Discussion

Infectious diseases and the golden age of phylogenetics: An E-debate

W. Fitch\textsuperscript{a}, S. Brisse\textsuperscript{b}, J. Stevens\textsuperscript{c}, M. Tibayrenc\textsuperscript{d,}\textsuperscript{*}

\textsuperscript{a} Department of Ecology and Evolutionary Biology, UC Irvine, Irvine, CA 92717, USA
\textsuperscript{b} Eijkman-Wincle Institute, Utrecht University, Academic Hospital Utrecht, AZU G04614, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands
\textsuperscript{c} Hatherley Laboratories, Department of Biological Sciences, University of Exeter, Prince of Wales Road, Exeter EX4 4PS, UK
\textsuperscript{d} UMR CNRS/IRD 9926, Genetics of Infectious Diseases, IRD, BP 5045, 911 Avenue Agropolis, 34032 Montpellier Cedex 1, France

Accepted 4 May 2001

1. First question: Many researchers working on molecular epidemiology (molecular epidemiology, CDC definition (1997): “the various techniques derived from immunology, biochemistry, and genetics for typing or subtyping pathogens”) of infectious diseases think they can interpret their data with no contribution from evolutionary genetic concepts. Do you find this approach misleading, and if so, why?

WF: I am not sure I have addressed the question you have in mind. I could easily throw the following away and do something more useful in the way of responding. Perhaps, if I understood the question better, I could write something more useful for the community. I don’t find the view of those who believe they need no help from evolutionary considerations, genetic and otherwise, when interpreting their data so much misleading as it is self-strangulation. If one has a limited question then perhaps one may justify such limited data collecting and analysis. But if one is interested in mechanism and evolution, in the selective forces that the pathogen is responding to and how it is doing so, in what characteristics of these changes are predictable and which useful for treatment, etc. then one wants all the appropriate tools and information affecting these questions that one can get. Our group has recently enumerated 18 individual amino acid positions of the human influenza type A virus hemagglutinin H(3) HA1 that are under strong positive selection to change, presumably to attenuate immune selection (Bush et al., 1999b) and we then went on to use that information to successfully predict, in 9 of 11 years tested, which of the influenza strains present in the population 1 year would be the progenitors of future generations of the virus (Bush et al., 1999a). This work has not yet advanced to the point where clinical practice has been greatly and usefully altered, but it is hard to imagine how such studies will not help our understanding of the pathogens being studied. A different example is the obtaining of the information that a pathogen is evolving clonally (Rodriguez et al., 1996) an important piece of information in determining the nature of a potential vaccine. There is another sense in which those with blinders on can say they have no need for phylogenies. It is true that everything one might prove with a tree can be learned, proved, and written without the use of such a tree. But that would lead in phylogenetic and other papers to impenetrable forests of thousands of words where one tree would give a much more lucid picture of the situation.

JS: A few points and observations in response to the question: “Do many Researchers try to interpret their data with no reference to underlying evolutionary genetic concepts?” . . . may be I’m showing my naivety, but I wouldn’t have thought so. Practitioners and epidemiologists possibly, but this then brings us on to a further consideration: in Britain, we use the expression “horses for courses”, meaning something that is appropriate for one use or question may not necessarily be suited to another use. Here I think we may have a case of this: surely not every practitioner (or, arguably, epidemiologist) needs to know the full evolutionary history of the organism which he/she is characterizing for some medical purpose; nor, in all probability, will they have the time (or money) to consider such relationships.

As Walter states: “If one has a limited question then perhaps one may justify such limited data collecting and
analysis. But if one is interested in mechanism and evolution, in the **selective forces** that the pathogen is responding to and how it is doing so, in what characteristics of these changes are predictable and which useful for treatment, etc. then one wants all the appropriate tools and information affecting these questions that one can get”. [My emphasis on selective forces].

Accordingly, this raises the point that maybe we (as research scientists) are not the right people to be responding to this question.

From my own view, I think that although one may not immediately see the value of a phylogenetic approach, one can probably never run the risk of having too much information. For example, in forensic entomology, a discipline in which one might imagine that it would suffice only to be able to positively identify fly species, the use of a phylogenetic approach has allowed a finer degree of confidence to be placed on the identity of different fly species using molecular methods in cases when it is not possible to obtain a definitive ID by morphology (e.g. Wells and Sperling, 1999; Stevens and Wall, 2001).

