

Chapter 3

Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases

Michel Tibayrenc

Abstract This chapter does not attempt to provide a comprehensive catalogue of the various methods of analysis and technologies used to characterize pathogens (molecular epidemiology). Rather it argues why molecular epidemiology should not be conceived without an extensive use of the concepts of population genetics and evolutionary biology. Moreover, it stresses that characterizing pathogens should open up to evaluating the impact of pathogens' genetic diversity on their relevant medical properties (downstream studies). Lastly, it presents the foreseeable future developments in this field, which has been upset by the exponential development of megatechnologies (massive sequencing, postgenomic studies, and bioinformatics).

3.1 Introduction

Before attempting to precisely define molecular epidemiology, it is interesting to observe it as a today's hot item. Using the term for a search in the SCOPUS database produces more than 15,000 references, most of them cited in the last few years. The term chiefly refers to transmissible diseases, but not exclusively. Some references concern other diseases such as cancer.

Considering the rapid growth of this field, there is no doubt that the publication of this book is quite timely. Actually, the premises of this burst onto the front stage were already apparent years ago. The field is at the crossroads of two major lines in modern biomedical research: the exponential development of molecular technologies and the preoccupying wave of emerging and reemerging infectious

M. Tibayrenc (✉)
Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle,
MIVEGEC/IDVEGEC (UM1-CNRS 5290-IRD 224), IRD Center,
BP 64501, 34374 Montpellier Cedex 5, France
e-mail: Michel.Tibayrenc@ird.fr

diseases (“The golden age of genetics and the dark age of infectious diseases” Tibayrenc 2001). As for the threat of infectious diseases, it is worth noting that the alarms sounded these last few years (SARS, avian flu) have not, in the end, been confirmed as major pandemics. Still, the fact remains that transmissible diseases remain by far the major killers of the human species on a worldwide basis. It can also be said that they more than ever constitute the major selection pressure of humankind (Haldane 1949), since many of them kill at the age of reproduction, or before it. Even when considering industrial countries, much more than 40 years ago, today infectious diseases are a major concern with the problems of AIDS, nosocomial infections, antibiotic resistance and multidrug-resistant pathogens. Obviously, considering that transmissible diseases are under control would be quite unrealistic. Can it be considered therefore that molecular epidemiology is a major contribution toward reaching this goal? The answer to this question is the very topic of this book, and more specifically, of the present chapter. The goal is neither to draw a comprehensive review of everything known in this field, nor to provide the reader with a handbook of molecular epidemiology. It is rather to give my personal views on how molecular epidemiology should be conceived today and the future avenues this field may take.

3.2 How to Define Molecular Epidemiology?

Reading the literature, it is obvious that there is no consensus on how to define the term. The classical definition of the Centers for Disease Control (CDC) in Atlanta is “the various biochemical and molecular techniques used to type and subtype pathogens” (CDC 1994). The goal in this definition is very clear: identify the species (type) and subspecies/strains (subtype) of pathogenic organisms. The method is also very clear: the use of all techniques offered by biochemistry and molecular biology toward reaching this goal. In itself, this definition is quite satisfactory. The goal of molecular epidemiology is to help elucidate the routes traveled by epidemics by tracking the relevant units of analysis, the entities that are responsible for the epidemics. The difficulties lie in the very definition and delimitation of these entities, as we will see below.

It should be emphasized that molecular epidemiology is no more than a particular case of molecular systematics applied to the specific case of the identification of pathogenic agents. It would be highly desirable that the strict rules used to identify other organisms with molecular markers (Avisé 2004) be applied to pathogens as well. As we will see, this is far from being the case.

3.3 Modern Technologies Have Upset the Field

Although there is no ultimate solution to all problems, there is no doubt that the arrival of powerful new technologies, and their lowering costs, have revolutionized the field. I will not detail here all the techniques that have been and are available for

pathogen characterization but rather explain how the techniques have made a major contribution to the field but are by no means a panacea.

In heroic times, the characterization of pathogenic agents and their strains relied on proteic markers, chiefly multilocus enzyme electrophoresis* (MLEE*, for a recent review see Tibayrenc 2009). These protein markers proved to be fine genetic markers and have allowed impressive progress, not only in routine strain typing, but also in basic bacterial population genetics (Selander et al. 1986). They have also been widely used for parasitic protozoa strain identification (Godfrey and Kilgour 1976). However, MLEE* has many drawbacks that definitely make it an outdated technology: (i) it is a time-consuming and delicate technique, (ii) it lacks resolution when the microevolutionary scale is considered, (iii) it is subject to homoplasy*, although this drawback has been exaggerated, (iv) and lastly, it requires bulk strain culturing, which leads to culturing bias* (see below).

