gene flow (either clonality or cryptic speciation). Explanations by either physical separation or natural selection are considered as biases which must be evaluated (see below).

A most important point is that inability to reject the null hypothesis (panmixia) is by no means a confirmation of this hypothesis. Such a result can very often be due to lack of resolution of the tests employed, or too small a sample, or both. This point is too often forgotten in statistics.

No population is perfectly panmictic, and the biases due to physical separation are discussed in Section 3.4.1. Throughout the remainder of this review, the term "panmictic" refers to potentially panmictic situations, in which the only obstacles to gene flow are physical ones — the situation obtaining in "normal" sexual species, such as humans, fruitflies, etc. Similarly, the term "sexual" refers to organisms in which gene exchange is obligatory, occurring at each generation. It is not used of organisms that are merely capable of gene exchange. Conversely, "non-panmictic" refers to a species subdivided into discrete genetic units (either cryptic species or clones) between which free gene flow is inhibited by biological obstacles.

The two main consequences of gene exchange in natural populations are segregation* of alleles* at given loci and recombination* of genotypes from one locus to another. Various statistical tests have been proposed by Tibayrenc *et al.* (1990) to explore these two biological phenomena. All these tests, listed in Table 2, are related to either the Hardy–Weinberg equilibrium* (segregation tests) or linkage disequilibrium* (recombination tests).

Table 2 Statistical tests (a–g) used to reveal departures from panmictic expectations (after Tibayrenc *et al.* 1990); for more details see Section 3.2.4.

Criterion	Description
Segregation (within locus)	
a	Fixed heterozygosity
b	Absence of segregation genotypes
c	Deviation from Hardy–Weinberg expectations
Recombination (between loci)	
d1, d2	Over-represented, identical genotypes widespread
e	Deficit of recombinant genotypes
f	Classical linkage disequilibrium analysis
g	Correlation between two independent sets of genetic markers

3.2.2. *Segregation Tests*

The classical Hardy–Weinberg statistics are applicable only when (i) the ploidy level of the organism under study is known, (ii) this level is greater than unity, and (iii) the alleles can be identified. These requirements are difficult to meet in the case of microorganisms. Bacteria have a haploid* genome, as do the stages of *Plasmodium* spp. which occur in humans. Hardy–Weinberg statistics are hence not valid for them. The level of ploidy is difficult to ascertain in most microorganisms. Even for the genera *Trypanosoma* and *Leishmania* diploidy* remains a mere working hypothesis (Lanar *et al.*, 1981; Maazoun *et al.*, 1981; Tibayrenc *et al.*, 1981a). Finally, even if a working hypothesis exists concerning the ploidy level of the organism involved, alleles are often difficult to discriminate in genetic studies of microbes. Even with isoenzymes, an allelic interpretation of the zymograms is always tentative since control mating experiments are either difficult (Jenni *et al.*, 1986) or impossible. For these reasons, segregation tests should be interpreted cautiously and used only as a complement to recombination tests.

3.2.3. *Recombination Tests*

These can be considered as more reliable than segregation tests for the reasons listed above. They can be used whatever the ploidy level of the organism, and even without identifying individual alleles and loci* (Tibayrenc *et al.*, 1990, 1991a, 1993; Stevens and Tibayrenc, 1995; Tibayrenc, 1995).

(a) *General procedure.* In random recombination, the expected frequency of a given genotype composed of n individual genotypes occurring at n different loci is the product of the observed frequencies of the individual genotypes which constitute it (the probability of occurrence of independent events). With isoenzymes, which remain the most widely used genetic markers, for a given enzyme system (which can be equated generally to an individual genetic locus), when it is difficult or impossible to discriminate individual alleles, each distinct and reproducible enzyme pattern is equated to a distinct genotype of which the allelic composition remains unknown. It is then possible to estimate the observed frequency of each genotype at given loci. Even when allelic interpretation is possible, this "blind" approach may be used, for it is the most parsimonious.

(b) *A practical example.* If two enzyme loci, A and B, are studied in two parasite strains, at each locus two different genotypes, 1 and 2, will be observed, and each will have an observed frequency of 0.5. The frequencies of the individual genotypes A1, A2, B1 and B2 are hence all 0.5, and the expected frequency of the composite genotype A1/B1 is $0.5 \times 0.5 = 0.25$ (see allelic frequency*), as are the frequencies of the other possible

combinations, A1/B2, A2/B2, and A2/B1. If the number of loci examined increases, the procedure still is the same. For example, if 10 loci, A to J, are studied, with two genotypes of equal frequency (0.5) at each locus, the frequency of any composite genotype will be only $(0.5)^{10}$, which makes this approach extremely powerful. Indeed, studies involving 15 to 20 different enzyme loci are commonplace, and the expected frequency of individual genotypes then becomes very low. The mere repetition of genotypes can thus be quite improbable, and becomes in itself a telling indication of departure from panmixia.

(c) *Cases in which identification of individual loci is impossible.* Even a given isoenzyme system is not always attributable to a unique locus. Several enzyme systems involve more than one locus. Sometimes it is easy to separate the study of the different loci involved in a given enzyme system; sometimes it is not. Linkage disequilibrium tests nevertheless remain possible. For a given enzyme system, each distinct and reproducible pattern is equated to a given elementary genotype, of which the composition in terms of alleles and loci remains unknown. The observed frequency of each elementary genotype is estimated, and the expected probability of the composite genotypes is the product of the observed probabilities of the elementary genotypes of which each is composed.

This procedure does not introduce any bias into the statistical tests, only a loss of information (for it is impossible to check for linkage among the possibly different loci that are plotted together for a given enzyme system).

In the case of RAPD, this procedure is the only one that can be used. Each primer amplifies DNA fragments whose relationships to identifiable loci are impossible to establish. It is even uncertain whether a given primer will amplify DNA segments of the same loci in different microbial stocks. Nevertheless, even in this extreme case, linkage disequilibrium tests (Table 2, f) can provide useful information (Tibayrenc *et al.*, 1993). As with isoenzymes (see above), for a given primer each distinct and reproducible pattern is equated to a given elementary genotype whose composition in terms of alleles and loci remains unknown. The expected probability of the composite genotypes is calculated as described above for isoenzymes.

(d) *A general principle for linkage disequilibrium analysis.* It does not matter whether linkage analysis is performed between loci or between groups of loci (see above). The only requirement is that it must be done between genetically independent sets of loci. If the loci or sets of loci are not independent of each other, considerable bias favouring linkage disequilibrium is introduced.

When individual loci can be discriminated, the problem is not large. In a sexual organism, it is generally considered that two loci must be tightly linked on the same chromosome to generate a statistically detectable linkage in natural populations (Hartl and Clark, 1989). This risk is considered very low if loci are randomly selected (Pujol *et al.*, 1993), and decreases geometrically when the number of loci studied increases.

When identification of individual loci becomes a problem, one has to be cautious. For example, the enzyme systems of the kinase family (hexokinase, fructokinase, etc.) can have overlapping specificities, and hence be partly redundant. It could therefore be misleading to perform linkage statistics between these systems; this could amount to doing the tests on the same loci, and the bias favouring linkage could hence be considerable.

This risk should not exist while performing linkage statistics involving different RAPD primers, as the probability of different primers involving the same locus is considered negligible (M. McClelland, personal communication).

This is not the case with RFLP studies involving the same hybridizing probe; the variability generated by the use of different restriction endonucleases cannot safely be considered to relate to totally independent sets of loci. The situation is different when different probes involving independent sequences are used.

3.2.4. *Some Peculiarities of the Tests*

(a) *Segregation tests.* These are all related to Hardy–Weinberg statistics. Nevertheless, the results may be so extreme as not to require statistical verification. Fixed heterozygosity is one of these cases: some genetic lines of *Trypanosoma cruzi*, for example, show constant heterozygous* patterns at certain isozyme loci (Tibayrenc and Ayala, 1988), which is incompatible with segregation (with random mating, even if the parents are heterozygous, there is an obligatory 50% of homozygous* segregants among the offspring). Fixed heterozygosity can help in deciding whether the genetic line under survey is a cryptic sexual species or a clone (Tibayrenc *et al.*, 1991a; see also Sections 3.3 and 9.3), with the reservation that such an assumption is dependent upon the working hypothesis of diploidy in *Trypanosoma* (see above).

(b) *Recombination tests.* These are all related to linkage disequilibrium, but explore different facets of it. They should hence be considered complementary to each other rather than redundant. In given situations and given data structures, some of them will be negative, while others will be positive. Only a single positive result is, in itself, a sufficient indication of statistical departure from panmixia.

Test d1 (Table 2) specifically checks for the spread over given geographical areas of genotypes that are over-represented in the sample according to panmictic expectations (Tibayrenc *et al.*, 1990). Indeed, as recalled recently by Maynard Smith *et al.* (1993), the mere observation of repeated genotypes is not in itself evidence of clonal propagation, and can be

statistically compatible with panmictic predictions. However, genotypes that are widespread and over-represented are strongly suggestive of clonal propagation, and this is especially telling when the geographical area involved is vast. The d1 test is performed either by a simple χ^2 test (when expected sizes are sufficient), or by a combination analysis according to the following formula:

$$P = \sum_{i=m}^{n} \frac{n! x^i (1-x)^{n-i}}{i!(n-i)!}$$

in which x is the expected probability of the multilocus genotype, as stated above (product of the observed frequencies of the single genotypes of which it is composed), n is the number of individuals sampled, and m is the numbers of individuals in the sample with the particular genotype.

Other tests listed in Table 2 include: d2, the probability of observing any genotype as often as, or more often than, the most common genotype in the sample; e, the probability of observing as few or fewer genotypes than actually observed (identical to the test designed by Cibulskis, 1988); f, the probability of observing a linkage disequilibrium level as high, or higher than, as that actually observed in the sample; and g, the correlation between independent sets of genetic markers. Tests d2, e and f are based on Monte Carlo simulations with 10^4 runs.

Tests d1, d2 and e are all based upon the observation of repeated genotypes. They are hence not directly usable when each stock represents a distinct genotype, and tend to be negative as data tend towards this extreme situation, which is not in itself evidence for sex. Clonal variability can be considerable and, if the genetic marker used has high resolution, every stock can exhibit a distinct genotype. The only condition to take into account in this case is linkage disequilibrium.

Even when repeated genotypes are lacking, tests d1, d2 and e can be performed by discarding the most discriminative loci. These tests can then be conveniently used to explore linkage among the rest of the loci, if repeated multilocus genotypes are revealed by this procedure.

Tests f and g can directly detect linkage disequilibrium, even if there are as many genotypes as individuals in the sample.

Test g (correlation between independent sets of genetic markers) is an especially telling example of linkage disequilibrium. Indeed, in the case of panmixia, the data related to a given marker (for example, isoenzymes) should have no predictive value on the data taken from another marker (for example, RAPD or RFLP), since the genes governing these distinct classes of marker should recombine independently. A convenient way to perform the g test is to estimate, between all possible stock pair-wise comparisons in the sample, the genetic distances for the two sets of markers (for

example, isoenzymes and RAPD), and to test their correlation by either a classical correlation test (Tibayrenc and Ayala, 1988; Tibayrenc *et al.*, 1993) or by a Mantel test (Mantel, 1967), as proposed by Stevens and Tibayrenc (1995) and Tibayrenc (1995). The second procedure (Mantel test) is more rigorous for, with the classical correlation test, it is difficult properly to evaluate the degree of freedom (the cells of the two matrixes of genetic distance are not totally independent of each other).

Test g can be further extended (Stevens and Tibayrenc, 1995; Tibayrenc, 1995) as a very general linkage disequilibrium test, by considering not only sets of data generated by technically distinct markers (for example, isoenzymes and RAPD), but also any two sets of distinct loci or groups of loci (for example, two sets of isoenzyme loci, or two sets of RAPD primers). To take a practical example, for a given set of strains, if results obtained from four different isoenzyme loci 1, 2, 3 and 4 are available, the correlation measured between the distances estimated from 1 and 2 on the one hand, and 3 and 4 on the other, will represent a measure of linkage disequilibrium in the same way as the correlation measured between isoenzyme and RAPD distances. It is possible to increase the resolution of this procedure by testing the correlation between all pairs of distances possible in the sample under study: not only 1–2/3–4, but also 1–3/2–4, 1–4, 2–3, etc. With populations at equilibrium, in which genetic recombination occurs at random (null hypothesis), these correlations will remain non-significant.

This extended g test makes it possible to avoid a major loss of information inherent in tests d1, d2, e and f. These last statistics take into account only two different classes of genotype: either identical, or non-identical. As an example, in Figure 1 it is obvious that genotypes A and B are more

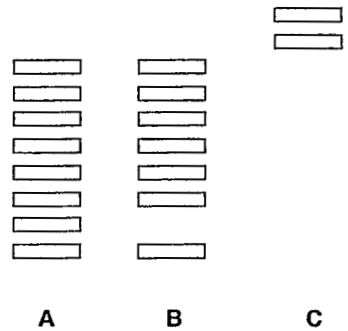

Figure 1 Three hypothetical genotypes corresponding to three different microbial strains. Profiles A and B are obviously more similar to each other than they are to C. Nevertheless, in the linkage disequilibrium tests d1, d2, e and f (see Table 2), they all fall into the same category: "non-identical genotype". The extended g test (Table 2) avoids this loss of information (see Section 3.2.3.(e)).

closely related to each other than they are to genotype C. With tests d1, d2, e and f, all three fall into only one category, "distinct genotypes". Test g, which is based upon the estimation of genetic distances, will take into account the closer similarity of genotypes A and B.

3.3. Possible Biological Obstacles to Gene Flow

3.3.1. *Clonality**

A clonal* population structure is the main working hypothesis that has been tested in microbe population genetics and that will be discussed in the present article (see Sections 4–8 for practical examples in various microorganisms). The tests described above have been designed chiefly to test for clonality.

Two important points about the clonal model proposed for several parasitic protozoa (Tibayrenc *et al.*, 1990) should be noted (see also Section 4.1.3): (i) the term "clone"* here has a broad genetic definition (see Appendix) and is not limited to mitotic propagation; and (ii) the "classical" clonal model does not imply that recombination never occurs in natural populations of microorganisms, but only that it is too rare an event to disrupt a prevalent pattern of clonal population structure. It is stated in the framework of this model that the natural clones exhibit a considerable degree of stability in space and time (Tibayrenc *et al.*, 1990). This has recently been a point of debate, which will be extensively discussed in Section 9.3, mainly in relation to the possible existence of "pseudoclonal models" differing slightly from the "classical" one in their inferred mechanisms, but also to a small extent with respect to the evolutionary consequences.

3.3.2. *Cryptic Speciation*

An alternative explanation of departures from panmixia, less frequently considered than clonality in the case of microbes, is cryptic speciation (Mayr, 1940). "Classical" speciation leads to genetic isolation, and can mimic many aspects of clonality. If two or more cryptic species are wrongly considered as a single panmictic unit, both departures from Hardy–Weinberg expectations and linkage disequilibrium will be observed. The evolutionary and epidemiological implications of cryptic speciation and clonality are in fact rather similar (see the full discussion of clonality vs. cryptic speciation in Section 9.3).

3.4. Possible Biases

All the tests listed above indicate nothing but departures from panmixia. In addition to the "true" biological obstacles to gene exchange (Section 3.3), the factors able to lead to departures from panmixia, and hence to cause positive results to be obtained in these tests, may be physical or biological in nature.

3.4.1. *Physical Obstacles to Gene Flow*

When populations are separated either by geographical distance or time, or both, they tend to accumulate different allelic frequencies (genetic drift). When such separated populations are wrongly considered a single panmictic unit, the tests listed above will indicate apparent departures from panmixia — deviations from Hardy–Weinberg expectations and linkage disequilibrium; this is referred to as the "Wahlund effect". The best way to avoid this bias is to design sampling conditions so that the stocks are sympatric* and are collected during a short period of time (Tibayrenc *et al.*, 1991a; Souza *et al.*, 1992). This is not always feasible, for many analyses of microbe population genetics have been performed in retrospect, using data from the literature that had not been collected for that purpose. Moreover, even with purpose-designed samples, the definition of sympatry* is not easy in the case of microorganisms; at the levels usually accepted to assess sympatry in higher organisms, it is not clear whether microorganisms have an actual opportunity for mating. This is most probably highly dependent upon the ecological behaviour and transmission cycle of each microbe species. Finally, strictly sympatric conditions make it impossible to evaluate the spread of given microbe genotypes over vast geographical areas and long periods of time, which is one of the main goals of microorganism genetic epidemiology.

If sympatry is not strictly ascertained, some tricks make it possible to evaluate the role of physical separation in generating departures from panmixia (Tibayrenc *et al.*, 1991a). (i) When segregation tests are considered, a Wahlund effect leads to a deficit in heterozygotes, rather than the converse (see Table 3, showing how an extreme case of allelic frequency difference in each of two geographically distinct populations can lead to a total absence of heterozygotes). An excess of heterozygotes, with its extreme case of fixed heterozygosity (see Section 3.2.4.(a)), is therefore an indication of biological obstacles to gene flow rather than of physical separation. (ii) When physical separation is responsible for a departure from panmixia, the over-represented genotypes identified by either segregation or recombination tests tend to be localized in restricted parts of the sampling area, whereas they can be widespread in the case of biological

Table 3 A model with two geographical locations (X and Y), one locus (A) and two possible alleles (1 and 2) at this locus, showing how extreme allelic frequency differences among populations can generate extreme departures from Hardy–Weinberg expectations when the two populations are unwittingly plotted together and considered as a unique population.

	Observed frequency		
	Population X	Population Y	Population X + Y
Locus/allele			
A1	1	0	0.5
A2	0	1	0.5
Genotype			
A1/1	1	0	0.5 (expected: 0.25)
A2/2	0	1	0.5 (expected: 0.25)
A1/2	0	0	0 (expected: 0.5)

obstacles to gene flow. Table 4 shows how an extreme case of fixed genotypes in each of two geographically distinct populations can lead to apparently total linkage between loci A and B.

Tables 3 and 4 show that there will be a tendency for the excess genotypes (either unilocus or multilocus) to be strictly localized, proportionally to the genetic drift and differences in allele or genotype frequencies among localities. In other words, to explain extreme departures from Hardy–Weinberg expectations or extreme linkage disequilibriums by geographical separation alone, one has to assume the occurrence of extreme

Table 4 A model with two geographical locations (X and Y), two loci (A and B) and two possible genotypes (1 and 2) at each locus, showing how extreme genotype frequency differences among populations can generate an extreme linkage disequilibrium when the two populations are unwittingly plotted together and considered as a unique population.