In short, while I would always advocate the use of a phylogenetic and evolutionary approach, I am prepared to accept that those who have other objectives (for which they can convincingly justify the lack of a phylogenetic perspective) may be able to achieve their immediate aims without a phylogenetic interpretation.

p.s. I repeat: If there are _MANY_ researchers working on molecular epidemiology of infectious diseases who interpret their data with no contribution from evolutionary genetic concepts, it would be interesting to hear from them.

**SB:** Molecular epidemiology is widely used to track the source and route of contamination of pathogens, in part for the academic purposes of understanding better their ecology, epidemiology and evolution, but mostly for operational purposes, namely to control the spread of these pathogens. The principal fields of application of molecular epidemiology include nosocomial (hospital-acquired) infections, food-borne infections, and community infections (not to mention veterinary and plant biology applications). If one is to understand why the data are not always analyzed according to the canonical approaches of evolutionary genetics, it is important to realize that the major focus of molecular epidemiology is on exploitation of results for practical purposes. Molecular epidemiology has been ‘reinvented’ many times, driven by similar necessity in communities of researchers working on different pathogens (e.g. viruses, bacteria, fungi or parasites) and in different application fields. Population and evolutionary biologists have, on their side, started only later to use pathogens as models (Levin et al., 1999), and calls for unifying efforts and applying conceptual approaches of evolutionary genetics to molecular epidemiology are recent (Tibayrenc, 1996; Lal and Tibayrenc, 1997; www.cdc.gov/ncidod/EID/vol3no2/news244.htm). The gap between practical and academic approaches to the study of the genetic diversity of pathogens is illustrated by the fact that the ‘clone concept’ has been developed independently by clinical microbiologists and by population geneticists (Ørskov et al., 1976; Selander and Levin, 1980). Has the gap filled in, and should it?