Techniques that directly take DNA as the target molecule have definitely surpassed protein/MLEE* characterization. However, it should be stressed that the many valuable results gathered by MLEE* studies, especially when population genetics and the evolution of pathogens are considered, remain entirely valid and have been thoroughly confirmed by DNA techniques.

The major technological step that has upset the field of molecular epidemiology is polymerase chain reaction* (PCR). Indeed, PCR* makes it virtually possible to amplify the DNA of a single cell. Its potential resolution is therefore very high. Moreover, by using specific primers*, one can selectively amplify the DNA of the organism under study, even if it is mixed with foreign DNAs. These two properties obviate the need for bulk culturing, thus avoiding the culturing bias*.

Many different techniques rely on PCR* amplification, including random primed amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP). However, it can be said that two techniques presently dominate the field of molecular epidemiology, namely microsatellites* and multilocus sequence typing* (MLST). The main advantage of microsatellites* is their high resolution. MLST* has the advantage of its high standardization and perfect portability. In terms of resolution power, MLST* is not at the top: it is often surpassed by mere restriction fragment length polymorphism* (RFLP) of antigen genes. Lastly, thanks to the recent progress in automatic sequencing, whole genome sequencing (WGS) of many strains in bacterial species tends to become routinely accessible, and makes it possible to design many SNP (single nucleotide polymorphisms) to type strains of pathogens (Pearson et al. 2009) It should be stressed that in and of themselves, these techniques are mere labeling tools and say nothing about the biological and evolutionary properties of the organisms they aim to characterize (see below).

3.4 Has Molecular Epidemiology Helped Clinical Practice?

The answer to this question depends on what molecular epidemiology is considered to be. Strictly speaking, if we take the basic definition given by the CDC (1994), routine serological tests can be considered molecular epidemiologic tools. The same

holds true for PCR* identification of pathogen species. If we limit ourselves to these very basic tools, there is no doubt that molecular epidemiology has made a considerable contribution to routine medical practice. However, when a more common meaning is considered (that is to say, high-resolution identification of strains and clones), molecular epidemiology has not entered the routine daily practice of medical doctors (Humphreys 2004). It should instead be considered an epidemiological research tool. Still, the fact remains that it is not confined to basic research. It does help health professionals make practical decisions. The survey and control of the H5N1 epidemics, for example, would not have been possible without the contribution of molecular epidemiology.

Molecular epidemiology is at the very boundary between basic research, which might remain speculative, and operational biomedical research, in which sufficiently robust results should be unflinchingly sought, because decisions need to be made.

3.5 Can Molecular Epidemiology Be Conceived of Without the Evolutionary Concepts?

This is apparently possible, since molecular epidemiology articles are published in high-impact journals that do not contain a word of evolutionary biology or population genetics (Foxman 2007; Sintchenko et al. 2007). Usually these articles rely on the implicit working hypothesis that natural populations of bacteria are composed of clones that undergo no genetic modifications over time. We will see that this is often untrue.

Logically speaking, evolutionary concepts should appear to be indispensable to analyze molecular data. The molecular polymorphism of organisms has been shaped by evolution. It is therefore rational to interpret it in terms of evolutionary genetics. When plants, mammals, insects, reptiles, etc. are investigated, it would appear nonsensical not to use population genetics and evolutionary biology concepts to analyze their molecular diversity (Avisé 2004). Pathogens should not constitute a special case. Quite the opposite, I would say, because their formal genetics and mating strategies still remain mysterious, although considerable progress has been achieved in the last 20 years.

For years I have called for the inclusion of population genetics and evolutionary biology in the very definition of molecular epidemiology (Tibayrenc 1995, 2005).

According to this vision, molecular epidemiology should not limit itself to lazy band counting and mere dendrograms automatically generated by appropriate software. It should rely on classical population genetics (de Meeûs et al. 2007) and phylogenetic approaches (Hall and Barlow 2006).

This vision advocates molecular epidemiology as an exploration, using evolutionary biology's conceptual tools, of how pathogen populations are distributed and how they evolve.

3.6 Downstream Research

Additionally, I have long proposed that beyond this indispensable inventory, the consequences of pathogens' genetic diversity on their relevant biomedical properties (a disease's clinical diversity, sensitivity to drugs and vaccines, serological and molecular diagnosis) be explored (downstream studies, see Tibayrenc 1995). The minimum would be, for example, that a new drug or a new vaccine be tested on a set of selected strains that are representative of the entire genetic diversity of the species. With micropathogens, if technically feasible, it appears indispensable to work with laboratory-cloned stocks, with verification of the cloning under the microscope. Too many studies rely on noncloned stocks that are often composed of various different genotypes. This makes these mixtures of stocks highly unstable, and hence the results of such studies are poorly reproducible.

3.7 Reticulate Evolution of Pathogens and Its Implications

In brief, this is the main result yielded by evolutionary studies dealing with pathogens. This chapter is not the place for a long discourse on the evolutionary biology of microbes. To make a long story short, this result has been reached with two complementary means: population genetic statistics and phylogenetic analysis.