	Observed frequency		
Genotype	Population X	Population Y	Population X + Y
A1	1	0	0.5
A2	0	1	0.5
B1	1	0	0.5
B2	0	1	0.5
A1B1	1	0	0.5[a]
A2B2	0	1	0.5[a]
A1B2	0	0	0[a]
A2B1	0	0	0[a]

[a] Expected values are 0.25 in each case.

genetic drift, which can be often refuted by the observation of widespread genotypes.

3.4.2. *Biological Factors: Natural Selection*

Apart from factors (either physical or biological) that interrupt gene flow "upstream" from the actual gene exchange, natural selection could interfere "downstream", by selecting for or against some genotypes, so that genotype distribution no longer meets panmictic expectations. It is probable that natural selection interferes with genotype distribution, but it is hardly conceivable that this alone is able to explain extreme departures from panmixia (Tibayrenc *et al.*, 1990, 1991a). Indeed, to explain the maintenance of strong linkage disequilibriums over generations, one would have to accept that most of the possible multilocus combinations are eliminated in every generation (genetic load). Table 5 shows that, even for a limited number of loci, the genetic load makes this a not very parsimonious explanation.

Two particular cases of the natural selection hypothesis are the inferences that apparently extreme linkage disequilibriums are due to the elimination of many genotypes either by immunological defences or by culture medium selection. Although it is very probable that these two factors do interfere with genotype distribution, it is again difficult to accept that they would be able to maintain by themselves the considerable extent of linkage observed in many microbial species (see below). Indeed, most of the possible genotypes, and always the same ones, would have to be systematically eliminated either by the immunological response or during cultivation *in vitro*. Possible ways to evaluate the impact of these

Table 5 A model with two possible genotypes at each locus, showing the proportion of genotypes that has to be eliminated in every generation to maintain complete linkage disequilibrium.

No. of loci[a]	No. of possible multilocus combinations	Proportion of genotypes eliminated
2	4	0.5
3	8	0.75
4	16	0.875
5	32	0.9375
6	64	0.9687

[a] With two loci, there are four possible genotype combinations, A1B1, A2B2, A1B2, and A2B1; to maintain complete linkage (only the two first genotypes observed), 50% of the genotypes (A1B2 and A2B1) have to be eliminated; similarly for higher numbers of loci, as shown.

factors would be (i) to study microbe genotype distribution in immunocompromised patients (see Section 9.2) or (ii) to omit the culture step by typing strains isolated directly from the host with the aid of the PCR. The second procedure is made difficult by the facts that (i) multilocus analysis is required for population genetic analysis and (ii) if microbe DNA from a given patient is amplified by several primers involving different loci, there is a high risk that the patient will actually be harbouring several genotypes of the microbe under study, and that the different primers will not amplify DNA from the same genotype.

4. A PARADIGM OF THE CLONAL MODEL: *TRYPANOSOMA CRUZI*

Pioneering work by Miles *et al.* (1977, 1978, 1981) revealed considerable isoenzyme variability in *T. cruzi*, which provided a favourable background for further extensive population genetics studies.

T. cruzi exhibits the clasical manifestations of clonal propagation, namely drastic departures from Hardy–Weinberg expectations and extreme linkage disequilibrium. This was soon recognized (Tibayrenc *et al.*, 1981b), and has been subsequently confirmed with many larger samples (Tibayrenc and Desjeux, 1983; Tibayrenc *et al.* 1984a, 1985, 1986, 1993; Tibayrenc and Ayala, 1988).

4.1. Circumstantial Evidence for Clonal Propagation of *T. cruzi*

4.1.1. *Lack of Segregation*

An example of lack of segregation in *T. cruzi* is shown in Figure 2 (Tibayrenc *et al.*, 1981b): among 73 Bolivian isolates, many genotypes that could occur by segregation were lacking: *Pgm 1/2, 2/2, 1/3, 3/3, Me 1/2, Gpi 1/1*. Similar "missing" genotypes have been repeatedly identified in more than 500 stocks to date, and careful calculations have been made on ample sympatric samples from southern Bolivia (Tibayrenc *et al.*, 1984a) (Tables 6 and 7). The value of χ^2 for the *Gpi* locus is 198, with 8 degrees of freedom, giving a value of $P << 10^{-3}$. For the same locus, the expected overall number of observable genotypes was 37.5, while the observed number was 99. Conversely, the expected global number of absent genotypes was 61.5. Interestingly, the heterozygous genotypes *2/4* and *3/4* were heavily over-represented, which does not support the hypothesis that these results were due to genetic drift and the Wahlund effect (see Section 3.4.1. (a)). Similar conclusions can be reached about the *Pgm* locus. It is worth

noting that these large departures from panmixia are still observed (Table 7) when the tests are performed only on zymodemes 2, 2a and 2c of Tibayrenc *et al.* (1984a), which roughly correspond to genotypes 32, 33, 39 and 43 in Figure 3. The hypothesis thus tested was that this group of

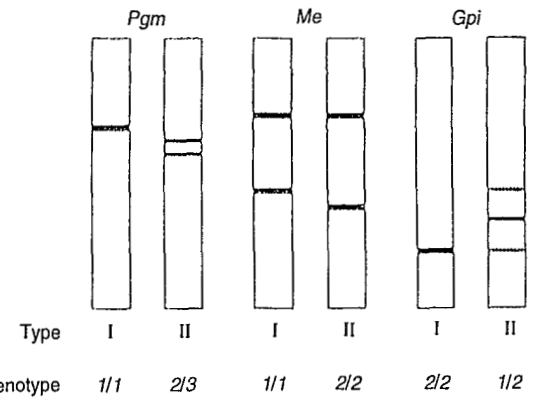

Figure 2 Diagrammatic representation of three different *T. cruzi* genotypes characterized for three enzyme systems, and showing lack of both segregation and recombination (linkage disequilibrium). (After Tibayrenc *et al.*, 1981b.)

Table 6 Hardy–Weinberg calculations for the *Gpi* and *Pgm* loci in 99 *T. cruzi* stocks from southern Bolivia. (After Tibayrenc *et al.*, 1984a.)

Genotype	Number observed	Number expected
Gpi locus		
3/3	9	2.5
5/5	45	21
2/4	31	6.8
3/4	14	7.2
2/2	0	2.3
4/4	0	5.3
2/3	0	4.8
2/5	0	13.6
3/5	0	14.6
4/5	0	20.9
Pgm locus		
1/1	45	20.9
3/3	9	10.1
2/3	45	13.9
2/2	0	4.9
1/2	0	20.1
1/3	0	29.1

Table 7 Hardy–Weinberg calculations on the *Gpi* and *Pgm* loci on 54 *T. cruzi* stocks from southern Bolivia, pertaining to the cluster of closely related zymodemes 2, 2a and 2c (see Figure 3). (After Tibayrenc *et al.*, 1984a.)

Genotype	Number observed	Number expected
Gpi locus		
3/3	9	4.9
2/4	31	12.8
3/4	14	13.2
2/2	0	4.6
4/4	0	9.1
2/3	0	23.1
Pgm locus		
3/3	9	18.2
2/3	45	26.3
2/2	0	9.5

zymodemes, which are more closely inter-related than they are to other zymodemes, might represent a distinct cryptic species (see Sections 3.3.2 and 9.3) rather than a cluster of clones; the hypothesis was not, however, corroborated.

4.1.2. *Lack of Recombination*

Natural populations of *T. cruzi* consistently show high levels of linkage disequilibrium. The *Pgm 1/1*, *Me 1/1* and *Gpi 2/2* genotypes are consistently associated with each other, as are *Pgm 2/3*, *Me 2/2* and *Gpi 1/2* (Figure 2). Although the method has been refined, the overall pattern has been verified with more than 500 stocks to date. Cross genotypes such as *Pgm 1/1 + Gpi 1/2* have never been observed. In other words, knowing the genotype at one of these loci makes it possible to predict the genotypes at the two other loci with a high probability of success, which is not the case in a sexual organism.

Linkage tests (see Table 2) performed on the 99 sympatric southern Bolivian stocks of *T. cruzi* (Tables 6 and 7) gave values of $P = 4.3 \times 10^{-15}$ for test d1 and $P < 10^{-4}$ for tests d2, e and f. In other words, the probability of sampling the dominant genotype more frequently than actually observed, assuming that the null hypothesis of free genetic exchange was valid, would be only 4.3×10^{-15}. Clearly, a considerable degree of linkage must exist in this parasite population, although the sampling was reasonably sympatric (within a circle of 20 km diameter).

Two facts are worth emphasizing concerning linkage in *T. cruzi*. (i) It has been verified in Amazonian sylvatic cycles also. In a study in my

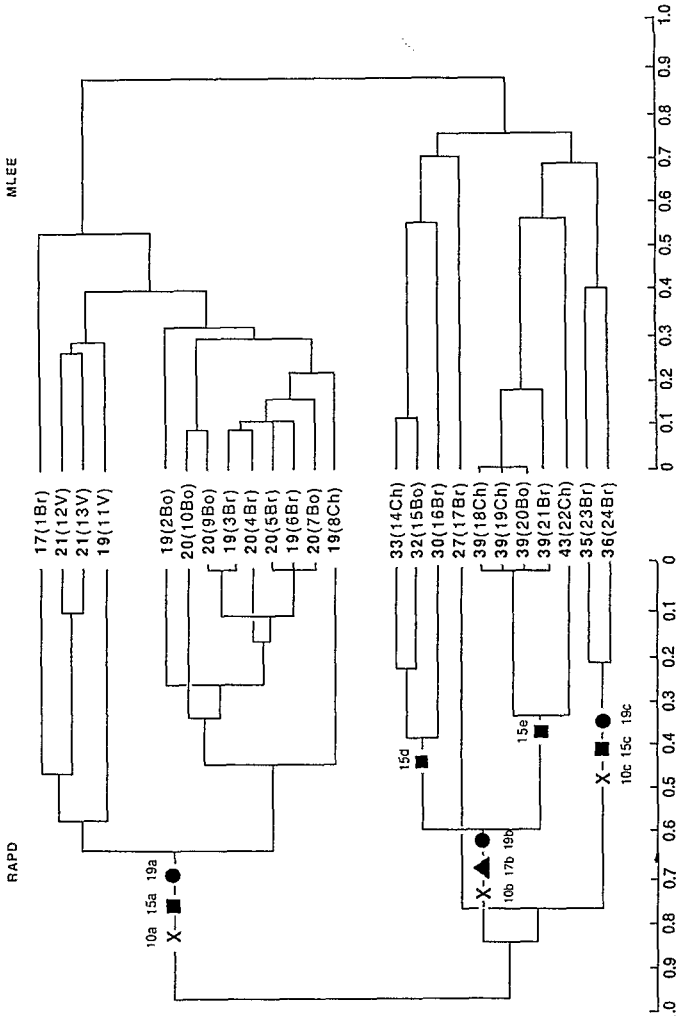

Figure 3 Dendrograms derived from genetic distances obtained by multilocus enzyme electrophoresis (MLEE) and random amplification of polymorphic DNA (RAPD) of 24 stocks of *T. cruzi*. Fair agreement between the two dendrograms is evidence of linkage disequilibrium. The symbols on the left dendrogram refer to RAPD fragments that have a synapomorphic value: they mark all the genotypes that pertain to a given evolutionary cluster. (After Tibayrenc *et al.*, 1993.)

laboratory on 26 *T. cruzi* stocks from French Guiana (Lewicka, 1991), the linkage tests d1, d2, e and f gave values of $P = 4.6 \times 10^{-13}$, 5×10^{-4}, 8×10^{-4}, and $< 10^{-4}$, respectively, supporting the hypothesis that clonal propagation was active in this situation also. (ii) It persists even in a situation of extreme sympatry, since the most distantly related *T. cruzi* zymodemes are currently some isolated in Bolivia from the same individual host, whether a triatomine bug or a human patient (Brenière *et al.*, 1985; Tibayrenc *et al.*, 1985). Another observation suggesting clonal propagation is the presence of over-represented, identical genotypes throughout vast geographical areas and long periods of time. For example, zymodeme 19 (Tibayrenc and Ayala, 1988) was isolated in Venezuela in 1976 and in Bolivia in 1983; zymodeme 20 was isolated in Bolivia in 1984 and in São Paulo, Brazil (date unknown); zymodeme 39 was isolated in Chile in 1977 and in Bolivia in 1983, etc. Such dominant, ubiquitous clonal genotypes of *T. cruzi* have been referred to as major clones (Tibayrenc and Brenière, 1988). This notion is quite different from the concept of "principal zymodeme" (Ready and Miles, 1980), which refers instead to predominant phylogenetic subdivisions of the species.

The final piece of persuasive evidence for clonal propagation of *T. cruzi* is the correlation between independent sets of genetic markers, a striking case of linkage disequilibrium (test g, Table 2). This has been shown to be true between (i) isoenzymes and kDNA RFLP (Tibayrenc and Ayala, 1988) and (ii) isoenzymes and RAPD (Tibayrenc *et al.*, 1993). Two phylogenetic trees constructed for 24 *T. cruzi* stocks from various sources are shown in Figure 3. Clustering patterns within each of the two trees are quite similar, due to the linkage between isoenzyme and RAPD characters. The correlation between the genetic distances inferred from the two sets of data was highly significant ($P < 10^{-4}$) (Tibayrenc *et al.*, 1993). PFGE data also show a high correlation with isoenzyme results, which has been taken as additional evidence for a clonal population structure in *T. cruzi* (Sanchez *et al.*, 1993).

4.1.3. *General Conclusions*

(i) *T. cruzi* appears to be composed of two major phylogenetic lineages, each extensively polymorphic and subdivided into smaller clusters (see Figure 3). The first major cluster, at the top of the trees in Figure 3, includes all those stocks more or less related to the formerly described zymodeme I (Miles *et al.*, 1977, 1981; Ready and Miles, 1980). The original zymodeme I stock corresponds to genotype No. 17 in Figure 3. The second cluster includes zymodemes II and III (genotypes No. 30 and 27, respectively). The genetic distances* separating the major lineages are

considerable, with values of Nei's standard genetic distance (Nei, 1972) up to 2 (Tibayrenc *et al.*, 1986). This value is about four times greater than that between humans and chimpanzees. Recent studies (C. Barnabé and M. Tibayrenc, unpublished data) using more markers (22 enzyme loci) and considerably more stocks (384), although they indicated considerably more variation (identifying a total of 258 different zymodemes), fully corroborated this overall pattern of two major lineages.

(ii) The most parsimonious hypothesis about the origin of the discrete phylogenetic lines of *T. cruzi* is long-term clonal evolution. Nevertheless, other hypotheses, such as short-term clonal propagation or the existence of two cryptic biological species, are also possible (Tibayrenc *et al.*, 1984b; see also Maynard Smith *et al.*, 1993), and will be discussed later (see Section 9.3.4.(c).(i)).

(iii) Circumstantial evidence for uniparental propagation does not mean that gene exchange never occurs in *T. cruzi*, but rather that it is not frequent enough to prevent the propagation of clones that are stable in space and time. This temporal stability could even reach an evolutionary scale. Nevertheless, even if this is generally true, occasional bouts of sex could still interfere with the evolutionary fate of the clones. This is a general caveat, applicable to all population genetic analyses. Throughout the rest of this review, the term "clonal" must be understood to mean "predominantly clonal".

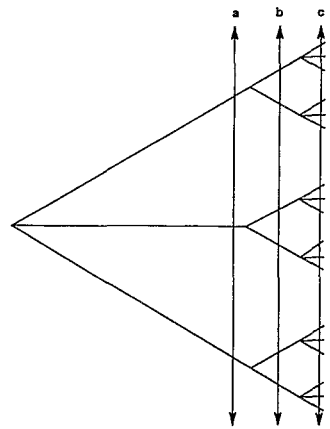

Figure 4 A hypothetical phylogenetic tree depicting the evolutionary divergence among some microbial genotypes. If a marker with a low molecular clock (= low resolution) is used, only three different genotypes will be distinguished (level a). If a more discriminative marker is used, six different genotypes will be detected (level b), and if a fast-evolving marker is used, the number will be increased to 18 (level c).

(iv) It is misleading to consider genetic lines characterized by a limited set of markers as "true" clones. They should rather be regarded as families of closely related clones. Improving the genetic labelling is certain to reveal additional variability within each of the previously characterized "clones" (see Figure 4). This has been verified in our laboratory in the case of the major "clone" 39 (Tibayrenc and Ayala, 1988), which has been split into 20 minor genotypes by using 22 isoenzyme loci instead of 15 (C. Barnabé and M. Tibayrenc, unpublished data). It is most important to emphasize this point in the context of recent debates on the perennialty of microbial natural clones (see Section 9.3).

4.2. Impact of Clonal Evolution on the Biological Properties of *T. cruzi*

It is reasonable to expect that the extent of phylogenetic divergence which has accumulated between the natural clones of *T. cruzi* will have an impact on this parasite's biological properties. Studies in my laboratory have been based on this working hypothesis. Possible correlations between genetic distances (= phylogenetic divergence) and biological properties were calculated by quantifying the biological properties, estimating the absolute differences between all pairs of stocks, and then calculating the correlation between these differences and the genetic distances for the same pairs of stocks. Sixteen stocks with various origins, representing three major clones (Tibayrenc and Brenière, 1988), were used in all these experiments. The main results are summarized below.

(i) There is a highly significant correlation between genetic distances on the one hand and the following properties on the other: epimastigote growth in culture medium, differentiation from epimastigote to trypomastigote, maximum parasitaemia in mice, infectivity to mice, mortality of mice (Laurent, 1994), and sensitivity *in vitro* of epimastigotes to both benznidazole and nifurtimox (S. Revollo and M. Tibayrenc, in preparation).

(ii) While the differences between distantly related genotypes are highly significant, the properties studied show great diversity among stocks within each genotype (= major clone). These results are in agreement with previous data presented by Andrade *et al.* (1983, 1985) and Andrade (1985). They are fully consistent with the working hypothesis, but they do show that, within each natural clone, there is considerable variability of the biological properties.

5. OTHER PARASITIC PROTOZOA

The methodology developed for *T. cruzi* has been applied to many other parasitic protozoa, as few species, apart from *T. brucei s.l.*, had been studied from the population genetic aspect. The main results are summarized in this section.