It is easy to find, among the clinical microbiology studies making use of molecular typing, examples of misuse of approaches and concepts of evolutionary genetics. For example, in part due to the popularization of ‘push-button’ DNA fingerprinting analysis software such as BioNumerics (and its ancestor GelCompar, created by Applied Maths, Kortrijk, Belgium) and Taxotron (Institut Pasteur, Paris, France), dendrograms have become convenient ways to illustrate the genetic diversity of a set of strains. Conclusions on the identification of strain groups, clusters or lineages and strain relatedness are drawn, often with no concern for statistical significance, from the visual inspection of dendrograms mostly based on Pearson correlation of fingerprint density curves. This practice is not only an excellent means to irritate a evolutionary geneticist familiar with bootstrapping and parsimony approaches, but it can also leads to misleading conclusions. The concept of clonality is another clear example of the incomplete convergence of molecular epidemiology and clinical microbiology. The term ‘clonal’ covers a variety of meanings. When used by a population geneticist, it will refer to a population structure where genetic exchange and recombination are restricted to such an extent that bacterial clones are genetically stable and can spread over vast geographic areas and long periods of times. Differently, a set of strains will be referred to as clonal upon molecular epidemiology analysis when all strains are indistinguishable, within defined levels of accepted variation, by a given set of fingerprinting data. In that sense, in the frame of an outbreak investigation, clonal relatedness will mean that cross-transmission has occurred. For a molecular epidemiologist, two samples do not belong to the same ‘clone’ if they show more than three PFGE band differences upon pulse field gel electrophoresis analysis, whereas a clone, in the sense of a bacterial lineage derived by asexual propagation from an ancestral cell, may be composed of many different ‘types’, some of them possibly showing no relatedness at all when analyzed with fast-evolving markers, if the clone is ancient. The term is also used to signify that a microbial population, or a whole species, has limited variability (*Mycobacterium tuberculosis* is clonal), regardless of whether genetic recombination played a role in generating the limited variation observed or not. Finally, a ‘clonal population structure’ is often employed by the molecular epidemiologist to mean that the population is composed of a few, over-dominant and widespread clones, whereas such observation may correspond, in fact, to an epidemic clonality situation (Maynard Smith et al., 1993) and would not necessarily mean that recombination is restricted in the long run. Debate on the clone concept is still active in the clinical microbiology field (Dijkshoorn et al., 2000). It is unlikely that different usage (which may be seen as misuse by some) of given approaches or concept
will ever be resolved; but is it worrisome? Awareness of
the existence of diversity of usage is sufficient to avoid
misunderstanding in most of the cases. Moreover, most
molecular epidemiological studies are conducted over the
time scale of the nosocomial outbreak, typically a few
weeks to a few months, period during which the stability
of clones of most bacterial species is probably sufficient
to warrant a purely typological approach. On the other
hand, a too simplistic approach to the problem of ‘what is
a case of clonal transmission’ is often seen in molecular
epidemiology studies, with statements such as ‘the iso-
lates obtained from different patients are identical’ being
possibly misleading when the discriminatory power of the
typing method employed is not sufficient, ‘the isolates were
indistinguishable’ being the only firm conclusion possible.
Similarly, the hypothesis of transmission inferred from lack
of distinction among isolates is possibly wrong when the
amount of genetic diversity, population structure and preva-
ience of clones in the direct environment is not taken into
consideration. Finally, for studies involving microbes iso-
lated over long periods of time and across wide geographic
areas, it is often relevant not only to discriminate genotypes,
but also to establish their phylogenetic relationships and
evaluate the impact of genetic recombination on their sta-
bility and evolution. It is important to take this into account
in this era of development of standardization of typing tech-
niques, with the aim of constructing library typing systems
and centralized fingerprint databases rendering it possible
to compare bacterial isolates characterized by many inde-
pendent users and thus to better follow clones over space
and time (www.ewi.med.uu.nl/gene). Knowledge on how
often antigens and resistance genes may transfer among
clonal lineages is also fundamental to predict response of
pathogens to selective pressure imposed by the host im-
munity or antimicrobial agents. The situation of bacterial
taxonomic revision, with its sometimes slow acceptance by
the clinical microbiology community, may provide a useful
comparison. Taxonomy, seen as the working technical lan-
guage of microbiologists, is more easily adopted by them
if it presents practical advantages (Magee, 1993). Similarly,
typing practice may incorporate those evolutionary genetics
concepts that will prove useful in the daily routine and will
leave the others aside, respectfully ignoring their academic
correctness.

JS: Firstly, may I congratulate Sylvain for a very eloquent
and thorough piece, which has prompted me to a couple of
further observations and questions.

Should we always expect to be able to derive evolutionary
knowledge from all data pertaining to molecular epidemi-
ology of infectious diseases? For example, markers such
as isoenzymes and RAPDs, which make excellent tools for
molecular epidemiology at the level of disease outbreaks
(generally the population genetics level) are plagued with
homoplasy which can make them next to useless for re-
constructing species phylogenies, e.g. the case of African
trypanosomes. This isn’t to say that we shouldn’t strive
to collect data in a way that it can aid an evolutionary
interpretation, merely that perhaps we cannot expect it to
be the case all the time.

SB: Thanks for your congratulations. I guess your
comment refers to the general fact that every molecular
tool has its defined range of applications, in terms of level
of discriminatory power, maximal evolutionary level of
resolution, and risk of homoplasy. I fully agree with this.

MT: Evolutionary genetics is not limited to species phy-
logeny. Every genetic, molecular, biochemical marker has
its own range of excellence, linked to its power of resolu-
tion, itself linked to its molecular clock. Isoenzymes and
RAPD generate homoplasy and lack resolution for higher
levels of phylogenetic divergence. For lower levels, they
remain powerful phylogenetic tools. They are therefore not
limited to epidemiological tracking and population genet-
ic studies. When a species is not too “big” (the case for many
Leishmania species for example), it is quite reliable to de-
limitate it with such markers. In this case, they are much
more reliable than conserved gene sequences, which show
no variability at low levels of phylogenetic divergence. On
the contrary, for “picoevolutionary” levels, isoenzymes and
RAPD lack resolution by comparison with markers which
have a faster molecular clock such as microsatellites.