The term "reticulate evolution" by itself refers to the situation where separate evolution of distinct genetic lines is occasionally countered by limited genetic exchange. This situation is found in many plant species (Avice 2004).

Pathogenic microbes have evidenced a similar pattern in many, if not most, species (Tibayrenc et al. 1990, 1991; Awadalla 2003; Heitman 2006). Unexpectedly, many micropathogens exhibit sex in the broad sense of the term (i.e., exchange of genetic material between different cells, whatever the cytological mechanism involved). Its precise mechanisms vary considerably: conjugation, transduction and transformation in bacteria, classical meiosis in *Trypanosoma brucei*, the agent of human African trypanosomiasis (Jenni et al. 1986), and nonmeiotic hybridization in *Trypanosoma cruzi*, the agent of Chagas disease (Gaunt et al. 2003). However, the results in evolutionary terms are the same: departures from panmictic* expectation and from a purely phylogenetic view.

Departures from panmictic* expectations can be evaluated with classical segregation tests relying on Hardy-Weinberg and F-statistics (de Meeûs et al. 2007). However, these tests are based on the assumption that the organism under survey is diploid. This makes the use of these approaches impossible for haploid organisms (bacteria, the blood stage of *Plasmodium*) and debatable for those organisms for which diploidy is not fully ascertained (*Leishmania*, *Trypanosoma*). Recombination tests (linkage disequilibrium* analysis, Tibayrenc et al. 1990; Maynard Smith et al. 1993) avoid this shortcoming and are applicable whatever the ploidy of the organism may be, even if it is unknown, or if the molecular profiles make it impossible to

identify individual alleles (Tibayrenc et al. 1991). A considerable advantage of recombination tests is that, by definition, they are the only ones able to test for the stability of multilocus associations. This is the very goal of molecular epidemiology, as we will emphasize later.

Phylogenetic analysis, with its many different approaches and specialized softwares, looks for the presence and age of strictly separated genetic units. Its very definition makes departures from such an ideal image in themselves an indication of genetic exchange and its intensity. With highly recombining pathogens such as *Neisseria gonorrhoeae* and *Helicobacter pylori*, any phylogenetic analysis is virtually impossible. Strong evidence for rarity or absence of recombination is a fair agreement between phylogenetic trees designed from different genetic markers, as seen in *T. cruzi* (see Fig. 3.1). This is a particular case of the concordance principle, which states that if a hypothesis is valid, it is increasingly confirmed by accumulating evidence (Avice 2004).

Pathogens show a continuum between these two extreme pictures, from highly recombining (*Helicobacter*, *Neisseria*) to scarcely recombining (*T. cruzi*, *Salmonella*, *Escherichia coli*).

The implications of this for molecular epidemiology and downstream studies are far-reaching. The more the species under study is composed of discrete, stable genetic lines (structured species, Tibayrenc 1995), the more these lines will be convenient targets for epidemiological follow-up and downstream studies. On the contrary, if genetic recombination is frequent, such discrete lines will not be present, and typing of stable multilocus genotypes will be impossible. In this case, only identification and follow-up of individual genes will be possible (see below).

3.8 Units of Identification: Genes, Clones, Strains, Subspecies, and Species

Founding molecular epidemiology on firm evolutionary concepts requires clearly defining the units of analysis to be used (Dijkshoorn et al. 2000).

The least questionable of these units of analysis is the gene. Many molecular epidemiology studies are based on the identification and follow-up of genes of interest (for example, genes of virulence, of drug resistance), conveniently labeled by specific PCR primers*. Such studies are possible whatever the population structure and recombination level of the pathogen under survey may be. However, it is more propitious to base these studies on a convenient and thorough population genetics framework of the species under study.

“Clone” should be understood here in its strict genetic meaning, that is to say, durable multilocus association. With this sense, a “clonal” species refers to all cases where the offspring genotypes are identical or almost identical to the parental ones, and where the propagation of multilocus genotypes that are stable in space and time are observed. This definition includes not only species that propagate themselves by classical mitotic propagation (the case of most bacterial species), but also many