5.1. *Trypanosoma brucei sensu lato*

This species is the one which made sex in parasitic protozoa "fashionable", due to the innovative paper by Tait (1980), in which he proposed the hypothesis that *T. brucei* was a sexual, panmictic organism on the basis of the isoenzyme patterns of isolates sampled in the field. Although the hypothesis of panmixia has not been corroborated (see below), Tait's work opened the way to an entirely new field of research, since the occurrence of sex in *T. brucei* has undoubtedly been confirmed in the laboratory (Jenni *et al.*, 1986; Gibson and Garside, 1991).

However, under natural conditions in the field, strong departures from panmixia are apparent. Cibulskis (1988), noting that many of the possible recombinants were lacking from natural populations of *T. brucei*, proposed the existence of a potential for the evolution of distinct strains within this species.

Tibayrenc *et al.* (1990, 1991a), analysing various data from the literature (see especially extensive isoenzyme analyses by Gibson *et al.*, 1980), found both departures from Hardy–Weinberg expectations and strong linkage disequilibriums, and proposed that natural populations of *T. brucei* had a basically clonal structure.

Further studies in my laboratory, using genetic techniques previously used for *T. cruzi* (see Truc and Tibayrenc, 1993; Mathieu-Daudé and Tibayrenc, 1994), favoured this proposal and, although there were some peculiarities, it is beyond doubt that considerable linkage is a constant fact of populations of *T. brucei* (see Table 8). Nevertheless, while many loci showed departures from Hardy–Weinberg expectations, it is interesting to note that, at other loci, all possible allelic combinations were observed, an observation rarely if ever recorded with *T. cruzi*. Moreover, the phylogenetic divergence among *T. brucei* genotypes, although considerable (average Nei's standard genetic distance = 0.27 ± 0.14, maximum 1.15) (Mathieu-Daudé and Tibayrenc, 1994), remains much lower than the values observed with *T. cruzi* (0.76 ± 0.48, maximum 2.02) (Tibayrenc *et al.*, 1986). This difference, which is statistically significant, could be explained either by a less ancient evolutionary divergence of the clones, or

Table 8. Results of linkage disequilibrium tests for 14 different *T. brucei* populations. (After Mathieu-Daudé and Tibayrenc, 1994.)

Group	Sample size	Polymorphic loci	Observed genotypes	Size[a]		Probability[b]			
				Observed	Expected	d1	d2	e	f
East Africa	21	16	21	1	–	–	–	–	$< 10^{-4}$
Central/West Africa	59	14	40	6	0.03	1.9×10^{-12}	10^{-3}	$< 10^{-4}$	$< 10^{-4}$
West Africa	30	12	21	5	0.01	2.4×10^{-12}	2×10^{-2}	7×10^{-4}	$< 10^{-4}$
Central Africa	20	6	12	6	1.67	4.8×10^{-3}	9.2×10^{-3}	1.7×10^{-2}	1.9×10^{-3}
West Africa									
Human	30	9	20	5	0.32	1.5×10^{-5}	5.6×10^{-3}	$< 5 \times 10^{-4}$	$< 10^{-4}$
Domestic	179	7	59	29	5.25	10^{-13}	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
East Africa									
Human	409	8	92	45	1.06	2.1×10^{-57}	$< 10^{-4}$	$< 10-4$	$< 10^{-4}$
Domestic	120	10	54	11	0.13	3.4×10^{-18}	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
Wild	44	10	33	3	0.66	2.9×10^{-2}	0.18	$< 6 \times 10^{-4}$	$< 10^{-4}$
Tsetse flies	70	11	51	5	0.03	9.7×10^{-11}	2×10^{-3}	$< 10^{-4}$	$< 10^{-4}$
Animal (LV[c])	154	6	52	21	1.69	6.2×10^{-17}	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
Human (LV)[c]	24	7	6	12	4.54	5.7×10^{-4}	10^{-3}	6×10^{-4}	$< 10^{-4}$
1980									
1981	34	7	9	17	6.43	4.2×10^{-5}	$< 10^{-4}$	10^{-4}	$< 10^{-4}$
1982	26	6	5	12	4.73	10^{-3}	10^{-2}	10^{-2}	$< 10^{-4}$

[a] Of the most common genotype.
[b] See Table 2 for description of tests used; tests d2, e and f are based on computer simulations with 10^4 iterations for sample sizes < 50, and 10^3 iterations for the other groups.
[c] LV, Lambwe Valley.

by inhibition of genetic divergence by genetic recombination (see Section 9.3.3.(c)). These facts (the presence of segregant alleles and less genetic divergence) might, therefore, suggest that gene exchange interferes more actively in *T. brucei* than in *T. cruzi* with the long-term evolution of the clones. In this context, a recent hypothesis (Cibulskis, 1992) proposed that *T. brucei* undergoes short-term clonal propagation, but that the clones soon lose their genetic individuality due to active recombination (see also Maynard Smith *et al.*, 1993 and Section 9.3).

In relation to the same problem of gene exchange frequency, more genotype diversity is observed in "wild" *T. brucei* stocks than in stocks isolated from humans (Mathieu-Daudé and Tibayrenc, 1994). A similar observation has been reported for *T. cruzi* by Lewicka (1991). For both parasites, it is at present difficult to decide whether these results are due to more frequent bouts of sex (Gibson, 1990) or, simply, to greater clonal diversity (see Section 3.2.4) in response to higher ecological variability. With both parasites, it is worth noting that, despite the higher genotype variability in "wild" populations, the levels of linkage in "wild" and "non-wild" populations are comparable when estimated by tests that do not rely on genotype repetition (e.g. tests f and g in Table 2; see Section 3.2.4).

Unfortunately, rigorous statistical means to estimate precisely the frequency of occurrence of sex in basically clonal populations are presently lacking. Nevertheless, some observations make it possible to address the question. This problem will be more extensively considered in Section 9.3.

In summary, although temporal stability of *T. brucei* clones still is a matter of debate (see Section 9.3), it is hardly acceptable that this organism is a sexual, panmictic one. This illustrates a general fact that has to be kept in mind for any microorganism: successful mating experiments (e.g. those by Jenni *et al.*, 1986) show that the potentiality for gene exchange is present in the species under study, but say nothing about the actual impact of this in natural populations. This last question is better explored by population genetic means.

5.2. *Leishmania* spp.

From a population genetics viewpoint, this genus of parasites presents a special case, for it has been subdivided into many Linnean species. Originally, the existence of these species was inferred from clinical and/or epidemiological arguments. Nevertheless, genetic criteria, principally isoenzyme patterns, become more and more important in the identification of species of *Leishmania* (see Peters and Killick-Kendrick, 1987). This means that *Leishmania* species have increasingly been given a phylogenetic

basis (see Section 11). However, many of the clades that correspond to given species exhibit very limited genetic diversity. For example, when direct comparisons are made using the same genetic techniques, as has been done in my laboratory (Guerrini, 1993), the *L. donovani/infantum/chagasi* "complex" appears to be genetically far more homogeneous than *T. cruzi*. Therefore, given that the power of resolution of the tests used in population genetics, like any statistical test, is very dependent upon the richness of the available information, the chances of demonstrating departures from panmixia in *L. infantum* are less than in *T. cruzi*, all other things being equal. At the extreme, when the genetic unit studied is monomorphic, all tests are impossible. The more the situation approaches this extreme, the less is the chance of obtaining significant results. I have extensively discussed this methodological difficulty (Tibayrenc, 1993), and it will be further developed in Section 9.3.

Despite this obstacle, a survey of data from the literature (see especially Maazoun *et al.*, 1986; Moreno *et al.*, 1986; Pratlong *et al.*, 1986, Desjeux and Dédet, 1989) reveals ample circumstantial evidence for clonality within many *Leishmania* species (Tibayrenc *et al.*, 1990, 1991a). As an example, within the poorly polymorphic *L. infantum*, the results of tests d1, d2 and e (Table 2) are 10^{-6}, 10^{-4} and 10^{-2}, respectively (Tibayrenc *et al.*, 1990). These results have been criticized on the basis that the stocks had not been sampled in sympatric conditions (Bastien *et al.*, 1992). Nevertheless, the hypothesis that linkage is due to geographical separation is hardly compatible with the fact that the dominant genotype (zymodeme MON 1, Moreno *et al.*, 1986) is widespread throughout the area surveyed (see Section 3.4.1). Anyway, more recent data obtained in my laboratory from more sympatric isolates by Guerrini (1993) and A.L. Bañuls (unpublished data), using the same genetic techniques as those used with *T. cruzi* and *T. brucei*, revealed clear indications of clonality for several *Leishmania* species, including *L. amazonensis, L. braziliensis, L. guyanensis, L. panamensis* and *L. peruviana*, despite the smallness of the samples. In spite of the fact that a clonal structure is the most parsimonious hypothesis to account for genetic diversity among *Leishmania* spp., the results discussed above need much refinement: the samples considered were often limited and were never designed specifically for population genetic purposes. The possible impact of gene exchange events must definitely be explored in depth in *Leishmania* populations, since growing evidence suggests that these events do occur. The evidence for this includes: (i) the observation under the microscope of cell fusion in *L. infantum* (see Lanotte and Rioux, 1990); (ii) PFGE results, showing putative recombinant patterns in isolates from southern France (Bastien *et al.*, 1992) and Peru (J.C. Dujardin, personal communication); and (iii) repeated reports of biochemical evidence for the formation of "hybrids" between different species (Evans *et al.*, 1987; Darce *et al.*, 1991). Similar evidence for the existence of putative hybrids between *L. panamensis* and *L. braziliensis* has been obtained in my laboratory by isoenzyme and RAPD studies (Bañuls, 1993) (Figures 5 and 6).

5.3. *Giardia duodenalis*

Giardia duodenalis exhibits genetic variability patterns which are comparable to a large extent to those of *T. cruzi*, which led to its being included in the framework of the clonal model (Meloni *et al.*, 1989; Tibayrenc *et al.*, 1990; Tibayrenc, 1994b). All the classical manifestations of clonal propagation are apparent, including strong linkage disequilibrium, widespread over-representation of genotypes (Tibayrenc *et al.*, 1990), and correlation between independent sets of markers (Meloni *et al.*, 1989). Moreover, as in

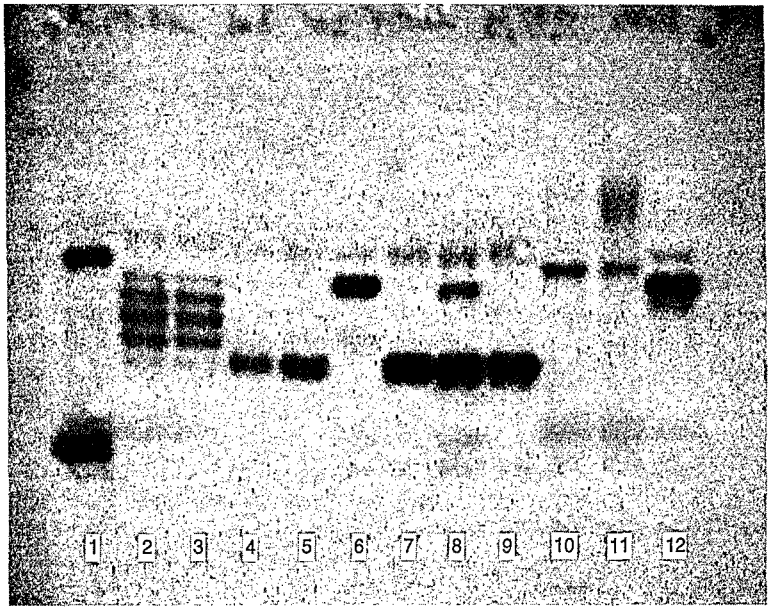

Figure 5 Isoenzyme analysis at the *Nh 1* locus, showing the existence of putative hybrids between *L. panamensis* and *L. braziliensis*. The *L. panamensis* putative parental profile is represented by samples 6 and 12 and the *L. braziliensis* putative parent by samples 4, 5, 7 and 9. The heterozygous patterns (putative hybrids; samples 2 and 3) exhibit five bands, for the enzyme is a tetramer. See also Figure 6. (After Bañuls, 1993.)

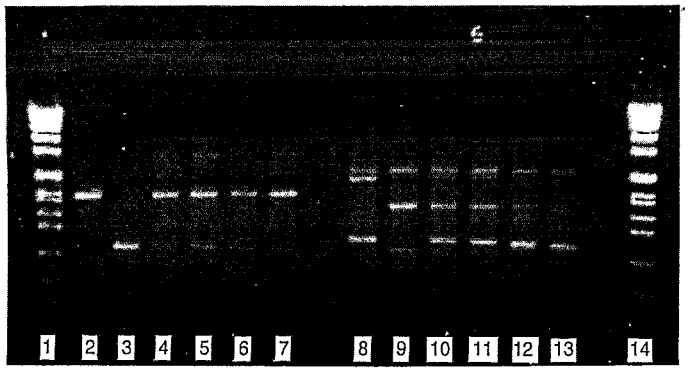

Figure 6 Random amplification of polymorphic DNA with two different primers, showing evidence of the existence of putative hybrids between *L. panamensis* and *L. braziliensis*. Lanes 1 and 14 are DNA scale ladders. Lanes 2–7, primer A5; lanes 8–13, primer A7. Lanes 2 and 8, *L. panamensis*; lanes 3 and 9, *L. braziliensis*; lanes 4–7 and 10–13, four putative hybrids. See also Figure 5. (After Bañuls, 1993.)

T. cruzi, the putative clonal lineages are separated from each other by large genetic distances (Andrews *et al.*, 1989).

Because of this last observation, an attempt has been made to subdivide *G. duodenalis* into four cryptic species defined by their phylogenetic divergence (Andrews *et al.*, 1989). The question of whether these putative species should be considered as sexual entities or as small clonal clades has been discussed by Tibayrenc (1993) (see Section 9.3).

Although clonality is a parsimonious hypothesis in *Giardia*, the question of this parasite's population structure should not be considered as definitively settled (nor is it settled for any species). The impact of culture selection on genotype distribution (see Section 3.4.2) seems especially high in *Giardia* (see Mayrhofer and Andrews, 1994) and its interference with population genetic statistics is an important topic for future studies.

5.4. *Plasmodium falciparum*

The malaria parasites obviously constitute a special case among parasitic protozoa. Sex is considered to be an obligatory event in each transmission cycle and recombinants have been successfully obtained in the laboratory (Walliker *et al.*, 1987). A panmictic population structure has been clearly postulated for this species, with all its logical epidemiological implications — in particular, the fact that panmixia would prevent the development of distinct strains (Walliker, 1985). This last inference was, however, based

on weak direct evidence, for few studies have been made of association between loci in natural populations (linkage disequilibrium), the only way to check for departures from panmixia. The stages developing in culture *in vitro* are haploid, which makes the application of Hardy–Weinberg statistics impossible (see Section 3.2.2). Two studies involving limited numbers of loci (Carter and McGregor, 1973; Carter and Voller, 1975) have produced results consistent with panmictic expectations.

On the contrary, the observation of statistically significant linkages reported in the literature led to the proposal that some kind of uniparental propagation could take place in certain populations of *P. falciparum* (see Tibayrenc *et al.*, 1990). An alternative explanation for these observations was proposed at the same time, namely the existence of cryptic species within the taxon *P. falciparum*. It must be borne in mind that, for this species as for any microorganism, the term "clonality" in population genetics is consistent with many reproductive systems that involve apparent mating, including self-fertilization (Tibayrenc and Ayala, 1991). Indeed, self-fertilization in a haploid organism generates genetic clones. It should hence be considered a specific case of clonality (Tibayrenc and Ayala, 1991), rather than an alternative model to it (Dye, 1991).

A lively debate followed the publication of these "non-panmictic" proposals (Dye *et al.*, 1990; Walliker *et al.*, 1990; Dye, 1991; Tibayrenc and Ayala, 1991; Tibayrenc *et al.*, 1991b; Walliker, 1991; Day *et al.*, 1992). This led to desirable additional investigations in the field.

Conway and McBride (1991) and Babiker *et al.* (1991) reported results supporting panmictic assumptions. The first study relied on the use of antigen markers, while the second was based upon isoenzymes, antigens and proteins. On the contrary, in my laboratory, Ben Abderrazak (1993) showed, by multilocus enzyme electrophoresis, the existence of statistically significant linkage disequilibriums in five populations among six. The only population showing agreement with panmictic expectations was that studied by Babiker *et al.* (1991). It has to be noticed that, in non-panmictic populations, the linkage observed involved usually three or four loci, while the rest of the loci, although fairly variable, showed limited linkage. This pattern is quite different from that seen in *Trypanosoma* or *Leishmania*, in which linkage is seen at all loci. For example, in a population of 31 stocks of *P. falciparum* from the Congo Republic, sampled under reasonably sympatric conditions (within a circle of 20 km diameter) and in which 12 enzyme loci were studied, linkage was almost restricted to the loci *Ldh*, *Gsr* and *Gdh*, which were almost totally linked (see Table 9).

A new way of addressing the problem has recently been developed, namely PCR amplification of oocyst genes isolated directly from the mosquito (Ranford-Cartwright *et al.*, 1993). The goal is to estimate the proportion of homozygous and heterozygous genotypes and to compare it

Table 9 Genotype distribution of three enzyme loci in a population of *P. falciparum* from the Congo Republic, with the probability of the results (assuming panmixia). (After Ben Abderrazak, 1993.)

Genotype	Size		Probability	Test used
	Observed	Expected		
1/1/1	18	11.68	10^{-2}	X^2
2/2/2	8	0.47	1.4×10^{-8}	d1[a]
Other	5	18.8	$<<10^{-3}$	X^2

[a] See Table 2.

with the expected proportions under panmictic expectations. Divergent results were obtained in Tanzania (Babiker *et al.*, 1994) and Papua New Guinea (R. Paul, M. Packer and K. Day, personal communication), with rates of heterozygosity of 64% and 7%, respectively. This difference could have been due to epidemiological peculiarities in each country: both are considered to be highly endemic areas, but the transmission rate in Tanzania is higher.

Both rates of heterozygosity, although that in Tanzania was high, were much lower than expected in a panmictic situation. Indeed, given the considerable number of alleles present at these loci, and their frequency, virtually all oocysts should be heterozygous. The explanation given by Babiker *et al.* (1994) was that the number of different *P. falciparum* genotypes that are present in a given human host is limited, and that multiple feedings by anopheline mosquitoes are rare.