The polymorphism of all genetic markers has been shaped
by evolution. They are therefore all able to yield evolutionary
information. The question remains: “is it always useful for
medical applications?”

2. Second question: The various genome sequencing
programs will generate considerable amounts of
sequence data. Do you expect from this technological
leap a revolution in molecular phylogeny?

JS: Quick answer: Yes, no, yes!

Yes, because more data will give the opportunity to
explore more evolutionary relationships.

No, because genome analyses typically focus on a single
model taxon from a group, which doesn’t provide enough
across-taxa coverage to allow meaningful phylogenetic
comparisons.

Yes, because as more is known about different genomes,
previously unexplored genes may become available as tar-
gets for phylogenetic analysis. There may also be lag phase
where phylogenetic methods developed initially for single
genotype analyses are refined to analyse true multi-gene
relationships.

SB: Whole genome sequence data will undoubtedly
increase our insight into how genomes evolve and will pos-
sibly lead to understand the origin of life. To achieve that,
comparative studies will be most useful. This is true at all
levels of phylogenetic divergence: for example, comparing

1 One would not expect all genes to be of equal utility/resolution for
resolving the phylogeny of a given set of species.
the genomes of organisms belonging to different domains of life (several genomes of both domains Archaea and Bacteria have been sequenced), to different species within a bacterial family (McClelland et al., 2000), or two genomes within a single species (Alm et al., 1999) revealed conserved genes, genome reorganisations (e.g. insertions/deletions), and more or less promiscuous horizontal transfer events. At the species level or below (Alm et al. op. cit.), genome comparison allow to identify potentially informative phylogenetic markers that may be useful for population genetics and molecular epidemiology as well. No doubt, these comparisons will reveal new patterns, and possibly new processes, on how microbes evolve.

WF: New methods find us discovering new things we didn’t dream of in their absence. The wealth of new data has spawned many new methods and that occurrence will continue to grow. And of course the data on which these programs can operate will also expand and foster new ideas and research directions. I can’t think of an easier call for the scientific future than that much new will be coming soon.

3. Third question: Phylogeny and the species concept in pathogenic microorganisms. Most pathogenic microorganisms do not undergo regular sexual recombination. The biological species concept or BSC (Dobzhansky, 1937) is therefore of little help to define species. There is no consensus on a species concept valid for parasitic protozoa, microfungi, bacteria or viruses. Many species have been defined on phenotypic characters dealing mainly with relevant medical properties (virulence, epidemiological parameters). Do you think that the phylogenetic species concept or PSC (Cracraft, 1983) should be used more widely or exclusively to define species of pathogenic microorganisms? You may cite practical examples based on your personal experience in your answer.

WF: I don’t think this should much of an issue. Long ago investigators needed a way to identify to others the organism that they were reporting on. This led to the field of taxonomy or systematics. Because one’s organism was normally clearly distinguishable from its closest relative, investigators came to think of their groups as “natural” units and called them species. The question then arose, “What is the reason for these natural groups; what do species signify?”

Mayr’s answer was the biological species concept which he defined as a population of individuals that could breed freely within the group but not between other groups. But breeding implies diplioidy and sexual reassortment of the genes. Since most of the eubacterial and Archaean organisms are haploid and have no sexual reassortment process, we have the dilemma posed by the question.

But perhaps the problem is in expecting groups to be “natural”. While the biological species concept is very useful in many cases, more and more we are observing that two species are not really differentially discernable at the level of gene sequences which is more of a problem than simply being able to mate lions and tigers to get ligers and tigons. While morphological and molecular gaps do exist between many species, gaps are not by nature obligatory. Thus, the use of the term species rests primarily on it’s utility as an identifier of organisms and should be accepted as such.

But taxonomy is more than an identification scheme. In increasing numbers, taxonomists have added another value to their work. They have striven to make their taxonomy reflect evolutionary relationships, which is wonderful for almost everyone who needs avail himself of their product. Perhaps we should be gratified by that and not worry about species except that they be recognizable. For that it is perhaps sufficient that they be a population all the members of which are more closely related to each other than any are to any organism not a member of the group. Such a position has the virtue that it applies equally well to both haploid and diploid organisms.