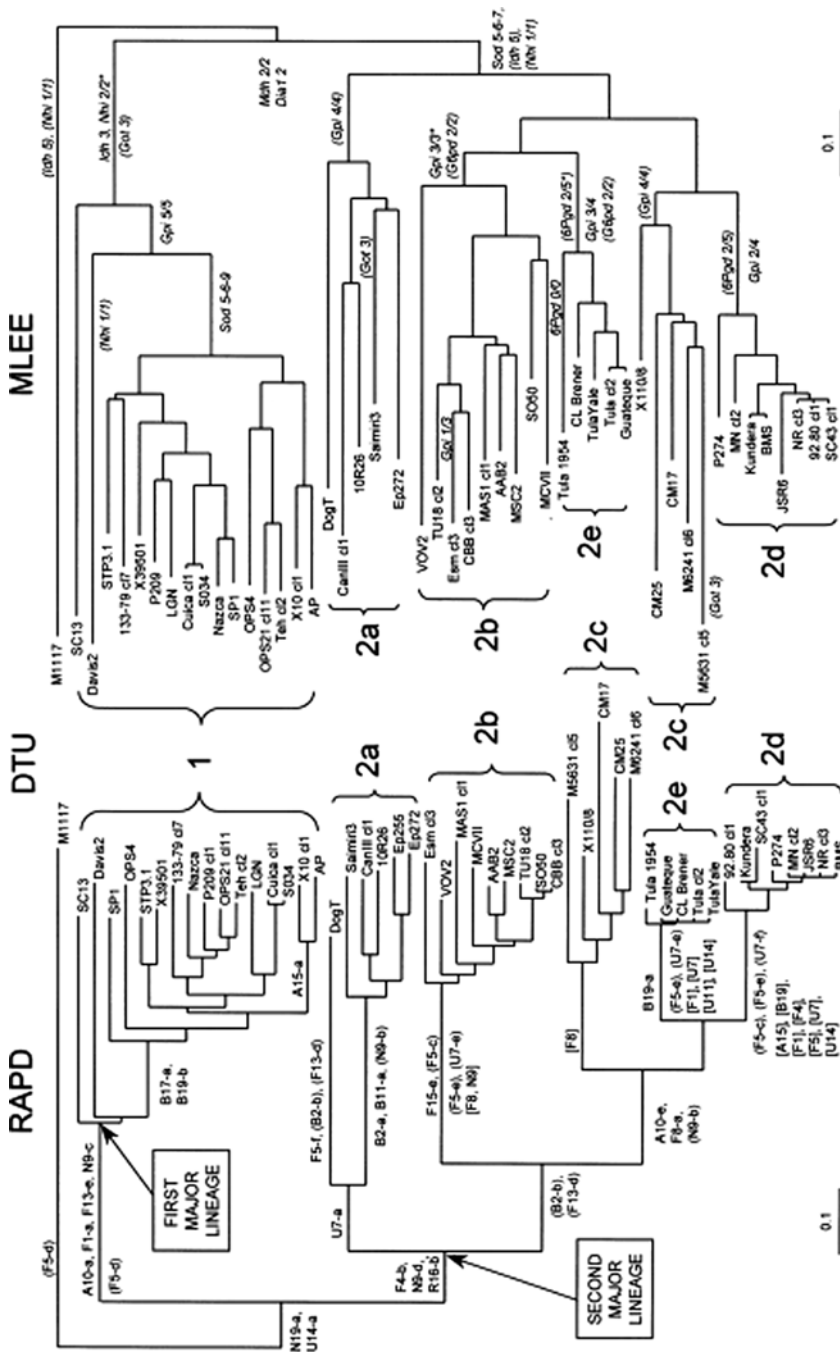


Fig. 3.1 Two phylogenetic trees depicting the evolutionary relationships among *T. cruzi* genotypes: isoenzymes (left) and RAPD (right). The strong similarity between the two trees is an extreme manifestation of linkage disequilibrium (After Brisse et al. 2000)

cases of parthenogenesis*, homogamy*, and self-fertilization* in homozygotes. Extreme inbreeding in this sense is not an alternative model to clonality (Rougeron et al. 2009), but rather a particular case of it (Tibayrenc et al. 2010).

All species in which genetic recombination is either absent or severely restricted should be considered clonal according to this definition, so that durable multilocus associations are not disrupted by recombination and behave like frozen genotypes or genetic photocopies. An important parameter to consider is the lifetime of such multilocus genotypes. Maynard Smith et al. (1993) distinguished between epidemic clonality and clonal evolution. The first refers to those species in which occasional bursts of genetic clones soon vanish in the common gene pool of a recombining species. *Neisseria meningitidis* is an example of this. Clonal evolution concerns those species in which clones remain stable at an evolutionary scale. *Trypanosoma cruzi* and *Salmonella* sp. are examples of this. Obviously, the consequences of the two models for molecular epidemiology and downstream studies are drastically different.

The term “strain” is widely used in the microbiology literature. However, its definition is a delicate one. Many specialists (including myself) use it with the sense of “stock”, that is to say, a culture of a pathogen with a given origin. The World Health Organization’s definition is quite different: it designates a set of stocks with various origins, but that share common defined properties (virulence, etc.). For molecular epidemiology, “strain” should designate a precise multilocus genotype, which is assumed to have a common, recent ancestry, in other words, a genetic clone. It is relevant to note that strains cannot exist in nonclonal, recombining species, although many epidemiologists maintain intuitively that such species have strains as well.