Several comments can be made concerning these results. (i) They (especially the data from Tanzania) confirm that outcrossing is a frequent event in *P. falciparum* natural populations. (ii) They nevertheless show (particularly in Papua New Guinea) the occurrence of a potential "bottle-neck" effect for outcrossing, since the self-fertilization rate, whatever may be the biological explanation, was much higher than expected. (iii) They provide information on crossing events in the mosquito, but are poorly informative about the actual consequences of these events "downstream", that is to say on the parasite population structure in humans, which is the relevant fact from an epidemiological point of view. The consequences of the "bottle-neck" effect due to self-fertilization, and of possible differential viability of the oocyst genotypes revealed by PCR, must be explored in depth by classical population genetics studies involving multilocus characterization of stocks from human hosts (Ben Abderrazak, 1993).

An innovative recent study is bound to relaunch the debate on *P. falciparum* population structure. Gupta *et al.* (1994a,b) and Gupta and Day (1994) have produced a model of antigenic variation in *P. falciparum*

with considerable implications for vaccine design. Although not primarily a consideration of the authors (Gupta *et al.*, 1994a,b), this model has potentially important implications for population structure. It implies the simultaneous presence, in close sympatry and in a highly endemic area (Papua New Guinea), of distinct strains of the parasite, a situation which is, at first sight, (Tibayrenc, 1994a) incompatible with the panmictic model (Walliker, 1985, 1991). Indeed, in the context of panmictic assumptions, in order to explain the maintenance of sympatric but distinct strains, it is necessary to infer that the antigen variability studied by Gupta *et al.* (1994a,b) and Gupta and Day (1994) is either monogenetic or governed by a few, tightly linked genes. If it is a polygenetic character, the various genes involved should be shuffled and separated by recombination in every generation (Tibayrenc, 1994a). Natural selection for certain variants could help to maintain some linkage among them but, if this were the only factor involved, it would be necessary to assume that most of the possible recombinants were eliminated in every generation (see Section 3.4.2 and Table 5).

In conclusion, the question of the population structure of *P. falciparum* remains unanswered. Preliminary results suggest that populations of this parasite are clearly less structured than those of *Trypanosoma* and *Leishmania* (see Ben Abderrazak, 1993). Nevertheless, the same results are hardly consistent with panmictic assumptions, which are furthermore questioned by the more recent results described above concerning both oocyst genetic variability (Babiker *et al.*, 1994; R. Paul, M. Packer and K. Day, personal communication) and antigenic variation (Gupta *et al.*, 1994a,b; Gupta and Day, 1994; Tibayrenc, 1994a). Given the considerable epidemiological implications of this debate, effort should be made to elucidate the population structure of *P. falciparum*, perhaps the most important parasite of humans.

5.5. *Toxoplasma gondii*

Among the limited data in the literature suitable for population genetic analysis (Dardé *et al.*, 1988, 1990), reports of extreme linkage disequilibrium and the existence of a dominant genotype distributed in both France and the USA led Tibayrenc *et al.* (1991a) to propose that *Toxoplasma gondii* has a clonal population structure. This proposal was corroborated by Sibley and Boothroyd (1992), who observed moreover that the two main clonal lineages they delimited by genetic analysis were associated with distinct pathogenic properties in mice.

It is at present not possible to decide definitely whether these two genetic categories correspond to sexual cryptic species or to two groups of closely

related clones, although preliminary indications favour the second hypothesis (Tibayrenc, 1993) (see Sections 5.3 and 9.3).

5.6. Other Species of Parasitic Protozoa

Data in the literature which are suitable for population genetic studies of other species of parasitic protozoa are limited (Tibayrenc *et al.*, 1990, 1991a; Tibayrenc and Ayala, 1991). It is possible to discuss only two interesting examples.

5.6.1. Entamoeba histolytica

Isoenzyme studies on *E. histolytica*, although at first based on a limited number of loci, made it possible to propose a new hypothesis, namely that some genotypes (zymodemes) were constantly linked to virulence, while others were constantly avirulent (Sargeaunt and Williams, 1979). This proposal was criticized on the assumption that *E. histolytica* zymodemes were actually plastic phenotypes subject to modification by culture conditions (Mirelman, 1987; Mirelman and Burchard, 1987), a hypothesis that itself received ample criticism (Sargeaunt, 1987; Clark *et al.*, 1992; Clark and Diamond, 1993). This point enters the framework of a broader debate that will be extensively discussed in Section 9.1. The existence of two distinct classes of *E. histolytica* genotypes linked to virulence, and attributable to distinct species, is a currently accepted working hypothesis (Diamond and Clark, 1993) that corroborates Sargeaunt and Williams' hypothesis (1979), as well as Brumpt's (1925) earlier views. In a phylogenetic study based on a fair range of isoenzyme markers, Blanc (1992) postulated that the virulent and avirulent species corresponded to two distinct clusters separated from one another by a large genetic distance. Specific molecular markers have now been developed to characterize pathogenic strains of this parasite (Garfinkel *et al.*, 1989; Tachibana *et al.*, 1992).

The population genetics of *E. histolytica*, although potentially able to throw some light on the above-mentioned problems, is still an unexplored field. Analysis of the limited data in the literature suggests clonality (Tibayrenc *et al.*, 1990, 1991a); there are clear linkage disequilibriums in *E. histolytica* populations from Canada (Proctor *et al.*, 1987) and South Africa (Sargeaunt *et al.*, 1982). This linkage persists when pathogenic and non-pathogenic genotypes are analysed separately, which suggests that the two putative cryptic species inferred by Diamond and Clark (1993) and Blanc (1992) are clonal rather than sexual (Tibayrenc, 1993). These assumptions should be considered as preliminary ones, due to the scarcity

of available information. Gene exchange in the laboratory has been suspected to occur in *E. histolytica* (see Sargeaunt, 1985; Sargeaunt *et al.*, 1988), which again is not inconsistent with a clonal population structure (see the *T. brucei* case, discussed in Section 5.1). Nevertheless, extensive studies are required to settle the question of the population structure of *E. histolytica*.

5.6.2. Naegleria *spp.*

Amoebae of the genus *Naegleria* constitute an interesting case, for they are predominantly free-living organisms that become parasitic only secondarily. Their population structure is hence informative, as a test of the hypotheses that clonality is either a fundamental feature of protozoa, or a secondary adaptation to parasitism (Tibayrenc *et al.*, 1991a).

Naegleria was one of the first parasites to be examined by a classical population genetic approach (Cariou and Pernin, 1987). This study led to the conclusion that the species *N. lovaniensis* was a sexual organism undergoing regular mating.

The situation seems to be different for other species of the same genus: *N. gruberi, N. australiensis* and *N. fowleri* show some classical indications of clonality, namely fixed heterozygosity and a widespread identical genotype (Tibayrenc *et al.*, 1990). This hypothesis, based on the limited data in the literature (Cariou and Pernin, 1987), has been amply corroborated in the case of *N. gruberi* by more extensive data (M.L. Cariou and P. Pernin, personal communication). *N. gruberi*, at least, is a well-documented case of a basically clonal free-living protozoon.

6. GENERAL CONCLUSION CONCERNING PARASITIC PROTOZOA

From this survey, it appears that "predominantly clonal", or at least "non-panmictic", situations constitute a kind of common denominator for several major species of parasitic protozoa, apart from the obviously peculiar case of *P. falciparum*. This overall conclusion should be considered as provisional rather than definitive. It needs much refinement, and "common denominator" does not mean that peculiarities do not exist in particular cases. Moreover, further debate is now questioning and enriching this approach (see Section 9). Nevertheless, it is reiterated here, as a parsimonious and falsifiable working hypothesis, that, apart from *P. falciparum*, a case that is still under discussion, natural populations of many parasitic protozoa are structured by severe biological obstacles to gene flow, and cannot be regarded as panmictic units.

7. EXTENDING THE CLONAL MODEL: PATHOGENIC YEASTS

As stated at the beginning of this review, it is desirable to use a common approach to address the question of genetic variability in any microorganism, for the problems involved are common, both in terms of basic and applied research.

Pathogenic yeasts are a medical problem mainly in immunocompromised patients. The AIDS epidemic has led to the development of research dealing with these microbes. As with parasitic protozoa, compartmentalization of this research has been an obstacle to rapid progress, and should be avoided in the future.

7.1. *Candida albicans*

Extensive efforts have been made to develop strain identification methods for this yeast, based on either isoenzyme electrophoresis (Lehmann *et al.*, 1989) or molecular markers (see, among many others, Stevens *et al.*, 1990; Soll *et al.*, 1991; Odds *et al.*, 1992). To my knowledge, few studies have relied on a population genetic approach to the study of strain diversity in *C. albicans*. Analysis of published isoenzyme results (Lehmann *et al.*, 1989) led to an unexpected result, since no sexual stage is known in *C. albicans*: the data showed no linkage disequilibrium, which was unique among many different microorganisms considered in that study (Tibayrenc *et al.*, 1991a). More recently, Caugant and Sandven (1993) characterized by enzyme electrophoresis a wide sample of stocks of *C. albicans* from northern Europe, and found only weak evidence for linkage. A totally different result was reported by Pujol *et al.* (1993), who analysed stocks of *C. albicans* isolated from patients in Montpellier (France) infected with human immunodeficiency virus (HIV), and applied to their isoenzyme data the tests developed by Tibayrenc *et al.* (1990) (see Table 2 and Section 3.2.4). Reasonable evidence of clonality was found, with considerable linkage disequilibrium. The level of significance of test d1 was 3×10^{-23}, while it was $<10^{-4}$ for tests d2, e and f (Table 2). Interestingly, test f remained at the same level of significance even if the only two repeated genotypes were removed from the sample, which is against the hypothesis that this population of *C. albicans* had an "epidemic" structure (Maynard Smith *et al.*, 1993; see the extensive discussion in Section 9.3).

It remains to be determined whether the sample examined by Pujol *et al.* (1993) constitutes a special case, or if the hypothesis of clonality can be extended to the whole species *C. albicans*. Preliminary results (C. Pujol, personal communication) have shown that clonality exists also in samples from countries other than France.

7.2. *Cryptococcus neoformans*

Similarly to *Candida albicans*, many studies have attempted to develop strain characterization tools for *Cryptococcus neoformans*, based either on isoenzyme electrophoresis (Safrin *et al.*, 1986; Brandt *et al.*, 1993) or molecular methods (Meyer *et al.*, 1993; Currie *et al.*, 1994). A first attempt to apply population genetic concepts to the study of this yeast, based on the isoenzyme data published by Safrin *et al.* (1986), showed clear indications of linkage disequilibrium within each of the two main serogroups of *C. neoformans* (see Tibayrenc *et al.*, 1991a). Brandt *et al.* (1993), using more extensive data, reached the conclusion that *C. neoformans* var. *neoformans* possessed a clonal population structure, while *C. neoformans* var. *gattii* was not clonal. This difference between *C. neoformans* var. *neoformans* and *C. neoformans* var. *gattii* could be questioned, for in the data used by Brandt *et al.* (1993) strong linkage is also apparent in the *gattii* group (electrophoretic types (ETs) 13 to 19), which appears to be far from panmictic (M. Tibayrenc, unpublished data). Actually, within the *gattii* set, two linkage groups appear, ET 13 to 16 on the one hand and ET 17 to 19 on the other hand. Many possible genotype compositions are lacking, while the genotypes actually observed are heavily over-represented. The probability of observing such a genotype distribution according to the d1 test (Table 2) is only 2.2×10^{-7}. Of course this result, based on only eight stocks, must be verified with more extensive samples.

8. THE POPULATION GENETICS OF BACTERIA

To a large extent, it is possible to find striking similarities between bacterial and parasitic protozoan population structures (Tibayrenc *et al.*, 1986, 1990; Tibayrenc and Ayala, 1988; Hartl, 1992; Maynard Smith *et al.*, 1993), and this has led to a call for a synthetic approach to construct a unified population genetics of microorganisms (Tibayrenc and Ayala, 1991).

Bacterial population genetics, initiated by Milkman (1973), has been far less compartmentalized than comparable studies dealing with eukaryotic microbes: standardized techniques (mainly starch gel electrophoresis) and common statistical methods of analysis have been applied to many different species, which makes it easier to draw direct comparisons between

them. Statistical analyses, although different from those propósed for parasitic protozoa (Tibayrenc *et al.*, 1990; see Table 2), rely on the same principles, and explore multilocus linkage disequilibrium, which is taken as major circumstantial evidence for clonal propagation. Hardy–Weinberg statistics are not applicable to bacteria, which are haploid organisms (see Section 3.2.2). In the literature of bacterial population genetics, stocks showing the same isoenzyme profile are referred to as "electrophoretic types" (ETs), a concept identical to "zymodemes" in medical protozoology.

The "clone concept" (Ørskov and Ørskov, 1983) has become a kind of paradigm in bacterial population genetics. Indeed, the clonal model, first established for *Escherichia coli* by Selander and Levin (1980), Whittam *et al.* (1983), Hartl and Dykhuizen (1984) and Ochman and Selander (1984), has been extended to many other bacterial species, including, for example, *Legionella pneumophila* (see Selander *et al.*, 1985), *Haemophilus influenzae* (see Musser *et al.*, 1985), *Neisseria meningitidis* (see Caugant *et al.*, 1986), *Yersinia enterocolitica* (see Caugant *et al.*, 1989) and *Borrelia burgdorferi* (see Dykhuizen *et al.*, 1993). The main conclusions of bacterial population genetics have been reviewed by Selander *et al.* (1987) and Young (1989).

Although many studies have led to the conclusion that the species investigated were clonal, different situations have been recognized even by classical population genetic approaches. As examples, absence of detectable linkage disequilibrium has been observed in *Neisseria gonorrhoeae* and *Pseudomonas aeruginosa*, which have therefore been considered as non-clonal species (Selander *et al.*, 1987). Lack of linkage disequilibrium has also been reported within each of the main subdivisions of *Rhizobium meliloti* by Eardly *et al.* (1990). In *Rhizobium leguminosarum*, Souza *et al.* (1992) explored in depth the different components of linkage disequilibrium, and reached the conclusion that, although some genetic isolation was apparent in close sympatry, a notable part of the overall linkage was due to geographical separation and genetic drift (see Section 3.4.1) rather than to biological obstacles to gene flow.

Moreover, analysis of fine gene structure has shown that some bacterial genes have a complex, "mosaic" structure due to recombination among different clonal lineages (Dykhuizen and Green, 1991; Maynard Smith *et al.*, 1992; Milkman and Bridges, 1993). So, despite the clonal paradigm, it became increasingly apparent that gene exchange (sex in a broad sense) played a major role in bacterial evolution, but one which probably differed from one species to another. This led Maynard Smith *et al.* (1993) to propose several distinct models of population structure to account for linkage disequilibrium in microbe populations. These proposals are discussed in the following section.

9. EMERGING DEBATES

9.1. Are Zymodemes and Electrophoretic Types Reliable Genotype Markers or Merely Plastic Phenotypes?

Isoenzymes are the "gold standard" in population genetics, and have been a major tool in the exploration of genetic diversity in microbes. Although isoenzyme profiles represent a phenotypic variation, their use in population genetics relies on the assumption that they reflect directly the variability of the genes involved in enzyme synthesis. Nevertheless, several recent studies dealing with parasitic protozoa overtly deviate from this consensual view, and could constitute, in the long term, a real school. These studies imply that, with the same underlying genotype, zymodemes can undergo drastic switches under environmental pressures as trivial as culture medium changes. This has been claimed for the main *T. cruzi* zymodemes (Romanha *et al.*, 1979; Alves *et al.*, 1993) and for those of *E. histolytica* (see Mirelman, 1987; Mirelman and Burchard, 1987). For both parasites, it is postulated that radical zymodeme switching is accompanied by drastic changes in biological properties such as virulence.

Obviously these puzzling results deserve attention, since they pose a challenge to the whole body of population genetic evidence in parasitic protozoa. Indeed, the two following proposals are strictly mutually exclusive: (i) *T. cruzi* zymodemes are plastic phenotypes and the same genotype can generate drastically different zymodemes under environmental pressures (Romanha *et al.*, 1979; Alves *et al.*, 1993); (ii) *T. cruzi* zymodemes are distinct genotypes that can be used conveniently for population genetic analyses and most probably represent distinct phylogenetic lineages (Tibayrenc *et al.*, 1981b, 1984a,b, 1986; Tibayrenc and Ayala, 1988).

The debate extends beyond the limited case of *T. cruzi*. A similar controversy exists about *E. histolytica*, with exactly the same alternatives. The proposals by Romanha *et al.* (1979), Alves *et al.* (1993), Mirelman (1987) and Mirelman and Burchard (1987) potentially cast doubt on any population genetic analysis of microbe isoenzyme variability.

There are several arguments against the hypothesis of zymodeme phenotypic plasticity in *T. cruzi*. First, in my personal experience of 14 years study of the isoenzymes of this parasite, my colleagues and I have never observed dramatic zymodeme switches in stocks whose cloning had been verified under the microscope. Such changes were sometimes observed in non-cloned stocks, but this is commonplace for a parasite in which mixed genotypes frequently occur in the same host (Brenière *et al.*, 1985; Tibayrenc *et al.*, 1985). Second, and most importantly, the

impressive correlation between isoenzyme markers on the one hand and molecular markers on the other makes the hypothesis of zymodeme phenotype plasticity hardly tenable. This correlation has been reported in studies involving kinetoplast DNA (Tibayrenc and Ayala, 1988), RAPD (Tibayrenc *et al.*, 1993; see Figure 3), and PFGE of DNA (Sanchez *et al.*, 1993). Acceptance of the hypothesis of zymodeme phenotype plasticity would necessitate accepting the view that changes in culture conditions are able to upset not only the isoenzyme profiles but also the patterns revealed by all these different DNA markers that probe diverse parts of the genome.

Clark and Diamond (1993) have suggested that the results reported for *E. histolytica* by Mirelman (1987) and Mirelman and Burchard (1987) could be accounted for by improper laboratory cloning. Nevertheless, before accepting this as a general explanation for the zymodeme phenotype plasticity hypothesis, work is in hand in my laboratory to check carefully the correlation between isoenzyme and molecular markers under various different conditions of culture, using the very stocks studied by Alves *et al.* (1993).

9.2. Opportunistic Infections in Persons Infected with HIV: a New Model for Microbial Population Genetics

The spread of HIV epidemics could lead to a reassessment of the population genetic models proposed for microorganisms. Immunodeficiency must have a major impact on the genetic diversity of the microbes that cause opportunistic infections in patients infected with HIV. It can be inferred that many of the microbial genotypes eliminated by the immune system in non-immunocompromised subjects will be able to survive in HIV patients. These microbial populations could therefore give a more "realistic" picture of the actual genetic diversity of the species under study. Although not a parsimonious hypothesis (see Section 3.4.2 and Table 5), it cannot be ruled out that linkage disequilibrium is generated in microbial natural populations, not by biological obstacles to mating (either clonality or cryptic speciation), but rather by elimination of most of the possible genotypes by immune defences. Comparing microbial populations from patients with and without HIV infection will make it possible to test this hypothesis under rigorous conditions.