JS: Walter has basically said it all, with (in my opinion) the key phrase being: “Thus, the use of the term species rests primarily on it’s utility as an identifier of organisms and should be accepted as such”. From the field of trypanosomatids, we only have to look at the mess which insistence on erecting a plethora of species names has caused in Leishmania. At the same time, it could be argued that the recognition of a single species name for Trypanosoma cruzi (T. cruzi) infecting humans was insufficient to reflect genuine differences in strain pathology and ecology. As we have seen with the recent meeting to amend T. cruzi terminology (Anonym, 1999), where the needs of the medical research community dictate, the erection of additional groupings can greatly assist and clarify communication and study. I guess, in the absence of a universal species concept which works equally well for viruses, bacteria, protozoa, elephants, plants, etc. I’m advocating and supporting the concept of a primarily utility-tailed biological species concept, underpinned with robust phylogenetic evidence. [Am I asking/expecting too much?].

SB: Some elements of thought: I am not sure that nowadays, many new species would be defined on the basis of their medical properties rather than on their phylogenetic positioning. This was true in the past, as phylogenetic approaches (such as those based on numerical taxonomy or 16S rRNA sequencing) were not commonly used. I guess that there is now a rather large consensus for the so-called polyphasic approach to the definition of species, where the concordance among both genotypic and phenotypic characters is considered before describing a new species. Although it is an operational species concept rather than a theoretical one, congruence of many markers generally equates to a monophyly of species, so in practice it is very close to a phylogenetic species concept. Apparently the view prevails that the nature of bacterial species is not simply a matter of philosophical discourse, but one of real practical significance. In that respect, non-respect of the phylogenetic species concept based on strong medical reasons (for example, Shigella
species which are particular clones within the *Escherichia coli* taxospecies) makes sense. But again, I guess that nowadays such distortions of the taxonomy compared to the phylogeny would not be accepted for new species, based on the principles of polyphasic taxonomy (we can suspect that *E. coli* O157:H7 would have been named a different species before the generalisation of numerical taxonomy and polyphasic approaches, but species distinction has not been given). The advantage of a species definition based on phylogenetic relationships is that it yields a stable nomenclature and is universal. Possibly a still debatable aspect is the cut-off level at which species are defined (70% DNA:DNA reassociation and five degrees DTm). Indeed, within many genomic species, clear-cut, phylogenetically well separated sequence clusters can be identified, based on the sequencing of housekeeping genes; and these clusters often deserve recognition as they correspond to ecologically separated populations. This is the case of three sequence clusters we identified within the genomic species *Klebsiella pneumoniae*, an important nosocomial opportunistic pathogen. All belong to the genomic species *K. pneumoniae*, and all contain a considerable amount of diversity, but there are clear differences in antibiotic susceptibility between them.

4. Fourth question: According to you, in the last 10 years, what have been the main milestones in molecular phylogenetics, either in general (more a question for Walter, although Jamie and Sylvain are welcome to comment) or for the specific case of pathogens (more a question for Jamie and Sylvain, although Walter is welcome to comment)?

**WF:** I do not think there were any major conceptual breakthroughs in the last 10 years. The sequencing of so many genomes and with more to come, is a monumental achievement although it is mostly technological improvements in a method already well known, however substantial the product may be. But that very largess has opened the eyes of many scientists to the need for software to answer the biologists’ questions. Mathematicians, physicists and computer people are all scrambling to get in on the ground floor of a new burgeoning growth industry, bioinformatics. I do expect to see many great advances with in the next 10 years however. Given so many miners for this intellectual gold rush, there are bound to be very substantial advances.