The definition of species is a never-ending story. An innumerable number of different definitions have been proposed by many authors (Hey 2001). However, one can distinguish between four main concepts to define species.

1. The biological species concept is the most classical one and refers to those communities of organisms that are potentially interfertile (Dobzhansky 1937). Obviously, this concept refers to classical, sexual species and is hardly useable for pathogenic microorganisms.
2. The phenetic species concept is based on the assumption that organisms pertaining to the same species are assumed to share many properties, while the contrary applies to organisms that belong to different species. To define species with this approach, as many characters as possible are taken into account without prioritizing them (Sneath and Sokal 1973). This concept could be useful to define pathogen species, although this approach is time-consuming.
3. The phenotypic concept is still widely used to define new species in insects, mammals, etc. In such cases, it is mainly based on morphological characters (color, shape, etc.). In microbiology, it can be said that many microbe species have been defined on phenotypic characteristics, either epidemiological traits or pathological properties. For example, *Mycobacterium leprae* is the causative agent of leprosy. *Leishmania infantum* is the agent of leishmaniosis in infants. *Leishmania panamensis* was first described in Panama. The phenotypic species concept applied to pathogenic microbes therefore aims at conveying relevant medical information.

4. The phylogenetic species concept considers that species should be equated to clades* or individual phylogenetic lines (Cracraft 1983). For the above-described reasons, this approach is not easy to use when micropathogens are considered due to the predominance of reticulate evolution in them. By definition, clades* do not exchange genes.

The definition of species for pathogens is a brain teaser. However, the extreme attitude that species are impossible to define in the case of microbes (de Meeûs et al. 2003) should be rejected. Health professionals need operational units of analysis to work on. Microbe species are not mere fantasies or video games: *Trypanosoma cruzi* cannot cause sleeping sickness. Even when considering the problem of reticulate evolution, one cannot see the microbial world as a genetic continuum. There are profound discontinuities, even if their boundaries are sometimes difficult to delineate. This means that the phylogenetic species concept can be used with some caution. The pragmatic approach I have recommended (Tibayrenc 2006) is to use the phylogenetic and phenotypic concepts jointly to describe new pathogen species. New species should be described only when they correspond to a phylogenetic reality and feature relevant medical/epidemiological properties. The use of the phylogenetic species concept means that species can be identified using the tools of molecular epidemiology.

In classical zoology, subspecies designate geographical variants of a given species that can be easily identified by specific traits, generally morphological particularities. The concept is of little value in medical microbiology. There are no strict criteria to define microbial subspecies; the tendency is for authors to describe subspecies when they dare not make real new species (Schönian et al. 2010).

3.9 Helpful, Although Not So Successful, Operational Concepts: Clonets and Discrete Typing Units/Tags

The classical concepts of population genetics and phylogenetic analysis are sometimes poorly adapted to the peculiarities of the microbial world and the specific demands of epidemiological follow-up.

3.9.1 Clones and Clonets

Let us consider a species that has been evidenced to be clonal by population genetic analysis. The use of a given set of genetic markers, say 20 microsatellite* loci, has individualized stocks that share the same multilocus combination. Do they represent genuine clones? Probably not. If we use 30 microsatellite* loci instead of 20, we will probably encompass additional genetic variability within each of these supposed

clones. To get around the difficulty, the term “clonet” has been coined to refer to those sets of stocks that appear to be identical to a given set of genetic markers in a basically clonal* species (Tibayrenc and Ayala 1991). The clonets are relevant units of analysis for epidemiological tracking. They should be delimited by a sufficient range of genetic markers, according to the time and space scale considered (see below).

3.9.2 *Discrete Typing Units and Tags*

I have already presented the implication of reticulate evolution for phylogenetic/ cladistic* analysis. The presence of some gene flow among genetic lines, and the fact that in hybridization events some genetic lines have two ancestors instead of one, make the clade* concept invalid. There are obvious genetic discontinuities between and within pathogen species that deserve to be described and delimited. To avoid the problem, I have proposed the descriptive and operational concept of discrete typing unit (DTU) (Tibayrenc 1998). DTUs are sets of stocks that are genetically closer to each other than to any other stock and are identifiable by common molecular, genetic, biochemical, or immunological markers called tags. DTUs correspond to reliable units of analysis for molecular epidemiology and downstream studies.

3.10 **Can This Field Be Unified?**

People designing and using molecular epidemiology sorely need to realize that they have common goals and that their approaches should converge to a large extent. Whatever the pathogen under survey – parasite, fungus, bacterium, virus, prion, be it of medical, veterinary or agronomical relevance – species, strains, clones, and genes of interest should be reliably identified using evolutionary concepts. The genetic variability of the pathogen considered should be taken into account for vaccine/drug design, epidemiological follow-up and clinical studies.