Apart from this question of basic knowledge, population genetic analysis of opportunistic microbial populations will provide valuable information for the epidemiological tracking of these pathogens in patients infected with HIV.

9.3. Does Linkage Disequilibrium Equate with Clonality?

If one discards the biases imputable to either geographical separation or natural selection (Section 3.4), and if one accepts that multilocus linkage is barely explicable by very close proximity on the same chromosome of multiple genes (Section 3.2.3.(d)), linkage indicates a departure from panmixia, the extent of which is proportional to the level of linkage observed. Although clonality has been the main working hypothesis considered in microbial population genetics (see Sections 4–8), it is not the only situation able to generate departures from panmixia within a given species. Moreover, even in the framework of the clonal model, strong linkage is not considered evidence of absolute clonality and, in the case of parasitic protozoa, it has been repeatedly stated that some recombination could interfere with the evolutionary fate of the clones (Tibayrenc *et al.*, 1986, 1990, 1991a; Tibayrenc and Ayala, 1988, 1991).

9.3.1. *Alternative Models of Microbe Population Structure*

Maynard Smith *et al.* (1993) have recently described the following four possible models of microbe population structure, and have proposed a statistical approach to distinguish them: (i) clonal model, with clones stable on an evolutionary scale; (ii) cryptic speciation, in which the species under study is actually subdivided into two or more genetic

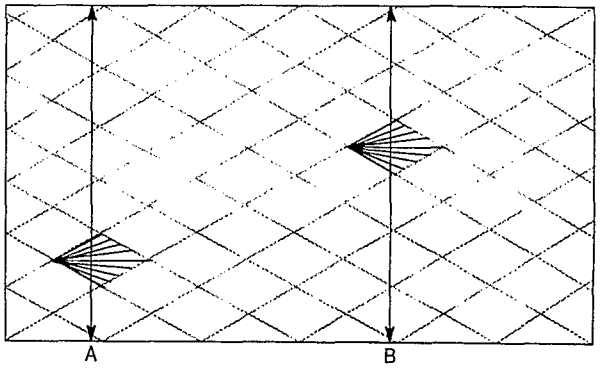

Figure 7 Evolutionary pattern of the epidemic model (Maynard Smith *et al.*, 1993; see also Tibayrenc *et al.*, 1984b). In a basically sexual species, occasional bouts of uniparental propagation generate "epidemic" clones (symbolized by dark lines), the lifetime of which does not exceed a few years; their genetic make-up is then shuffled in the common gene pool. If samples are examined at times A or B, the presence of repeated clonal genotypes will have increased the level of linkage disequilibrium of the population, although this population is basically sexual.

lineages, genetically isolated from one another, with each being panmictic; (iii) epidemic model, in which, in a basically sexual species, some genotypes propagate clonally for a short time (at most a few years; J. Maynard Smith, personal communication), after which their genetic make-up is broken by recombination (see Figure 7); (iv) sexual model, with a population as panmictic as that of a "normal" sexual species (humans and fruitflies). Cases (i), (ii) and (iii) are identical to the hypotheses A, B and C discussed by Tibayrenc *et al.* (1984b) in the case of *T. cruzi*, while models (i) and (ii) have been discussed by Tibayrenc (1993) in connection with *Entamoeba, Giardia* and *Toxoplasma*. Model (iii) is similar to that proposed by Cibulskis (1992) for *T. brucei*.

9.3.2. *Statistical Approach*

Maynard Smith *et al.* (1993), while describing a new linkage disequilibrium measurement, proposed more specific ways to distinguish between models (i), (ii) and (iii).

(i) Distinguishing model (i) from model (ii). The statistics are performed, not on the whole array of genotypes, but within each of the phylogenetic lineages that subdivide it (for example, in Figure 3, they might be performed separately within each of the two main *T. cruzi* clusters). If the linkage disappears within the subdivisions, this favours model (ii); that is to say these subdivisions should be regarded as panmictic units (sexual species), rather than groups of clones.

(ii) Distinguishing model (i) from model (iii). It is assumed that model (iii) will result in the recent propagation of "epidemic clones", with all the members of each clone having identical genotypes. If all individuals sharing this genotype are taken into account in the population genetics statistics, this will bias the data in favour of linkage. It is hence proposed to take as the unit not the individual (stock) but rather the genotype (= zymodeme = ET). If the linkage disappears, it is taken as evidence that this linkage was due to epidemic clonality rather than "true" clonality.

In the study by Maynard Smith *et al.* (1993), species of the model (i) type included *Salmonella* spp. and *T. cruzi*. The only species in model (ii) was *Rhizobium meliloti* (divisions A and B), with no parasitic example. Model (iii) species included, for example, *Neisseria meningitidis* and *T. brucei*. Lastly, sexual, panmictic species (model (iv)) included *N. gonorrhoeae* and *P. falciparum*. It is worth noting that the only *P. falciparum* population analysed by Maynard Smith *et al.* (1993) was also found to be panmictic using the linkage statistics listed in Table 2 (Ben Abderrazak, 1993).

9.3.3. *Discussion*

The approach proposed by Maynard Smith *et al.* (1993) raises several problems that are closely related.

(a) *Statistical type-1 error.* In this approach in general, impossibility to reject the null hypothesis of panmixia is taken as confirmatory evidence for the hypothesis. There is need for caution here: negative statistical results can be due to a lack of resolution rather than to an incorrect working hypothesis (see, for example, the detailed analysis of this problem in the case of *T. brucei* by Cibulskis, 1988).

(b) *Lack of data weighting.* The level of resolution of the linkage statistics proposed, as with any statistics, is heavily dependent upon the amount of information available. In particular, it is influenced by the number of alleles at each locus and the number of loci surveyed. Two observations can be made in this connection on the proposals by Maynard Smith *et al.* (1993).

(i) The cases analysed were not directly comparable, since the information available was not identical. As examples, only nine enzyme loci were considered in the case of *N. gonorrhoeae* while 24 were examined in *Salmonella*. A negative result in the tests could therefore be due to lack of resolution (statistical type-1 error) rather than to panmixia. Data should be weighted to make these comparisons fully informative.

(ii) The test to distinguish model (i) (clonality) from model (ii) (cryptic speciation) leads automatically to a reduction in the amount of genetic diversity available for comparison. This problem has been discussed by Tibayrenc (1993): for example, it is impossible to know whether the four main *G. duodenalis* lineages equated to four distinct species by Andrews *et al.* (1989) represent either sexual species (model (ii)) or groups of clones (model (i)), because there is no genetic variability within each of them. So, no statistics are possible. More generally, the more the population tends towards monomorphism, the more difficult it becomes to perform any test. This makes direct comparison between the species as a whole and its subdivisions difficult, for the amount of genetic diversity and the population sizes are automatically lowered by the process of subdivision. Data weighting is hence again needed in this case.

(c) *Difficulty in reliably identifying epidemic clones.* In distinguishing model (i) ("true" clonality) from model (iii) (short-term, epidemic clonality), counting each genotype only once, while there may be numerous individuals with each genotype, leads to the loss of a considerable amount of the information available (with the risk of a statistical type-1 error; see above), and proportionally lowers the level of linkage (for repeated genotypes favour the existence of linkage).

This approach is acceptable only if it is assumed that identical genotypes

must be the result of a very recent epidemic episode. This assumption may be misleading, for the distinction between identical and non-identical genotypes is very dependent upon the level of resolution of the markers used (see Figure 4); clones characterized by a given set of markers should be considered rather as families of closely related clones (Tibayrenc and Ayala, 1988). This difficulty led to the proposal of the notion of the clonet* (Tibayrenc and Ayala, 1991; see Section 12.3). Improving the level of resolution of the technique used generally reveals more genotypes.

To illustrate the relative, rather than absolute, nature of the notion of identical genotypes, it is informative to discuss the complementary way of distinguishing clonality (model (i)) from epidemic propagation (model (ii)) proposed by Maynard Smith *et al.* (1993). This consists of taking as the unit for the tests neither individuals nor genotypes, but clusters of closely related isoenzyme genotypes (ETs). This approach relies on the working hypothesis that the variability observed within each of these clusters of closely related genotypes (for example, the difference between levels b and c of resolution in Figure 4) has been generated in a few years (the time-scale of epidemic propagation).

Similarly, in a recent study by Hide *et al.* (1994), several *T. brucei* populations have been characterized by both multilocus enzyme electrophoresis (MLEE) and DNA "fingerprinting". Certain genotypes (zymodemes) were repeatedly observed by MLEE, while the "fingerprint" revealed as many different genotypes as there were individual stocks. This situation can again be illustrated by Figure 4, in which the isoenzymes could represent level b of resolution, while the "fingerprint" patterns could represent level c. In the linkage tests performed by Hide *et al.* (1994), the zymodemes were used as a unit rather than the "fingerprint" genotypes, and the results have been taken as evidence of epidemic rather than clonal structure (Maynard Smith *et al.*, 1993). This procedure again implies the underlying working hypothesis that, within each zymodeme, the additional variability revealed by DNA "fingerprinting" (represented by the difference between the b and c levels of resolution in Figure 4) has been generated in a few years, which is probably not the case.

9.3.4. *Proposed Additional Approaches*

(a) *Data weighting.* The ideal situation, which is difficult to attain, is when samples with the same amount of available information are compared. Faced with already published data, or when comparing subdivisions of a species with the entire species (distinguishing model (i) model (ii)), the only possible method is artificially to lower the quantity of information in the bigger sample, so that it is comparable to that of the smaller one. For example, if two species are compared, and the first has been surveyed at 20 loci while the second has been characterized by 8 loci only, statistics should be performed with only 8 loci for the first species and, if possible, these loci should be the same as those used for the second species.

Features that should be taken into account for statistical weighting are: (i) population size; (ii) number of variable loci; (iii) average genetic distance and its standard deviation (zero distances being eliminated); (iv) number of different genotypes at each locus; (v) relative frequency of the different genotypes observed at given loci (when one genotype is dominant, the situation tends towards monomorphism — the more informative situation is when the different genotypes at a given locus have the same frequency); and (vi) multilocus genotype diversity (number of different genotypes divided by the number of individuals).

(b) *Technical standardization.* Since data from different laboratories are sometimes difficult to compare, the ideal situation is when different species are compared in the same laboratory with similar techniques. This is the approach presently emphasized in my laboratory.

(c) *Alternative approaches to distinguish clonality from epidemic propagation.* (i) Looking for true phylogenetic divergence. In model (i), the clones can be equated to divergent phylogenetic lineages, while in model (iii) they amount to mere individual variants (they have exactly the same genetic status as identical twins in the human species). As noted by Tibayrenc *et al.* (1984b), in model (iii) (epidemic propagation in a basically sexual species) the genetic distances separating the genotypes have no phylogenetic value. Two conclusions can therefore be drawn.

First, it is improbable that large genetic distances will be observed, as regular genetic recombination will have inhibited much divergence among the genotypes. For example, if the different clones of *T. cruzi* were the result of short-term epidemic spread from the common gene pool of a basically sexual species, one would have to infer considerable genetic variability to account for the huge genetic distances that are commonplace among the genotypes of this species (Tibayrenc *et al.*, 1984b, 1986) (see Section 4.1.3.(b)).

Second, since genetic distances have no phylogenetic value in model (iii), there is no reason to expect a correlation between independent sets of genetic markers (test g in Table 2), such as MLEE and RAPD. The risk of this special approach to linkage being biased by the presence of short-term epidemic clones (which would have an identical genotype detected by both markers) can be avoided by counting repeated genotypes only once (based on the marker with the higher resolution), as recommended by Maynard Smith *et al.* (1993). From a statistical point of view, the approach would be made especially informative by selecting genotypes that represent a fair sample of the whole range of genetic distances recorded.

Correlation between different kinds of marker has been extensively

demonstrated in *T. cruzi* by Tibayrenc and Ayala (1988) and Tibayrenc *et al.* (1993), which supports the hypothesis that linkage in this species is not due to epidemic propagation (see Section 4.1.3.(b)). Interestingly, strict correlation between isoenzyme and RAPD diversities was also noted within each of the two main subdivisions of *T. cruzi* (see Tibayrenc *et al.*, 1993), which suggests that these two clusters represent sets of clones (the models (i) and A of Maynard Smith *et al.*, 1993 and Tibayrenc *et al.*, 1984b respectively), rather than cryptic panmictic species (models (iii) and C (see Section 4.1.3.(b)).

With *T. brucei*, there is also a highly significant correlation between genetic distances inferred from isoenzymes and RAPD (F. Mathieu-Daudé and M. McClelland, personal communication). In the *T. brucei* sample studied, every stock represented a distinct RAPD genotype, and so the risk of biasing the correlation by counting identical genotypes several times was avoided. This result again suggests that *T. brucei* genotypes have a certain stability at the evolutionary level. Again, this does not mean that sex is absent from this parasite's natural populations; but it is evidence against the hypothesis that linkage disequilibrium is the result of short-term epidemic spread in a basically sexual species (see Section 5.1).

(ii) Reliably identifying genotypes that result from a short-term epidemic spread. To use safely the means proposed by Maynard Smith *et al.* (1993) for distinguishing model (i) ("true" clonality) from model (iii) (epidemic propagation), it is crucial to ascertain to what extent genotypes or clusters of closely related genotypes characterized by a given genetic marker can be equated to short-term, epidemic clones. If this condition is not fulfilled, the approach proposed by Maynard Smith *et al.* (1993) could sometimes merely non-specifically evaluate the overall strength of linkage disequilibrium. This strength could equally well be reduced by more frequent, even though occasional, bouts of sex in a basically clonal species. The only means safely to identify epidemic clones, and hence conveniently to use the approach of Maynard Smith *et al.* (1993), is to select a marker with an appropriate molecular "clock" (level of resolution). This critical point is discussed in Section 11.

The approach proposed by Maynard Smith *et al.* (1993) appears most promising, and it is fairly probable that the biological situations described by models (i)–(iii) do exist in nature. Moreover, this approach is worth using even in its present state of development, provided that the results are interpreted cautiously. Hopefully, the refinements and alternative approaches proposed in Section 9.3.4 will make it possible to distinguish more sharply between the four kinds of population structure proposed by Maynard Smith *et al.* (1993).

10. TWO MAIN KINDS OF POPULATION STRUCTURE

It is important clearly to redefine the main alternatives that must be considered in relation to population structure. For example, although the occasionally heated debate on clonality vs. sexuality could obscure the fact, it is worth emphasizing that, in the cases considered by Maynard Smith *et al.* (1993), the epidemic model (iii) is closer to the panmictic model (iv) than to the clonal model (i). It is therefore proposed that future studies focus on the two main kinds of population structure discussed below (Sections 10.1 and 10.2).

10.1. Non-structured Species

This category contains those species which do not appear to be subdivided by real phylogenetic lineages. In this situation, the genetic distances observed among genotypes have no phylogenetic value, and only reflect individual variability (Tibayrenc *et al.*, 1984b). These genotypes can hence be equated to individual variants. The species grouped here are either panmictic or epidemic (with short-term clonality; Maynard Smith *et al.*, 1993). Although the evolutionary implications of the two models are not very different, for epidemiological tracking purposes it is important to distinguish panmictic from epidemic structures: the genotypes found in panmictic structures are quite ephemeral, while those of epidemic clones may last for a few years.

10.2. Structured Species

This category includes species of which the opposite is true: they are subdivided into true distinct phylogenetic lineages, and the genetic distances between these lineages convey real evolutionary information. These discrete lines might represent either cryptic biological species or clonal lineages. The implications of cryptic speciation and clonal evolution are largely comparable in terms of evolution and applied research.

Distinguishing between these two main divisions can be attempted by population genetic means, such as test g (Table 1: correlation between independent sets of genetic markers like isoenzymes and RAPD). Since divergent evolutionary lines occur in the second category, additional evidence can also be provided by the classical techniques of phylogenetics such as Wagner networking (Felsenstein, 1978) or "boot-strapping" analysis (Dodds, 1986).

10.3. Possible Additional Categories

Complex situations may cut across the schematic distinctions proposed above, making it even more difficult to design statistical approaches to enable the construction of a rigorous picture of microbe natural diversity.

10.3.1. *Long-lasting Epidemic Model*

It is easy to imagine cases in which the stability of clones would exceed by far the lifetime of a few years inferred for epidemic clones (Maynard Smith *et al.*, 1993), without attaining an evolutionary scale. If the stability of these clones could be counted in hundreds of years, it would have profound implications for epidemiology, which again raises the relevance of the level of resolution of genetic markers (see Section 11) in distinguishing such situations from the epidemic model proposed by Maynard Smith *et al.* (1993). With such time scales, especially when fast-evolving markers are considered, considerable genetic divergence could accumulate between clones before their genetic make-up were diluted in the common gene pool by genetic recombination.

Other possible cases, distinct from both cryptic speciation and clonal evolution, and with evolutionary consequences which may closely mimic those of either cryptic speciation or the clonal model, are considered below.

10.3.2. *Strict Homogamy*

It is possible to imagine cases in which homogamy is very strict, and only those individuals that share an identical genotype, or have very closely related genotypes, are capable of genetic exchange (Figure 8). From an evolutionary point of view, this model amounts to a particular case of cryptic speciation, in which each species is either monomorphic or extremely poorly polymorphic.

10.3.3. *Progressive Speciation*

As proposed for *Bacillus subtilis* by Roberts and Cohan (1993), it is possible that gene exchange is inhibited proportionately to the level of genetic divergence among microbial lineages (Figure 9).