**SB:** I would agree with Walter in the way that no particular conceptual change in our way to reconstruct phylogenies has appeared during the last 10 years, although refinements have been brought by the wider use of maximum likelihood versus parsimony or distance methods. To my opinion major recent milestones derived from molecular phylogenetics in the development of our knowledge of the microbial world diversity are: (i) the description of the three major domains of life (Eucarya, Bacteria, Archaea) by Carl Woese; (ii) the discovery that lateral (horizontal) transfer of genes among distantly related microbial lineages (including distinct major subdivisions) has been common in the evolutionary history of living forms, leaving us with the view that bacteria share a common microbial gene pool; this aspect includes major improvements of our understanding of the respective roles of clonality, genetic exchange and ecological specialisation in shaping bacterial population structure and on the evolution of important genes such as antibiotic resistance and virulence genes; (iii) the discovery of a hitherto unsuspected and tremendous biodiversity of non-cultivable microbial forms (much more numerous and diverse than the cultivable ones) by the use of direct ribosomal RNA amplification and sequencing from natural samples. Possibly a still awaited major milestone is the development of standardisation of molecular phylogeny data below the species level and the integration of data into library typing systems useful to catalogue strains biodiversity for the purposes of molecular epidemiology, biotechnological resources, and evolutionary studies. For this purpose, DNA chips and binary typing systems will probably play a major role in the near future.

**JS:** Nothing much to add. Just to say that I think we are now entering a possibly more stable period of molecular analysis where we understand the theoretical basis of the tools which we’re using.

In my own field, molecular epidemiology of trypanosomiasis, the last 10 years have seen the demise of isoenzymes (and their — for phylogenetic purposes — inherent homoplasy), the birth and (thankfully) rationalisation of RAPD analysis as researchers finally realised that just because a technique used DNA it wasn’t necessarily foolproof (and that it is actually quite a good idea to know the nature of the variation which one is looking at!), and the increasing use of microsatellite and sequencing analysis. As both Walter and Sylvain have stated however, throughout this period the concepts remained largely the same to identify alleles and to reconstruct relationships while the tools improved significantly.

**MT:** You speak as though isoenzymes were just artefacts. Do you forget that for phylogenetic studies under and around the species level, they provided us with 90% of the information yielded by more sophisticated markers? In the case of *T. cruzi*, the agent of Chagas’ disease, RAPDs, microsatellites and sequencing were just able to confirm the existence of (at least) 6 discrete phylogenetic lineages previously clearly shown by MLEE. The limitation of isoenzymes was for branching the upper levels of phylogenetic divergence, due to saturation of their level of resolution. RAPDs and microsatellites undergo exactly the same limitation. It is a matter of too fast a molecular clock. When *Leishmania* are concerned, all species were adequately identified by MLEE, with convenient specific markers. When we just want to identify a strain, we do not waste our money with molecular markers: generally one isoenzyme locus is enough (in other words: is synapomorphic). The “small” species of *Leishmania*, cherished by the “leishmaniacs”
are indeed right in the power of resolution of isoenzymes. In the bacteria studied by us (E. coli, Helicobacter pylori, Neisseria meningitidis), MLEE and molecular markers gave quite parallel results. The “demise of isoenzymes” you are talking about is rather a matter of fashion. Homoplasy is a drawback of MLEE, it is not a lethal one.

JS: Final point.

There is now a growing perception that the soon-to-be completed genome analysis of several major pathogens will lead to significant advances in combating the disease which they cause, to which I would answer: ‘potentially’. As Alan Fairlamb pointed out at the recent Royal Society parasite genome meeting, several perfectly good drugs for the treatment of trypanosomatid diseases have recently been removed from (Nifurtimox) or threatened with removal from (Efloritine) production. Thus, while more information may allow us to identify new targets for intervention and to design new drugs, ultimately it will be economics which determines whether or not this knowledge will be used. In the case of many globally important diseases caused by parasites (e.g. malaria, Chagas disease) the economic benefits of producing new (or old) drugs may never be cost-effective. . .

References


Lal, A., Tibayrenc, M., 1997. The first international workshop on molecular epidemiology and evolutionary genetics of pathogenic microorganisms. Emerging Infect. Diseases 3, 244.


W. Fitch (USA) is a senior evolutionist who has made well-known eminent contributions to Molecular Phylogenetics. According to himself, his experience with pathogens is limited. He works at the University of California at Irvine.

S. Brisse (France) is junior researcher, although already well-experienced in the field of phylogenetics. He is deeply involved in the study of pathogens, parasitic protozoa and bacteria. He is at Utrecht University, The Netherlands.

J. Stevens (UK) is junior researcher, although already well-experienced in the field of phylogenetics. He is deeply involved in the study of pathogens, parasitic protozoa and bacteria. He is at the University of Exeter, UK.

---