Yet the world of pathogens is too heterogeneous for building totally standardized approaches. Species concepts are not the same for parasites, bacteria, and viruses. What’s more, the molecular tools cannot be the same.

However, when the pathogens considered are closely related, it is nonsense to design radically different approaches. An example is the description of new species. With *Leishmania* parasites, many new species have been described, some of them based on doubtful criteria (Van der Auwera et al. 2011). On the other hand, until now, specialists of Chagas disease have agreed on considering *Trypanosoma cruzi*, the causative agent of Chagas disease, a unique, although genetically extremely heterogeneous, species (Zingales et al. 2009). *Leishmania* specialists are splitters, Chagas specialists are lumpers. These differences have little evolutionary or medical justification.

3.11 The Crucial Parameters of Time and Space Scales

Molecular epidemiology may act on very different time and space scales. For example, one may wish to survey the spread of one of the six *T. cruzi* DTUs over the entire American continent. A totally different case would be the identification and follow-up of a unique methicillin-resistant clone of *Staphylococcus aureus* that has contaminated a hospital intensive care unit. Obviously the molecular and statistical tools will have to be adapted to the case at hand. In the first case, classical phylogenetic analysis and markers that have a slow molecular clock* (multilocus sequence typing*, multilocus enzyme electrophoresis*) can be used. In the second case, markers with a finer resolution, in other words, a faster molecular clock*, such as microsatellites* or some RFLP* markers, will have to be used.

3.12 New Problems: Typing Noncultivable Pathogens

Koch's postulates state that:

- (a) The microorganism must be found in abundance in all organisms suffering from the disease but should not be found in healthy animals.
- (b) The microorganism must be isolated from a diseased organism and grown in pure culture.
- (c) The cultured microorganism should cause disease when introduced into a healthy organism.
- (d) The microorganism must be reisolated from the inoculated, diseased experimental host and identified as being identical to the original specific causative agent.

These postulates are now invalidated by the fact that many pathologies, such as Crohn disease, could be caused by pathogens that cannot be cultured.

One of the most remarkable achievements of molecular epidemiology is to make it possible to characterize these unknown pathogenic agents. Sequences of nonspecific genes such as ribosomal RNA genes are used to reach this goal (see Frank and Feldman 2007 for a recent, exhaustive review).

3.13 Conclusion: The Future of the Field: Total Pathogen Profiling, Whole Genome Sequencing, Integrated Genetic Epidemiology

Technological progress will continue at an increased speed, just as the cost of analyses will continue to be lowered. This will permit holistic approaches that were not possible with limited technological resources. Pathogen profiling is one of the new strategies that will emerge from use of these new tools

(Sintchenko et al. 2007). It integrates molecular data (genome, transcriptome, proteome, metabolome) with clinical and epidemiological data, assisted by geographic information systems (GIS).

Pathogen profiling will be even more valuable if it relies on the survey of pathogen's whole genomes, which becomes now more and more accessible, at least for viruses and bacteria (Pearson et al. 2009). This opens the way to the new field of population genomics (Tibayrenc 2005).

Another approach long called for (Tibayrenc 1998) is to concurrently consider the role played in the transmission and severity of infectious diseases, by the host's, the pathogen's and the vector's (in the case of vector-borne diseases) genetic diversity as well as the interactions between the three (coevolution phenomena). Such a holistic approach (integrated genetic epidemiology of infectious diseases) should be extended to the entire genome of hosts, pathogens and vector (population genomics). This is the very topic of the new journal *Infection, Genetics and Evolution* (Elsevier; <http://www.elsevier.com/locate/meegid>) and the *Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases* (MEEGID; <http://www.meegidconference.com/>) congresses. Such an approach has been proposed for the study of Chagas disease (Tibayrenc et al. 2010).

Glossary of Specialized Terms (* in text)

Clade evolutionary line defined by cladistic analysis. A clade is monophyletic* and is genetically separated (i.e., evolves independently) from other clades.

Cladistic analysis a specific approach of phylogenetic analysis based on the polarization of characters that are shared between ancestral (plesiomorphic) and derived (apomorphic) characteristics. Only those apomorphic characters shared by all members of a given clade* (synapomorphic character) are considered to convey relevant phylogenetic information (Hennig 1966).

Culturing bias This term refers to the situation where a host is infected by several different genotypes of a given pathogen. When culturing an isolate of this pathogen, some genotypes will tend to dominate to the detriment of others. At the end of the culturing process, the collection of genotypes does not reflect the original isolate (Tibayrenc 1995).

Homogamy preferential mating between individuals that are genetically identical or extremely similar to each other.