The complex situations described above could be all the more difficult to distinguish from "true" clonal evolution since, in the framework of this last model, various numbers of episodes of sex could interfere with the evolutionary fate of the clones (Tibayrenc *et al.*, 1990). These episodes of sex could vary in number from one species to another, and even within a given species, from one ecological cycle to another. As examples, although

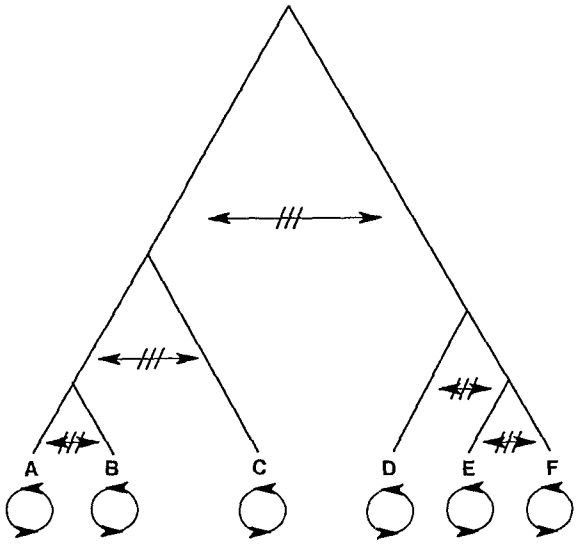

Figure 8 Homogamic model in which gene exchange is free (symbolized by circular arrows) among individuals sharing an identical or very closely similar genotype, while it is inhibited (symbolized by crossed horizontal arrows) between non-identical genotypes.

several lines of evidence suggest placing *T. brucei* in the framework of clonal evolution, it is likely that gene exchange is more active in this species than in *T. cruzi*. Moreover, the possibility that, for these two species, gene exchange is more active in cycles involving their natural, wild hosts than in those involving human hosts, cannot be ruled out.

Some of the means by which one can attempt to distinguish between the various kinds of population structure described above are summarized in Table 10. In this context, the following points are important.

1. A negative test result is never definitive confirmation that the null hypothesis is verified.
2. The selection of a marker with an appropriate resolving power (see Section 11) is crucial for identifying genotypes, the repetition of which results from epidemic propagation (see Section 9.3.4.(c).(ii)).
3. There is at present no precise limit to what can be called "large" genetic distances, although extreme values such as those estimated for *T. cruzi* do suggest phylogenetic divergence rather than individual diversity (see Section 9.3.4.(c).(i)).
4. If homogamy involves closely related, rather than identical, genotypes,

Table 10 Differential diagnosis between various kinds of population structure.

	Linkage disequilibrium[a]			Test g[b]	"Large" genetic distances	Phylogenetic analysis[c]
	1	2	3			
Lack of intraspecific phylogenetic subdivisions (non-subdivided species)						
Panmictic model	−	−	NA	−	−	−
Epidemic model	+	−	NA	−	−	−
Long-lasting epidemic model[d]	+	−	NA	−	−	−
Intraspecific phylogenetic subdivisions (subdivided species)						
Clonal evolution	+	+	+	+	+	+
Cryptic speciation	+	+	−	+	+	+
Strict homogamy[e]	+	+	+/−[f]	+	+	+
Progressive speciation	+	+	−	+	+	+

[a] 1, Whole sample, with the individual taken as unit. 2, Whole sample, with the genotype taken as unit (the marker used must have a level of resolution adequate to ensure that identical genotypes are the result of recent epidemic spread; the level of resolution required for the epidemic model (Maynard Smith *et al.*, 1993) and the long-lasting epidemic model are different: see Section 10.3.1). 3, Subdivisions of species in which panmixia could be suspected (for example, each of the two main clusters in Figure 3). NA, Not applicable.

[b] See Table 2: correlation between different kinds of genetic marker (e.g. isoenzymes and RAPD).

[c] Use of phylogenetic techniques such as Wagner networking (Felsenstein, 1978) or "boot-strapping" analysis (Dodds, 1986).

[d] See Section 10.3.1.

[e] See Section 10.3.2.

[f] Depends on the level of homogamy (see Section 10.3.2).

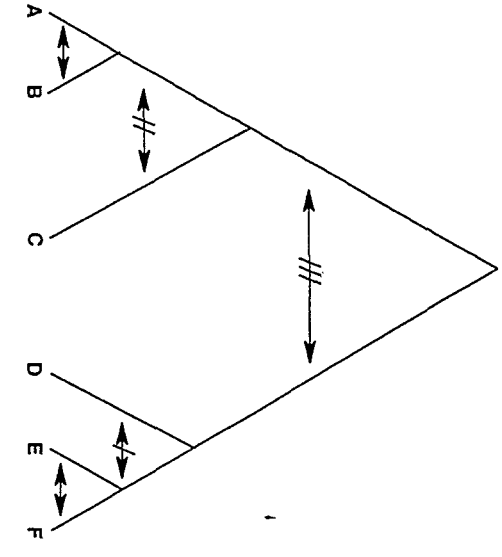

Figure 9 Model of progressive speciation, in which gene exchange is inhibited proportionally to the phylogenetic divergence observed between microbial lineages. Gene exchange is symbolized by horizontal arrows, while the strength of its inhibition is indicated by the number of slanting bars. Gene exchange is frequent between either A and B or E and F, and becomes scarcer (one bar) between D on one hand and E or F on the other, and between C and A or B (two bars). Inhibition is maximal between the two main clusters, A to C on one hand and D to F on the other.

the results could mimic either cryptic or progressive speciation (see above). If homogamy is very strict and involves only identical genotypes (at least with the marker used), the results will be identical to those obtained with the clonal evolution model.

11. THE RELEVANCE OF TIME AND SPACE FOR POPULATION GENETICS AND STRAIN TYPING OF MICROORGANISMS

From the points developed in Sections 9.3 and 10, it is apparent that the use of a given marker and a given statistical method must be determined by the question under study: it is artificial to speak about "good" and "bad" markers. A "good" technique is simply one which appears more suitable to answer a given problem. These problems, often not clearly distinguished from each other in studies dealing with microbe strain characterization, can be conveniently classified as follows, according to the time and space scale considered (Table 1).

11.1. Four Different Levels of Analysis

11.1.1. *Experimental Level*

The need to characterize stocks may be limited to a laboratory experiment dealing with immunology, virulence in mice, etc., and involving several stocks of a given microbe. The only requirement here is to select a marker with sufficient resolution to discriminate between the stocks under study. Population genetic analyses are obviously not required in this case.

11.1.2. *Nosocomial Level*

The purpose may be to follow the spread of a given microbe genotype in a very limited time and space scale: for example, infection of a resuscitation unit by a methicillin-resistant *Staphylococcus aureus* stock over a span of time of a few weeks or, at most, a few months. The level of resolution of the marker used here is critical in order to be able to make a reasonable assumption that identical genotypes are truly the result of very short-term nosocomial propagation. On the other hand, the probability that gene exchange will have interfered with the genotype distribution in such a short space of time is limited.

11.1.3. *Epidemic Level*

The goal here is to survey the spread of given strains over extended areas and over a scale of time of several years. It becomes critical to estimate the possible impact of genetic recombination on the genotype distribution observed (population genetic analysis). Moreover, the molecular "clock" of the marker used should also be carefully considered: if it is not fast enough, the "identical" genotypes may actually be collections of closely related genotypes (lack of resolution). If too fast, the "identical" genotypes could have been generated by the phenomena of reversion. A central epidemiological question that addresses this scale of resolution is the research into animal reservoirs of human diseases.

11.1.4. *Evolutionary Level*

Rather than following the spread of given genotypes (either at the nosocomial or the epidemic level), the purpose of these studies is to obtain a general picture of the population structure and evolutionary patterns within a given species. Population genetic analyses are hence definitely required to estimate the impact of gene exchange on population structure. When distinguishing a clonal structure from an epidemic one, it is critical to choose a marker with a level of resolution which makes it possible safely to identify epidemic clones. When drawing an overall picture of population structure, slowly evolving markers may be more convenient.

A major problem is that the levels of resolution (molecular "clocks") of the various markers available are far from finely calibrated, and it is difficult to select the appropriate marker ideally to fit the four situations described above. A few preliminary remarks can nevertheless be made.

11.2. Two Different Categories of Genetic Marker

Genetic markers (see Section 3.1) can be classified into "generalist" and "specialist".

11.2.1. *Generalist Markers*

This first category refers to those markers, such as isoenzymes or RAPD, which are usable whatever organism is under study. They can therefore be conveniently used to make direct comparisons about genetic diversity and population structure between different microorganisms. Comparisons dealing with the levels of genetic divergence observed should be tentative if made between very different kinds of organism (mammals and bacteria, for example), but can safely be attempted between more closely related organisms. As an example, the levels of intraspecific phylogenetic divergence appear to be far higher in *T. cruzi* (see Tibayrenc *et al.*, 1986) than in *T. brucei* (see Mathieu-Daudé and Tibayrenc, 1994). This comparison is all the more informative since it was made in the same laboratory with the aid of similar isoenzyme techniques.

11.2.2. *Specialist Markers*

This second category refers to those markers whose use is limited to a given species or a given group of species. RFLP or PFGE studies involving the hybridization of a specific sequence as probe are limited to the category of organisms that share this specific sequence (for example, RFLPs relying on the IS6110 probe in *Mycobacterium tuberculosis*; see Van Embden *et al.*, 1993). Kinetoplast DNA RFLP studies (schizodeme analysis) and PFGE karyotyping are also limited to the study of specific kinds of microorganisms. It is impossible to use specialist markers to make direct comparisons between the organisms to which they are specific, on the one hand, and other microorganisms on the other hand. Moreover, it is impossible to guess in advance the level of resolution of these markers: comparisons with the results obtained by generalist markers for the same

collection of strains will be necessary in order to obtain some idea of the level of their molecular "clock".

11.3. Setting the Molecular "Clock"

MLEE is most probably too slow a marker to fit the situations described in Sections 11.1.2 and 11.1.3 (nosocomial and epidemic levels; in particular, research into animal reservoirs of human diseases), even using a fair number of loci. The isoenzyme profiles in most cases reveal a set of related genotypes rather than a unique genotype, and hence can provide only a limited presumption that given genotypes are identical. On the other hand, MLEE is a convenient marker for estimating phylogenetic divergences among the main subdivisions of a species. Even slower markers (ribosomal ribonucleic acid gene sequences, for example) could be required to explore better the phylogenetic divergence within species such as *T. cruzi*, the diversity of which is considerable (Tibayrenc and Ayala, 1988). RAPDs seem to have a finer resolution than isoenzymes, and will be used to complement them and to refine the picture of genetic diversity within each subdivision of a species.

To address situations involving nosocomial and epidemic levels, it is possible that even RAPD will not be discriminative enough. RFLP "fingerprints" (Van Embden *et al.*, 1993) could provide a finer level of resolution but, again, this will have to be verified by direct comparison with the results obtained using generalist markers with the same sets of stocks.

Comparisons between the same stocks using various markers will make it possible to rank, in a relative manner, their speed of evolution. Nevertheless, to estimate the stability of these markers for the time scales relevant to the nosocomial and epidemic levels, long-term experiments with microbe cultures extended over several years are required. This will permit the better testing of the epidemic model proposed by Maynard Smith *et al.* (1993). As an example, it could be verified whether the additional "fingerprint" variability recorded by Hide *et al.* (1994) within *T. brucei* zymodemes (see Section 9.3.3.(c)) could be generated by mutation in a few years in laboratory-cloned stocks. If this is not verified, and the "fingerprint" patterns remain stable over several years, the "fingerprint" genotype, rather than the zymodeme, should be used as the unit in linkage disequilibrium tests to distinguish epidemic propagation from clonality (Maynard Smith *et al.*, 1993).

It has not yet been fully ascertained whether the molecular "clock" of the various markers is the same in the laboratory and in nature. Nevertheless, this approach represents the only means to estimate experimentally the speed of evolution of the markers to be compared. Such long-term

experiments involving several major parasite species (with stocks whose cloning has been verified under the microscope) and different kinds of genetic marker are presently in hand in my laboratory.

12. POPULATION GENETICS AND THE NOTION OF SPECIES IN MICROORGANISMS

12.1. Non-clonal Microorganisms

In microorganisms that are proved to be panmictic, as proposed for *N. gonorrhoeae* by O'Rourke and Stevens (1993), there should be no problem in applying the biological concept of species (Mayr, 1940), if the specialists involved find it desirable.

12.2. Basically Clonal Microorganisms

On the contrary, in those organisms in which clonal evolution is suspected or well established and, more generally, in all cases in which gene exchange is not frequent enough to generate panmixia, the biological concept of a species is hardly applicable. The definition and boundaries of species hence become a matter of convenience. This problem has been discussed by Tibayrenc (1993). Asexual species (agamospecies) can be defined on the basis of biological/medical criteria, phylogenetic divergence (cf. the concept of genospecies in bacteriology; Grimont, 1988), or, better, a combination of these two criteria. As an example, the existence of four cryptic species within *G. duodenalis* has been proposed on the basis of phylogenetic divergence only (Andrews *et al.*, 1989), while speciation within *E. histolytica* was inferred from both phylogenetic divergence and the particular virulence of some genotypes (Blanc, 1992).

Since the number of different clones in basically asexual organisms is virtually unlimited, caution should be used in describing new species from phylogenetic inferences. At present, there is a growing risk of "species inflation" in the genus *Leishmania*. New Linnean binary names should obviously be limited to major phylogenetic lineages, if possible linked to specific biological and/or medical properties.

It has been proposed to revive the biological concept of species (Mayr, 1940) in the case of those bacteria such as *Escherichia coli*, in which gene exchange noticeably interferes with clonal evolution (Dykhuizen and Green, 1991). Nevertheless "sex" in *E. coli* definitely has a different

evolutionary status than it does in "real" sexual species. Moreover, it is probable that there is no clear-cut limit to gene exchange among bacterial populations. So it is probable that the proposal by Dykhuizen and Green (1991) will be difficult to apply routinely to bacterial systematics.

12.3. Clonets and Major Clones

While it is not desirable to multiply Linnean binary names, some species of microorganisms as currently defined could be too broad for applied studies such as clinical research or immunology. As an example, it is well known that some *L. infantum* genotypes are more viscerotropic than others, at least in immunocompetent patients. The natural clone defined in the framework of the clonal model (Tibayrenc *et al.*, 1990) could provide a useful additional taxonomic unit for such applied studies. Since the level of resolution of a given marker is obligatorily limited, the new term "clonet" has been coined (Tibayrenc and Ayala, 1991) to refer to all isolates of a clonal species that share the same profile for a given set of genetic markers. Widespread clonets ("major clones"; Tibayrenc and Brenière, 1988) should be the subject of priority studies dealing with virulence, resistance to drugs, host specificity, etc.

13. CONCLUDING REMARKS

I have tried to show that, even for the rather simple purpose of epidemiological tracking, the need to analyse data in population genetic terms is crucial, and increases proportionally with the scale of time and space considered. The risk that gene exchange interferes with genotype stability is especially great in those genera having a well-known sexual cycle (*Babesia, Plasmodium* and *Toxoplasma*). With such organisms, mere empirical "typing" is especially questionable. However, the risk of sex interfering with the stability of the genotype cannot be ruled out for any microbe, without careful population genetic analysis.

Time scale is also a central problem in selecting the appropriate marker with a level of resolution which will make it possible to answer the appropriate question. Additional studies are required to calibrate properly the molecular "clocks" of the numerous markers available.

On the other hand, progress in theoretical interpretation of genetic data will most probably depend upon the possibilities of building bridges between previously separate lines of research studying different kinds of microbe. The development of this comparative population genetics of

microorganisms (Tibayrenc and Ayala, 1991), summarized in Table 1, will make it possible: (i) to perfect strain typing so that it becomes a really routine and reliable epidemiological tool; (ii) to improve and codify microbial taxonomy; (iii) to address the linked central questions which are still unanswered — what are the evolutionary forces that drive microbe genetic diversity? and what is the impact of this genetic diversity on those biological properties, such as virulence, resistance to drugs, etc., that are of interest for applied research (either agronomical or medical)?

ACKNOWLEDGEMENTS

Thanks are due to all members of my team who generated beautiful data and greatly contributed to theoretical progress through lively and enthusiastic discussions.

I also thank the eminent external collaborators who contributed notably to the development of this research: M.L. Cariou and M. Solignac (CNRS, Gif/Yvette, France), D. Le Ray (Tropical Medicine Institute, Antwerp, Belgium), F. Kjellberg and F. Renaud (CNRS, Montpellier, France) and F.J. Ayala (UC Irvine, California).

REFERENCES

Alves, A.M.B., Tanuri, A., de Almeida, D.F. and von Krüger, W.M.A. (1993). Reversible changes in the isoenzyme electrophoretic mobility pattern and infectivity in clones of *Trypanosoma cruzi*. *Experimental Parasitology* **77**, 246–253.

Andrade, S.G. (1985). Morphological and behavioural characterization of *Trypanosoma cruzi* strains. *Revista de Sociedade Brasileira de Medicina Tropical* **18**, 39–46.

Andrade, V., Brokskyn, C. and Andrade, S.G. (1983). Correlation between isoenzyme patterns and biological behaviour of different strains of *Trypanosoma cruzi*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **77**, 796–799.

Andrade, S.G., Magalhaes, J.B. and Pontes, A.L. (1985). Evaluation of chemotherapy with benznidazole and nifurtimox in mice infected with *Trypanosoma cruzi* strains of different types. *Bulletin of the World Health Organization* **63**, 721–726.

Andrews, R.H., Adams, M., Boreham, P.F.L., Mayrhofer, G. and Meloni, B.P. (1989). *Giardia intestinalis*: electrophoretic evidence for a species complex. *International Journal for Parasitology* **19**, 183–190.

Avise, J.C. (1994). *Molecular Markers, Natural History and Evolution*. London: Chapman & Hall.

Ayala, F.J. and Kiger, J.A. (1984). *Modern Genetics*, 2nd edn. Menlo Park, CA: Benjamin Cummings.

Babiker, H.A., Creasey, A.M., Fenton, B., Bayoumi, R.A.L., Arnot, D.E. and Walliker, D. (1991). Genetic diversity of *Plasmodium falciparum* in a village in Eastern Sudan. 1. Diversity of enzymes, 2D-PAGE proteins and antigens. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **85,** 572–577.

Babiker, H.A., Ranford-Cartwright, L.C., Currie, D., Charlwood, J.D., Billingsley, P., Teuscher, T. and Walliker, D. (1994). Random mating in a natural population of the malaria parasite *Plasmodium falciparum*. *Parasitology* **109,** 413–421.

Baker, J.R. (1989). Sexual processes in parasitic protozoa. *International Journal for Parasitology* **19,** 465–472.

Bañuls, A.L. (1993). Analyse génétique d'isolats naturels sud-américains de *Leishmania* montrant de possibles phénomènes de recombinaison. Ph.D. Dissertation, University of Montpellier, France.

Bastien, P., Blaineau, C. and Pagès, M. (1992). *Leishmania*: sex, lies and karyotypes. *Parasitology Today* **8,** 174–177.