Homoplasy Common possession by distinct evolutionary units of identical characters that do not originate from a common ancestor. The origin of homoplastic traits include the following: (a) convergence (possession of identical characters derived from different ancestral characters, due to convergent evolutionary pressure, for example, the fins of fish and dolphins); (b) parallelism (possession of identical characters derived from a single ancestral character and generated independently in different evolutionary units); and (c) reversion (restoration of an ancestral character from a derived character).

Isoenzymes Protein extracts of the biological samples under analysis are separated by electrophoresis. The gel is then processed with a histochemical reaction involving the specific substrate of a given enzyme. This enzyme's zone of activity is then specifically stained on the gel. From one sample to another, migration differences can be visible for this same enzyme. These different electrophoretic forms of a given enzyme are referred to as isoenzymes or isozymes. When given isoenzymes are driven by different alleles of a single gene, they are more specifically referred to as alloenzymes or allozymes. Differences in migration result from different overall electrical charges between isoenzymes. Overall electric charges are a resultant of the individual electric charges of each amino acid (AA) of a given enzyme. The AA sequence is the direct result of the DNA sequence of the gene that codes for this enzyme. It is therefore considered (and verified) that isoenzyme polymorphism is a faithful reflection of the genetic polymorphism of the organism under study.

Linkage disequilibrium Nonrandom association of genotypes occurring at different loci.

Microsatellite A short DNA sequence of DNA, usually 1–4 bp long, that is repeated together in a row along the DNA molecule. Microsatellites are fast-evolving markers, with a high resolution level and are found in many organisms, including pathogens.

Molecular clock In its strict, original sense (more correctly called the DNA clock hypothesis), the concept that the rate of nucleotide substitutions in DNA remains constant over time. In a broader sense, simply how fast the genomic part that codes for the variability of the marker considered evolves. This speed is driven by the mutation rate. It may be regular or irregular.

Monophyletic an evolutionary line that has a unique ancestor.

Multilocus sequence typing (Maiden et al. 1998) is a highly standardized approach based on the sequencing of 450-bp parts of a set of housekeeping genes (usually seven). It has been widely used for a high number of bacterial species and some eukaryotic pathogens as well. The main advantage of MLST is its perfect portability (possibility of reliably communicating results among different laboratories), since sequences can be simply emailed. Strains that share the same combination of alleles are referred to as sequence types (ST). Strains that share 7/7 alleles = consensus group; strains that share 6/7 alleles = single-locus variants (SLV); strains that share 5/7 alleles are double-locus variants (DLV). SLV and DLV = clonal complexes.

Panmixia panmictic: Situation in which genetic exchanges occur randomly in the population under survey.

Parthenogenesis Reproduction by the development of a single gamete without fertilization by a gamete of the opposite sex.

Polymerase chain reaction (PCR) A technique that copies the complementary strands of a target DNA chain through a set of cycles until the needed DNA amount is produced. PCR uses synthesized primers* whose nucleotide sequences are complementary to the DNA flanking the target region. The DNA is heated to separate the complementary strands, then cooled to have the primers bind

to the flanking sequences. The enzyme Taq DNA polymerase is added and the reaction is left to pass through the required number of replication cycles.

Primer a short DNA sequence used in PCR* technologies, that anneals to a single strand of DNA and acts as a starting point to initiate DNA polymerization mediated by the enzyme Taq DNA polymerase.

Restriction fragment length polymorphism (RFLP) variability in the DNA of a given organism evidenced by the use of restriction endonuclease bacterial enzymes. The endonuclease cuts the DNA at a specific restriction site with a given sequence, and the polymorphism of the DNA fragments thus obtained can be visualized on gels, either directly by ethidium bromide staining or by Southern blot hybridization with specific probes.

Self-fertilization fertilization by the union of male and female gametes from the same individual