Ben Abderrazak, S. (1993). Variabilité génétique des populations de *Plasmodium falciparum*. Ph. D. Thesis, University of Montpellier, France.

Ben Abderrazak, S., Guerrini, F., Mathieu-Daudé, F., Truc, P., Neubauer, K., Lewicka, K., Barnabé, C. and Tibayrenc, M. (1993). Isozyme electrophoresis for parasite characterization. *Methods in Molecular Biology* **21,** 361–382.

Blanc, D.S. (1992). Determination of taxonomic status of pathogenic and non-pathogenic *Entamoeba histolytica* zymodemes using isoenzyme analysis. *Journal of Protozoology* **39,** 471–479.

Brandt, M.E., Bragg, S.L. and Pinner, R.W. (1993). Multilocus enzyme typing of *Cryptococcus neoformans*. *Journal of Clinical Microbiology* **31,** 2819–2823.

Brenière, S.F., Tibayrenc, M., Antezana, G., Pabon, J., Carrasco, R., Selaès, H. and Desjeux, P. (1985). Résultats préliminaires en faveur d'une relation faible ou inexistante entres les formes cliniques de la maladie de Chagas et les souches isoenzymatiques de *Trypanosoma cruzi*. *Comptes Rendus de l'Académie des Sciences, Paris* **300,** 555–558.

Brumpt, E. (1925). Etude sommaire de l'*Entamoeba dispar* n. sp., amibe à kystes quadrinucléés parasite de l'homme. *Bulletin de l'Académie de Médecine* **94,** 942–952.

Cariou, M.L. and Pernin, P. (1987). First evidence for diploidy and genetic recombination in free-living amoebae of the genus *Naegleria* on the basis of electrophoretic variation. *Genetics* **115,** 265–270.

Carter, R. and McGregor, I.A. (1973). Enzyme variation in *Plasmodium falciparum* in The Gambia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **67,** 830–837.

Carter, R. and Voller, A. (1975). The distribution of enzyme variation in populations of *Plasmodium falciparum* in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **69,** 371–376.

Caugant, D.A. and Sandven, P. (1993). Epidemiological analysis of *Candida albicans* strains by multilocus enzyme electrophoresis. *Journal of Clinical Microbiology* **31,** 215–220.

Caugant, D.A., Froholm O., Bovre, K., Holten, E., Frasch, C.E., Mocca, L.F., Zollinger, W.D. and Selander, R.K. (1986). Intercontinental spread of a genetically distinctive complex of clones of *Neisseria meningitidis* causing epidemic disease. *Proceedings of the National Academy of Sciences of the USA* **83,** 4927–4931.

Caugant, D.A., Aleksic, S., Mollaret, H.H., Selander, R.K. and Kapperud, G. (1989). Clonal diversity and relationships among strains of *Yersinia enterocolitica*. *Journal of Clinical Microbiology* **27,** 2678–2683.

Cibulskis, R.E. (1988). Origins and organization of genetic diversity in natural populations of *Trypanosoma brucei*. *Parasitology* **96,** 303–322.

Cibulskis, R.E. (1992). Genetic variation in *Trypanosoma brucei* and the epidemiology of sleeping sickness in the Lambwe Valley, Kenya. *Parasitology* **104,** 99–109.

Clark, C.G. and Diamond, L.S. (1993). *Entamoeba histolytica* – an explanation for the reported conversion of nonpathogenic amebae to the pathogenic form. *Experimental Parasitology* **77,** 456–460.

Clark, C.G., Cunnick, C.C. and Diamond, L.S. (1992). *Entamoeba histolytica* – is conversion of nonpathogenic amebae to the pathogenic form a real phenomenon? *Experimental Parasitology* **74,** 307–314.

Conway, D.J. and McBride, J.S. (1991). Population genetics of *Plasmodium falciparum* within a malaria hyperendemic area. *Parasitology* **103,** 7–16.

Currie, B.P., Freundlich, L.F. and Casadevall, A. (1994). Restriction fragment length polymorphism analysis of *Cryptococcus neoformans* isolates from environmental (pigeon excreta) and clinical sources in New York City. *Journal of Clinical Microbiology* **32,** 1188–1192.

Darce, M., Moran, J., Palacios, X., Belli, A., Gomez-Urcuyo, F., Zamora, D., Valle, S., Gantier, J.C., Momen, H. and Grimaldi, G., jr (1991). Etiology of human cutaneous leishmaniasis in Nicaragua. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **85,** 58–59.

Dardé, M.L., Bouteille, B. and Pestre-Alexandre, M. (1988). Isoenzymic characterization of seven strains of *Toxoplasma gondii* by isoelectrofocusing in polyacrylamide gels. *American Journal of Tropical Medicine and Hygiene* **39,** 551–558.

Dardé, M.L., Bouteille, B. and Pestre-Alexandre, M. (1990). Caractérisation isoenzymatique de 15 souches clonées de *Toxoplasma gondii*. *Bulletin de la Société Française de Parasitologie* **8,** 238 (abstract).

Day, K.P., Koella, J.C., Nee, S., Gupta, S. and Read, A.F. (1992). Population genetics and dynamics of *Plasmodium falciparum*: an ecological view. *Parasitology* **104,** 335–352.

Desjeux, P. and Dédet, J.P. (1989). Isoenzyme characterization of 112 *Leishmania* isolates from French Guiana. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **83,** 610–612.

Diamond, L.S. and Clark, C.G. (1993). A redescription of *Entamoeba histolytica* Schaudinn, 1903 (emended Walker, 1911) separating it from *Entamoeba dispar* Brumpt, 1925. *Journal of Eukaryotic Microbiology* **40,** 340–344.

Dodds, K.G. (1986). *Resampling Methods in Genetics and the Effect of Family Structure in Genetic Data.* Institute of Statistics Mimeograph Series No. 1684T. Raleigh, NC: North Carolina State University.

Dujardin, J.C., Llanos-Cuentas, A., Caceres, A., Arana, M., Dujardin, J.P., Guerrini, F., Gomez, J., Arroyo, J., De Doncker, S., Jacquet, D., Hamers, R., Guerra, H., Le Ray, D. and Arevalo, J. (1993). Molecular karyotype variation in *Leishmania (Viannia) peruviana*: indication of geographical populations in Peru distributed along a north–south cline. *Annals of Tropical Medicine and Parasitology* **87,** 335–347.

Dye, C. (1991). Population genetics of nonclonal, nonrandomly mating malaria parasites. *Parasitology Today* **7,** 236–240.

Dye, C., Davies, C.R. and Lines, J.D. (1990). When are parasites clonal? *Nature* **348,** 120.

Dykhuizen, D.E. and Green, L. (1991). Recombination in *Escherichia coli* and the definition of biological species. *Journal of Bacteriology* **173,** 7257–7268.

Dykhuizen, D.E., Polin, D.S., Dunn, J.J., Wilske, B., Preacmursic, V., Dattwyler R.J. and Luft, B.J. (1993). *Borrelia burgdorferi* is clonal — implications for taxonomy and vaccine development. *Proceedings of the National Academy of Sciences of the USA* **90,** 10163–10167.

Eardly, B.D., Materon, L.A., Smith, N.H., Johnson, D.A., Rumbaugh, M.D. and Selander, R.K. (1990). Genetic structure of natural populations of the nitrogen-fixing bacterium *Rhizobium meliloti. Applied and Environmental Microbiology* **56,** 187–194.

Evans, D.A., Kennedy, W.P.K., Elbihari, S., Chapman, C.J., Smith, V. and Peters, W. (1987) Hybrid formation within the genus *Leishmania. Parassitologia* **29,** 165–173.

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27,** 401–410.

Garfinkel, L.I., Giladi, M., Huber, M., Gitler, C., Mirelman, D., Revel, M. and Rozenblatt, S. (1989). DNA probes specific for *Entamoeba histolytica* possessing pathogenic and nonpathogenic zymodemes. *Infection and Immunity* **57,** 926–931.

Gibson, W. (1990). Trypanosome diversity in Lambwe Valley, Kenya — sex or selection? *Parasitology Today* **6,** 342–343.

Gibson, W. and Garside, L. (1991). Genetic exchange in *Trypanosoma brucei brucei* — variable chromosomal location of housekeeping genes in different trypanosome stocks. *Molecular and Biochemical Parasitology* **45,** 77–90.

Gibson, W.C., Marshall, T.F. de C. and Godfrey, D.G. (1980). Numerical analysis of enzyme polymorphism: a new approach to the epidemiology and taxonomy of trypanosomes. *Advances in Parasitology* **18,** 175–246.

Grimont, P.A.D. (1988). Use of DNA reassociation in bacterial classification. *Canadian Journal of Microbiology* **34,** 541–547.

Guerrini, F. (1993). Génétique des populations et phylogénie des Leishmania du Nouveau-Monde. Ph.D. dissertation, University of Montpellier, France.

Gupta, S. and Day, K. (1994). A strain theory of malaria transmission. *Parasitology Today* **10,** 176–181.

Gupta, S., Hill, A.V.S., Kwiatkowski, D., Greenwood, A.M., Greenwood, B.M. and Day, K.P. (1994a). Parasite virulence and disease patterns in *Plasmodium falciparum* malaria. *Proceedings of the National Academy of Sciences of the USA* **91,** 3715–3719.

Gupta, S., Trenholme, K., Anderson, R.M. and Day, K.P. (1994b). Antigenic diversity and the transmission dynamics of *Plasmodium falciparum. Science* **263,** 961–963.

Hartl, D. (1992). Population genetics of microbial organisms. *Current Biology* **2,** 937–942.

Hartl, D. and Clark, A.G. (1989). *Principles of Population Genetics*, 2nd edn, pp. 45–57. Sunderland, MA: Sinauer Associates.

Hartl, D.L. and Dykhuizen D.E. (1984). The population genetics of *Escherichia coli. Annual Review of Genetics* **18,** 31–68.

Hide, G., Tait, A., Maudlin, I. and Welburn, S.C. (1994). Epidemiological relationships of *Trypanosoma brucei* stocks from south east Uganda: evidence for different population structures in human infective and non-human infective isolates. *Parasitology* **109,** 95–111.

Hillis, D.M. and Moritz, C. (1990). *Molecular Systematics*. Sunderland, MA: Sinauer Associates.

Jaccard, P. (1908). Nouvelles recherches sur la distribution florale. *Bulletin de la Société vaudoise de Sciences naturelles* **44,** 223–270.

Jenni, L., Marti, S., Schweitzer, J., Betschart, B., Le Page, R.W.F., Wells, J.M., Tait, A., Paindavoine, P., Pays, E. and Steinert, M. (1986). Hybrid formation between African trypanosomes during cyclical transmission. *Nature* **322,** 173–175.

Kilgour, V. and Godfrey, D.G. (1973). Species-characteristic isoenzymes of two aminotransferases in trypanosomes. *Nature New Biology* **244,** 69–70.

Lanar, D.E., Levy, L.S. and Manning, J.E. (1981). Complexity and content of the DNA and RNA in *Trypanosoma cruzi. Molecular and Biochemical Parasitology* **3,** 327–341.

Lanotte, G. and Rioux, J.A. (1990). Fusion cellulaire chez les *Leishmania* (Kinetoplastida, Trypanosomatidae). *Comptes Rendus de l'Académie des Sciences, Paris* **310,** 285–288.

Laurent, J.P. (1994). Comparaison des propriétés biologiques de différents clones naturels de *Trypanosoma (Schizotrypanum) cruzi* (Chagas, 1909), agent de la maladie de Chagas. Ph.D. dissertation, University of Montpellier, France.

Lehmann, P.F., Kemker, B.J., Hsiao, C.B. and Dev, S.J. (1989). Protein and enzyme electrophoresis profiles of selected *Candida* species. *Journal of Clinical Microbiology* **27,** 2514–2521.

Lewicka, K. (1991). Etude de la variabilité génétique de *Trypanosoma (Schizotrypanum) cruzi* en Guyane française. Ph.D. dissertation, University of Montpellier, France.

Maazoun, R., Lanotte, G., Rioux, J.A., Pasteur, N., Killick-Kendrick, R. and Pratlong, F. (1981). Signification du polymorphisme enzymatique chez les leishmanies. A propos de trois souches hétérozygotes de *Leishmania infantum, Leishmania* cf. *tarentolae* et *Leishmania aethiopica. Annales de Parasitologie humaine et comparée* **56,** 467–475.

Maazoun, R., Pratlong, N., Lanotte, G. and Rioux, J.A. (1986). Le complexe *Leishmania major.* A propos de l'analyse numérique de 35 souches identifiées par la méthode enzymatique. *In*: Leishmania: *Taxonomie et Phylogénèse: Applications Éco-épidémiologiques (Colloque Internationale CNRS/INSERM, 1984)*, pp. 119–129. Montpellier: IMEEE.

Mantel, N. (1967). The detection of disease clustering and a generalised regression approach. *Cancer Research* **27,** 209–220.

Mathieu-Daudé, F. and Tibayrenc, M. (1994). Isozyme variability of *Trypanosoma brucei s.l.*: genetical, taxonomical and epidemiological significance. *Experimental Parasitology* **78,** 1–19.

Maynard Smith, J., Dowson, C.G. and Spratt, B.G. (1992). Localized sex in bacteria. *Nature* **349,** 29–31.

Maynard Smith, J., Smith, N.H., O'Rourke, M. and Spratt, B.G. (1993). How clonal are bacteria? *Proceedings of the National Academy of Sciences of the USA* **90,** 4384–4388.

Mayr, E. (1940). Speciation phenomena in birds. *American Naturalist* **74,** 249–278.

Mayrhofer, G. and Andrews, R.H. (1994). Selection of genetically distinct *Giardia* by growth *in vitro* and *in vivo*. *In*: Giardia: *from Molecules to Disease* (R.C.A.

Thompson, J.A. Reynoldson and A.J. Lymbery, eds), pp. 52–53. Wallingford, UK: CAB International.

Meloni, B.P., Lymbery, A.J. and Thompson, C.A. (1989). Characterization of *Giardia* isolates using a non-radiolabelled DNA probe, and correlation with the results of isoenzyme analysis. *American Journal of Tropical Medicine and Hygiene* **40,** 629–637.

Meyer, W., Mitchell, T.G., Freedman, E.Z. and Vilgalys, R. (1993). Hybridization probes for conventional DNA fingerprinting used as single primers in the polymerase chain reaction to distinguish strains of *Cryptococcus neoformans*. *Journal of Clinical Microbiology* **31,** 2274–2280.

Miles, M.A., Toyé, P.J., Oswald, S.C. and Godfrey, D.G. (1977). The identification by isoenzyme patterns of two distinct strain-groups of *Trypanosoma cruzi*, circulating independently in a rural area of Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **71,** 217–225.

Miles, M.A., Souza, A., Povoa, M., Shaw, J.J., Lainson, R. and Toyé, P.J. (1978). Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochtonous patients with Chagas's disease in Amazonian Brazil. *Nature* **272,** 819–821.

Miles, M.A., Povoa, M. De Souza, A.A., Lainson, R., Shaw, J.J. and Ketteridge, D.S. (1981). Chagas's disease in the Amazon Basin: II. The distribution of *Trypanosoma cruzi* zymodemes 1 and 3 in Pará State, north Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **75,** 667–674.

Milkman, R. (1973). Electrophoretic variation in *Escherichia coli* from natural sources. *Science* **182,** 1024–1026.

Milkman, R. and Bridges, M.M. (1993). Molecular evolution of the *Escherichia coli* chromosome. 4. Sequence comparisons. *Genetics* **133,** 455–468.

Mirelman, D. (1987): Effect of culture conditions and bacterial associates on the zymodemes of *Entamoeba histolytica*. *Parasitology Today* **3,** 37–40.

Mirelman, D. and Burchard, G.D. (1987). *Entamoeba histolytica*: is it necessary to characterize pathogenic strains? Reply (contra). *Parasitology Today* **3,** 353.

Morel, C.M., Chiari, E., Plessmann Camargo, E., Mattei, D.M., Romanha, A.J. and Simpson, L. (1980). Strains and clones of *Trypanosoma cruzi* can be characterized by pattern of restriction endonuclease products of kinetoplast DNA minicircles. *Proceedings of the National Academy of Sciences of the USA* **77,** 6810–6814.

Moreno, G., Rioux, J.A, Lanotte, G., Pratlong, F. and Serres, E. (1986). Le complexe *Leishmania donovani s.l.* Analyse enzymatique et traitement numérique. Individualisation du complexe *Leishmani infantum*. Corollaires biogéographiques et phylétiques. A propos de 146 souches originaires de l'Ancien et du Nouveau Monde. In: Leishmania. *Taxonomie et phylogénèse. Applications éco-épidémiologiques (Colloque Internationale CNRS/INSERM, 1984)*, pp. 105–117. Montpellier: IMEEE.

Musser, J.M., Granoff, D.M., Pattison, P.E. and Selander, R.K. (1985). A population genetic framework for the study of invasive diseases caused by serotype b strains of *Haemophilus influenzae*. *Proceedings of the National Academy of Sciences of the USA* **82,** 5078–5082.

Nei, M. (1972). Genetic distance between populations. *American Naturalist* **106,** 283–292.

Ochman, H. and Selander, R.K. (1984). Evidence for clonal population structure in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the USA* **81,** 198–201.

Odds, F.C., Brawner, D.L., Staudinger, J., Magee, P.T. and Soll, D.R. (1992).

Typing of *Candida albicans* strains. *Journal of Medical and Veterinary Mycology* **30,** 87–94.

O'Rourke, M. and Stevens, E. (1993). Genetic structure of *Neisseria gonorrhoeae* populations — a non-clonal pathogen. *Journal of General Microbiology* **139,** 2603–2611.

Ørskov, F. and Ørskov, I. (1983). Summary of a workshop on the clone concept in the epidemiology, taxonomy, and evolution of the Enterobacteriaceae and other bacteria. *Journal of Infectious Diseases* **148,** 346–357.

Pasteur, N., Pasteur, G., Bonhomme, F., Catalan, J. and Britton-Davidian, J. (1987). *Manuel Technique de Génétique par Électrophorèse des Protéines*. Technique et Documentation Lavoisier, Paris.

Peters, W. and Killick-Kendrick, R. (1987). *The Leishmaniases in Biology and Medicine*. London: Academic Press.

Pratlong, F., Lanotte, G., Ashford, R.W. and Rioux, J.A. (1986). Le complexe *Leishmania tropica*. A propos de l'analyse numérique de 29 souches identifiées par la méthode enzymatique. In: Leishmania: *Taxonomie et Phylogénèse: Applications Éco-épidémiologiques (Colloque Internationale CNRS/INSERM, 1984)*, pp. 129–139. Montpellier: IMEE.

Proctor, E.M., Wong, Q., Yang, J. and Keystone, J.S. (1987). The electrophoretic isoenzyme patterns of strains of *Entamoeba histolytica* isolated in two major cities in Canada. *American Journal of Tropical Medicine and Hygiene* **37,** 296–301.

Pujol, C., Reynes, J., Renaud, F., Raymond, M., Tibayrenc, M., Ayala, F.J., Janbon, F., Mallié, M. and Bastide, J.M. (1993). The yeast *Candida albicans* has a clonal mode of reproduction in a population of infected HIV+ patients. *Proceedings of the National Academy of Sciences of the USA* **90,** 9456–9459.

Ranford-Cartwright, L., Balfe, P., Carter, R. and Walliker, D. (1993). Frequency of cross-fertilization in the human malaria parasite *Plasmodium falciparum*. *Parasitology* **107,** 11–18.

Ready, P.D. and Miles, M.A. (1980). Delimitation of *Trypanosoma cruzi* zymodemes by numerical taxonomy. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **74,** 238–242.

Revollo, S. (1995). Impact de l'évolution clonale de *Trypanosoma cruzi*, agent de la maladie de Chagas, sur certaines propriétés biologiques médicalement importantes du parasite. Ph.D. dissertation, University of Montpellier, France.

Richardson, B.J., Baverstock, P.R. and Adams, M. (1986). *Allozyme Electrophoresis. A Handbook for Animal Systematics and Population Studies*. New York: Academic Press.

Roberts, M.S. and Cohan, F.M. (1993). The effect of DNA sequence divergence on sexual isolation in *Bacillus*. *Genetics* **134,** 401–408.

Romanha, A.J., Da Silva Pereira, A.A., Chiari, E. and Kilgour, V. (1979). Isoenzyme patterns of cultured *Trypanosoma cruzi:* changes after prolonged subculture. *Comparative Biochemistry and Physiology* **62B,** 139–142.

Safrin, R.E., Lancaster, L.A., Davis, C.E. and Braude, A.I. (1986). Differentiation of *Cryptococcus neoformans* serotypes by isoenzyme electrophoresis. *American Journal of Clinical Pathology* **86,** 204–208.

Sanchez, G., Wallace, A., Muñoz, S., Venegas, J. and Solari, A. (1993). Characterization of *Trypanosoma cruzi* populations by several molecular markers supports a clonal mode of reproduction. *Biological Research* **26,** 167–176.

Sargeaunt, P.G. (1985). Zymodemes expressing possible genetic exchange in

Entamoeba histolytica. Transactions of the Royal Society of Tropical Medicine and Hygiene **79,** 86–89.

Sargeaunt, P.G. (1987). *Entamoeba histolytica*: is it necessary to characterize pathogenic strains? Reply (pro). *Parasitology Today* **3,** 353.

Sargeaunt, P.G. and Williams, J.E. (1979). Electrophoretic isoenzyme patterns of the pathogenic and non-pathogenic intestinal amoebae of man. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **73,** 225–227.

Sargeaunt, P.G., Williams, J.E., Jackson, T.F.H.G. and Simjee, A.E. (1982). A zymodeme study of *Entamoeba histolytica* in a group of South African schoolchildren. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **76,** 401–402.

Sargeaunt, P.G., Jackson, T.F.H.G., Wiffen, S.R. and Bhojnani, R. (1988). Biological evidence of genetic exchange in *Entamoeba histolytica. Transactions of the Royal Society of Tropical Medicine and Hygiene* **82,** 862–867.

Selander, R.K. and Levin, B.R. (1980). Genetic diversity and structure in *Escherichia coli* populations. *Science* **210,** 245–247.

Selander, R.K., McKinney, R.M., Whittam, T.S., Bibb, W.F., Brenner, D.J., Nolte, F.S. and Pattison, P.E. (1985). Genetic structure of populations of *Legionella pneumophila. Journal of Bacteriology* **163,** 1021–1037.

Selander, R.K., Musser, J.M., Caugant, D.A., Gilmour, M.N. and Whittam, T.S. (1987). Population genetics of pathogenic bacteria. *Microbial Pathogenesis* **3,** 1–7.

Sibley, L.D. and Boothroyd, J.C. (1992). Virulent strains of *Toxoplasma gondii* comprise a single clonal lineage. *Nature* **359,** 82–85.

Soll, D.R., Galask, R., Schmid, J., Hanna, C., Mac, K. and Morrow, B. (1991). Genetic dissimilarity of commensal strains of *Candida* spp. carried in different anatomical locations of the same healthy women. *Journal of Clinical Microbiology* **29,** 1702–1710.

Souza, V., Nguyen, T.T., Hudson, R.R., Piñero, D. and Lenski, R.E. (1992). Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: evidence for sex? *Proceedings of the National Academy of Sciences of the USA* **89,** 8389–8393.

Stevens, D.A., Odds, F.C. and Scherer, S. (1990). Application of DNA typing methods to *Candida albicans* epidemiology and correlations with phenotype. *Reviews of Infectious Diseases* **12,** 258–266.

Stevens, J.R. and Tibayrenc, M. (1995). Detection of linkage disequilibrium in *Trypanosoma brucei* isolated from tsetse flies and characterized by RAPD analysis and isoenzymes. *Parasitology* **110,** 181–186.

Tachibana, H., Kobayashi, S., Paz, K.C., Aca, I.S., Tateno, S. and Ihara, S. (1992). Analysis of pathogenicity by restriction-endonuclease digestion of amplified genomic DNA of *Entamoeba histolytica* isolated in Pernambuco, Brazil. *Parasitology Research* **78,** 433–436.

Tait, A. (1980). Evidence for diploidy and mating in trypanosomes. *Nature* **237,** 536–538.

Tibayrenc, M. (1993). *Entamoeba, Giardia* and *Toxoplasma:* clones or cryptic species? *Parasitology Today* **9,** 102–105.

Tibayrenc, M. (1994a). Antigenic diversity and the transmission dynamics of *Plasmodium falciparum:* the clonality/sexuality debate revisited. *Parasitology Today* **10,** 456–457.

Tibayrenc, M. (1994b). Population genetics and systematics — how many species of *Giardia* are there? *In*: Giardia: *from Molecules to Disease* (R.C.A. Thompson, J.A. Reynoldson and A.J. Lymbery, eds), pp. 41–48. Wallingford, UK: CAB International.

Tibayrenc, M. (1995). Population genetics and strain typing of microorganisms: how to detect departures from panmixia without individualizing alleles and loci. *Comptes Rendus de l'Académie des Sciences, Paris* **318,** 135–139.

Tibayrenc, M. and Ayala, F.J. (1988). Isozyme variability of *Trypanosoma cruzi*, the agent of Chagas' disease: genetical, taxonomical and epidemiological significance. *Evolution* **42,** 277–292.

Tibayrenc, M. and Ayala, F.J. (1991). Towards a population genetics of microorganisms: the clonal theory of parasitic protozoa. *Parasitology Today* **7,** 228–232.

Tibayrenc, M. and Brenière, S.F. (1988). *Trypanosoma cruzi:* major clones rather than principal zymodemes. *Memorias do Instituto Oswaldo Cruz* **83,** 249–255.

Tibayrenc, M. and Desjeux, P. (1983). The presence in Bolivia of two distinct zymodemes of *Trypanosoma cruzi*, circulating sympatrically in a domestic transmission cycle. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **77,** 73–75.

Tibayrenc, M., Cariou, M.L. and Solignac, M. (1981a). Interprétation génétique des zymogrammes de flagellés des genres *Trypanosoma* et *Leishmania. Comptes Rendus de l'Académie des Sciences, Paris* **292,** 623–625.

Tibayrenc, M., Cariou, M.L., Solignac, M. and Carlier, Y. (1981b). Arguments génétiques contre l'existence d'une sexualité actuelle chez *Trypanosoma cruzi*; implications taxonomiques. *Comptes Rendus de l'Académie des Sciences, Paris* **293,** 207–209.

Tibayrenc, M., Echalar, L., Dujardin, J.P., Poch, O. and Desjeux, P. (1984a). The microdistribution of isoenzymic strains of *Trypanosoma cruzi* in southern Bolivia. New isoenzyme profiles and further arguments against Mendelian sexuality. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **78,** 519–525.

Tibayrenc, M., Solignac, M., Cariou, M.L., Le Ray, D. and Desjeux, P. (1984b). Les souches isoenzymatiques de *Trypanosoma cruzi*: origine récente ou ancienne, homogène ou hétérogène? *Comptes Rendus de l'Académie des Sciences, Paris* **299,** 195–198.

Tibayrenc, M., Cariou, M.L., Solignac, M., Dédet, J.P., Poch, O. and Desjeux, P. (1985). New electrophoretic evidence of genetic variation and diploidy in *Trypanosoma cruzi*, the causative agent of Chagas' disease. *Genetica* **67,** 223–230.

Tibayrenc, M., Ward, P., Moya, A. and Ayala, F.J. (1986). Natural populations of *Trypanosoma cruzi*, the agent of Chagas' disease, have a complex multiclonal structure. *Proceedings of the National Academy of Sciences of the USA* **83,** 115–119.

Tibayrenc, M., Kjellberg, F. and Ayala, F.J. (1990). A clonal theory of parasitic protozoa: the population structure of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas* and *Trypanosoma*, and its medical and taxonomical consequences. *Proceedings of the National Academy of Sciences of the USA* **87,** 2414–2418.

Tibayrenc, M., Kjellberg, F., Arnaud, J., Oury, B., Brenière, S.F., Dardé, M.L. and Ayala, F.J. (1991a). Are eucaryotic microorganisms clonal or sexual? A population genetics vantage. *Proceedings of the National Academy of Sciences of the USA* **88,** 5129–5133.

Tibayrenc, M., Kjellberg, F. and Ayala, F.J. (1991b). Clonal defence. *Nature* **350,** 385–386.

Tibayrenc, M., Neubauer, K., Barnabé, C., Guerrini, F., Sarkeski, D. and Ayala, F.J. (1993). Genetic characterization of six parasitic protozoa: parity of random-primer DNA typing and multilocus isoenzyme electrophoresis. *Proceedings of the National Academy of Sciences of the USA* **90,** 1335–1339.

Truc, P. and Tibayrenc, M. (1993). Population genetics of *Trypanosoma brucei* in Central Africa: taxonomic and epidemiological significance. *Parasitology* **106,** 137–149.

Van Embden, J.D.A., Cave, M.D., Crawford, J.T., Dale, J.W., Eisenach, K.D., Gicquel, B., Hermans, P., Martin, C., Mcadam, F.R., Shinnick, T.M. and Small, P.M. (1993). Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting — recommendations for a standardized methodology. *Journal of Clinical Microbiology* **31,** 406–409.

Wallicker, D. (1985). Characterization of *Plasmodium falciparum* of different countries. *Annales de la Société belge de Médecine tropicale* **65,** 69–77.

Walliker, D. (1991). Malaria parasites: randomly interbreeding or "clonal" populations? *Parasitology Today* **7,** 232–235.

Walliker, D., Quakyi, I.A., Wellems, T.E., McCutchan, T.F., Szarfman, A., London, W.T., Corcoran, L.M., Burkot, T.R. and Carter, R. (1987). Genetic analysis of the human malaria parasite *Plasmodium falciparum. Science* **236,** 1661–1666.

Walliker, D., Beale, G. and Luzzatto, L. (1990). When are parasites clonal? *Nature* **348,** 120.

Welsh, J. and McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acid Research* **18,** 7213–7218.

Whittam, T.S., Ochman, H. and Selander, R.K. (1983). Multilocus genetic structure in natural populations of *Escherichia coli. Proceedings of the National Academy of Sciences of the USA* **80,** 1751–1755.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acid Research* **18,** 6531–6535.

Young, J.P.W. (1989). The population genetics of bacteria. In: *Genetics of Bacterial Diversity* (D.A. Hopwood and K.F. Chater, eds), pp. 417–438. London: Academic Press.

APPENDIX: GLOSSARY

Allele: one of the different molecular forms of the same gene.

Allelic frequency: the ratio of the number of a given allele to the total number of alleles in the population under survey. For example, in a fruitfly population (fruitflies are diploid and sexual) there are two alleles, *a* and *b*, of a given gene. Among 100 individuals, there are 36 homozygous genotypes *a/a* 16 homozygous genotypes *b/b*, and 48 homozygous genotypes *a/b*. Since the organism is diploid, in 100 individuals there are 200 alleles. In a homozygous individual *a/a*, there are two alleles *a*. In a homozygous individual *b/b*, there are two alleles *b*. In a heterozygous individual *a/b*, there is one allele *a* and one allele *b*. The total number of alleles *a* is hence twice the number of *a/a* individuals (36 × 2) plus the number of *a/b* individuals, which equals 120. The frequency of this allele is 120/200 = 0.6. Similarly, the total number of alleles *b* is twice the number of *b/b* individuals (16 × 2) plus the number of *a/b* individuals = 80. The frequency of this allele is 80/200 = 0.4. Since there are only two alleles, this last result could also have been deduced from 1 minus the frequency of allele *a* = 1– 0.6 = 0.4.

Allopatry: living in different geographical locations (cf. *sympatry*).

Clone, *clonal*, *clonality*: clonal propagation is not equivalent to mitotic propagation; in population genetics, this term is used when the individuals of the progeny are genetically identical to one another and to the reproducing individual (Tibayrenc and Ayala, 1991). Apart from mitotic reproduction, this may include parthenogenesis and self-fertilization in haploid organisms. Hence a clonal population structure can be seen in animals exhibiting apparent meiosis and even mating. From a population genetics point of view, the term clonality does not refer to the mating behaviour, but rather to the population structure.

Clonet: all the isolates of a clonal species which appear to be genetically identical on the basis of a particular set of markers.

Diploid: the condition in which there are two copies of each chromosome, and hence of each gene (diploid is frequently indicated as 2*N*).

Gene: a DNA sequence coding for a given polypeptide or, more broadly, any given DNA sequence. This broad sense is adopted in the population genetic tests described in this review.

Genetic distance: various statistical quantities inferred from genetic data, estimating the genetic dissimilarities among individuals or populations. The most widely used are Nei's standard genetic distance (Nei, 1972) and the Jaccard distance (Jaccard, 1908). Although the statistics differ, most genetic distances are derived from an estimation of the percentage of band mismatch on electrophoresis gels.

Haploid: the condition in which there is only one copy of each chromosome and hence of each gene (haploid is frequently indicated as 1*N* or *N*).

Hardy–Weinberg equilibrium: see *segregation.*

Heterozygote: in a diploid organism, the two copies of a given gene in one individual have a different molecular structure: this individual harbours two different alleles of the same gene.

Homozygote: in a diploid organism, the two copies of a given gene in one individual have an identical molecular structure.

Isoenzymes: enzymes extracted from certain samples, for example various parasite stocks, are separated by electrophoresis. The gel is then subjected

to a histochemical reaction involving the specific substrate of a given enzyme and this enzyme's zone of activity is specifically stained. The same enzyme from different samples may migrate at different rates (see Figure 5). These different electrophoretic forms of the same enzyme are referred to as isoenzymes or isozymes. When given isoenzymes are governed by different alleles of the same gene, they are referred to as alloenzymes or allozymes. For detailed information about isoenzyme characterization of parasites, see Ben Abderrazak *et al.* (1993).

Karyodeme: a set of individuals or stocks that share the same pulse field gel electrophoresis (q.v.) profile.

Kinetoplast DNA: an extranuclear genome specific to the order Kinetoplastida *(Trypanosoma, Leishmania,* etc.), situated near the base of the flagellum, and representing 10–20% of the total DNA.

Linkage disequilibrium: non-random reassortment of genotypes occurring at different loci (see *recombination*).

Locus: the physical location of a given gene on the chromosome. By extension, in population genetics jargon, the gene itself.

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

Pulse field gel electrophoresis (PFGE): separation of large DNA fragments by electrophoresis using alternately pulsed, perpendicularly oriented electrical fields.

Rapdeme: a set of individuals or stocks that share the same random amplification of polymorphic DNA (RAPD; q.v.) profile for a given set of primers.

Random amplification of polymorphic DNA (RAPD): a method simultaneously proposed by Williams *et al.* (1990) and Welsh and McClelland (1990) to study genetic variability (also known as arbitrarily-primed polymerase chain reaction or AP-PCR). While, in the classical polymerase chain reaction (PCR), the primers used are known DNA sequences, the RAPD technique relies on primers whose sequence is arbitrarily determined (usually 10-mer primers are used). Under low-stringency conditions, the PCR reaction generates fragments whose polymorphism (see Figure 6) can be analysed on either ethidium bromide-stained agarose gels (Williams *et al.*, 1990), or polyacrylamide sequence gels with radiolabelling of the fragments (Welsh and McClelland, 1990).

Recombination, linkage disequilibrium: free recombination implies that the expected probability of a given multilocus genotype is the product of the observed probabilities of the single genotypes of which it is composed. 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 individuals who are both AB and Rh(+) is $0.5 \times 0.5 = 0.25$. Inhibition of recombination leads to linkage disequilibrium, or non-random association among loci (the predictions of expected probabilities for multilocus genotypes are no longer satisfied). As an example, if the observed frequency of the AB and Rh(+) individuals was greater than 0.25, it would suggest that the two loci were linked (not transmitted independently). If this frequency were 0.5, this would indicate total linkage between AB and Rh (the two characters being transmitted as a unit). (Table 2 lists specialized tests of linkage disequilibrium.)

Restriction fragment length polymorphism (RFLP): variability in the DNA of a given species revealed by the use of restriction endonucleases. The endonuclease cuts the DNA at given restriction sites, and the polymorphism of the generated DNA fragments can be analysed on gels.

Schizodeme: a set of individuals or stocks that share the same kDNA RFLP profile (Morel *et al.*, 1980). More broadly, a set of individuals or stocks that share the same RFLP profile.

Segregation, Hardy–Weinberg equilibrium: in a panmictic population of a diploid organism, with a gene of 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 a/a, a/b and b/b will be p^2, $2pq$ and q^2, respectively. This is the case in the population described under *allelic frequency* above. If the observed frequencies are statistically different from the expected ones, it is evidence that gene flow is restricted in the population under survey.

Sympatric, sympatry: living in the same geographical location (cf. *allopatry*).

Zymodeme: a set of individuals or stocks that share the same profile for a given set of isoenzyme markers.