References

- Avisé JC (2004) Molecular markers, natural history and evolution, 2nd edn. Chapman & Hall, New York
- Awadalla P (2003) The evolutionary genomics of pathogen recombination. *Nat Rev Genet* 4:50–60
- Brisse S, Barnabé C, Tibayrenc M (2000) Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. *Int J Parasitol* 30:35–44
- Centers for Disease Control and Prevention (1994) Addressing emerging infectious diseases threat. A prevention strategy for the United States, 27, Atlanta
- Cracraft J (1983) Species concept and speciation analysis. In: Johnson RF (ed) Current ornithology. Plenum Press, New York, pp 159–187
- de Meeûs T, Durand P, Renaud F (2003) Species concepts: what for? *Trends Parasitol* 19:425–427
- de Meeûs T, McCoy K, Prugnolle F et al (2007) Population genetics and molecular epidemiology or how to “débusquer la bête”. *Infect Genet Evol* 7:308–332
- Dijkshoorn L, Ursing BM, Ursing JB (2000) Strain, clone and species: comments on three basic concepts of bacteriology. *J Med Microbiol* 40:397–401
- Dobzhansky T (1937) Genetics and the origin of species. Columbia University Press, New York
- Foxman B (2007) Contributions of molecular epidemiology to the understanding of infectious disease transmission, pathogenesis, and evolution. *Ann Epidemiol* 17:148–156
- Frank DN, Feldman RA (2007) Molecular-phylogenetic strategies for characterization of uncultured pathogens. In: Tibayrenc M (ed) Encyclopedia of infectious diseases- modern methodologies. Wiley, New York, pp 57–73
- Gaunt MW, Yeo M, Frame IA et al (2003) Mechanism of genetic exchange in American trypanosomes. *Nature* 421:936–939
- Godfrey DG, Kilgour V (1976) Enzyme electrophoresis in characterizing the causative agent of Gambian Trypanosomiasis. *T Roy Soc Trop Med H* 71:217–225
- Haldane JBS (1949) Disease and evolution. *La ricerca scientifica* 19(Suppl):68–76
- Hall BG, Barlow M (2006) Phylogenetic analysis as a tool in molecular epidemiology of infectious diseases. *Ann Epidemiol* 16:157–169
- Heitman J (2006) Sexual reproduction and the evolution of microbial pathogens. *Current Biol* 16:R711–R725
- Hennig W (1966) Phylogenetic systematics. Univ Illinois Press, Urbana
- Hey J (2001) The mind of the species problem. *Trends Ecol Evol* 16:326–329

- Humphreys H (2004) Does molecular typing make any contribution to the care of patients with infection? *Clin Microbiol Infect* 10:269–271
- Jenni L, Marti S, Schweizer J et al (1986) Hybrid formation between African trypanosomes during cyclical transmission. *Nature* 322:173–175
- Maiden MCJ, Bygraves JA, Feil E et al (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 95:3140–3145
- Maynard Smith J, Smith NH, O'Rourke M et al (1993) How clonal are bacteria? *Proc Natl Acad Sci USA* 90:4384–4388
- Pearson P, Richard T, Okinaka RT et al (2009) Phylogenetic understanding of clonal populations in an era of whole genome sequencing. *Infect Genet Evol* 9:1010–1019
- Rougeron V, De Meeùs T, Hide M et al (2009) Extreme inbreeding in *Leishmania braziliensis*. *Proc Natl Acad Sci USA* 25:10224–10229
- Schönián G, Mauricio I, Cupolillo E (2010) Is it time to revise the nomenclature of *Leishmania*? *Trends Parasitol* 26:466–469
- Selander RK, Musser JM, Caugant DA et al (1986) Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol* 51:873–884
- Sintchenko V, Iredell JR, Gilbert GL (2007) Pathogen profiling for disease management and surveillance. *Nature Rev Microbiol* 5:464–470
- Sneath PHA, Sokal RR (1973) Numerical taxonomy. In: Kennedy D, Park RB (eds) *The principle and practice of numerical classification*. Freeman, San Francisco, pp 537
- Tibayrenc M (1995) Population genetics of parasitic protozoa and other microorganisms. *Adv Parasitol* 36:47–115
- Tibayrenc M (1998) Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. *Int J Parasitol* 28:85–104
- Tibayrenc M (2001) The golden age of genetics and the dark age of infectious diseases. *Infect Genet Evol* 1:1–2
- Tibayrenc M (2005) Bridging up the gap between molecular epidemiologists and evolutionists. *Trends Microbiol* 13:575–580
- Tibayrenc M (2006) The species concept in parasites and other pathogens: a pragmatic approach? *Trends Parasitol* 22:66–70
- Tibayrenc M (2009) Multilocus enzyme electrophoresis for parasites and other pathogens. In: Caugant DA (ed) *Methods in molecular biology: molecular epidemiology of microorganisms*. Humana Press, Totowa, pp 13–25
- Tibayrenc M, Ayala FJ (1991) Towards a population genetics of microorganisms: the clonal theory of parasitic protozoa. *Parasitol Today* 7:228–232
- Tibayrenc M, Kjellberg F, Ayala FJ (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. *Proc Natl Acad Sci USA* 87:2414–2418
- Tibayrenc M, Kjellberg F, Arnaud J et al (1991) Are eucaryotic microorganisms clonal or sexual? A population genetics vantage. *Proc Natl Acad Sci USA* 88:5129–5133
- Tibayrenc M, Telleria J, Diosque P et al (2010) Integrated genetic epidemiology of Chagas disease. In: Tibayrenc M (ed) *Genetics and evolution of infectious diseases*. Elsevier Insights, Burlington
- Van der Auwera G, Fraga J, Montalvo AM et al (2011) *Leishmania* taxonomy up for promotion? *Trends Parasitol* 27:49–50
- Zingales B, Andrade SG, Briones MRS et al (2009) A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Mem Inst Oswaldo Cruz* 104:1051–1054