Stud. Hist. Phil. Biol. & Biomed. Sci. 39 (2008) 451–468

Contents lists available at ScienceDirect

Stud. Hist. Phil. Biol. & Biomed. Sci.

journal homepage: www.elsevier.com/locate/shpsc

History, objectivity, and the construction of molecular phylogenies

Edna Suárez-Díaz [a], Victor H. Anaya-Muñoz [b]

[a] National University of Mexico and Max Planck Institute for the History of Science, Facultad de Ciencias, Universidad Nacional Autónoma de México, Circuito Exterior Ciudad Universitaria, Coyoacán, DF 04510, México
[b] Institute for Theoretical Biology, Humboldt Universtiät zu Berlin, Invalidenstraße 43, Berlin 10115, Germany

ARTICLE INFO

Article history:
Received 18 November 2007
Received in revised form 10 July 2008

Keywords:
Molecular phytogenies
Objectivity
Character weighting
Quantification
Statistical Methods
Automation

ABSTRACT

Despite the promises made by molecular evolutionists since the early 1960s that phylogenies would be readily reconstructed using molecular data, the construction of molecular phylogenies has both retained many methodological problems of the past and brought up new ones of considerable epistemic relevance. The field is driven not only by changes in knowledge about the processes of molecular evolution, but also by an ever-present *methodological anxiety* manifested in the constant search for an increased *objectivity*—or in its converse, the avoidance of *subjectivity*.

This paper offers an exhaustive account of the methodological and conceptual difficulties embedded in each of the steps required to elaborate molecular phytogenies. The authors adopt a historical perspective on the field in order to follow the development of practices that seek to increase the objectivity of their methods and representations. These include the adoption and development of explicit criteria for evaluation of evidence, and of procedures associated with methods of statistical inference, quantification and automation. All these are linked to an increasing use of computers in research since the mid 1960s. We will show that the practices of objectivity described are highly dependent on the problems and tools of molecular phylogenetics.

When citing this paper, please use the full journal title *Studies in History and Philosophy of Biological and Biomedical Sciences*

> The largest concentration of information present in an organism, and perhaps also the largest amount of information, and the only organically transmissible information, is in its semantides. [S]emantides are potentially the most informative taxonomic characters and not ... just one type of characters among other, equivalent types ... (Zuckerkandl & Pauling, 1965, p. 98)

> Molecular phylogenies work not because DNA is 'better,' more real, or more basic than morphology, but simply because the items of a DNA program are sufficiently numerous and independent to ensure that degrees of simple matching accurately measure homology ... (Gould, 1986, p. 68)

1. Introduction

Reconstructing the history of life has been one of the most demanding activities of biologists and natural historians since the nineteenth century. After Darwin, the enterprise was definitively oriented towards the systematic elaboration of plant and animal phylogenies (represented in trees) that aimed to show the pattern of speciation. But as any scientist familiar with historical reconstruction knows, this is not an easy task. All kinds of difficulties arise, whether these are related to the lack of evidence from the past, to the nature of the historical records or, equally importantly, to the myriad conceptual, theoretical and methodological obstacles and decisions that lie between the available empirical evidence and the reconstruction of a pattern. In the case of biological evolution, some of the issues that stand in the way of the reconstruction of phylogenetic relationships are linked to the very different characterizations of *homology*, which has been said to be 'the central concept of *all* biology' (Hall, 1994, p. 1), or—at least implicitly—the root of all difficulties (Felsenstein, 2001). Although different approaches to homology may lie beneath many of the methodological differences among rival schools of systematics (and therefore of molecular phylogenetics), we are more concerned with the practices of objectivity—specifically concerning a

E-mail addresses: emsd@hp.fciencias.unam.mx (E. Suárez-Díaz), v.anaya@biologie.hu-berlin.de (V. H. Anaya-Muñoz).

1369-8486/$ - see front matter © 2008 Published by Elsevier Ltd.
doi:10.1016/j.shpsc.2008.09.002

historical science—that are shared by the members of this community, and not with the debates around the concept of homology or the different philosophies behind it.

Despite the promises made by molecular evolutionists since the early 1960s that phytogenies would be readily reconstructed using molecular data (Suárez, 2007), the construction of molecular phytogenies has both retained many methodological problems of the past and brought up new ones of considerable epistemic relevance. Although heavily equipped with new technologies and different statistical tools that enable the processing of very large amounts of sequencing data—in particular, a heavy reliance on computers and software—molecular phylogenetics does not escape the difficulties of *reconstructing the past* (in Elliot Sober's terms; Sober, 1988a). The field is driven not only by changes in knowledge about the processes of molecular evolution, but also by an ever-present *methodological anxiety* manifested in the constant search for an increased *objectivity*—or in its converse: the avoidance of *subjectivity*. Broadly speaking, scientists in this field are constantly confronted with the goal of having more objective sets of methods (of statistical inference) capable of delivering more objective representations of the phylogenetic relations between species.[1]

The search for objectivity is embodied in the growing complexity of the practices and tools used in this field, including the extended communication networks supported by computer networks and shared databases. This is the outcome of four decades of applying and developing the techniques and research strategies of molecular biology in problems of evolutionary biology, which has resulted in the massive production of amino acid and nucleotide sequences and their statistical analysis for the construction of phylogenies. The advent of automated sequencing of DNA and the overwhelming impulse of the Human Genome Project have greatly accelerated this process.

We begin the paper with a brief presentation of the debate between organism-centered and molecular evolutionists that took place during the 1960s, concerning what type of characters—molecular or morphological—constituted the best evidence for the reconstruction of phylogenetic relations between species. We focus on the different ideals and values involved in the contrasting approaches of molecular and organismal biologists, and establish their connections with different ideals of objectivity. Our long Section Three describes the methodological debates that characterize the field of molecular phylogenetics and points to the many problems and obstacles that exist between molecular data and knowledge of the physical processes of molecular evolution, on the one hand, and between molecular data and phylogenetic representations, on the other. We have sought to exhibit, without too many technicalities, the methodological issues at play in each and every step of the molecular evolutionists' job. At the same time we have tried to highlight the scientists' attempts to avoid subjectivity. Section Four establishes connections between the case of molecular evolution and broader reflections concerning practices of objectivity, quantification and automation in science. Our purpose is to present a challenging set of problems to be addressed by a philosophy of science interested in scientific practices. Finally, in the concluding remarks we revisit the questions posed in the early debates between classical and molecular evolutionists in light of the specific difficulties posed by the study of *historical* processes: In what sense is molecular evidence *cleaner* or *more direct* than morphological evidence to reconstruct the pattern of history? Has the ideal of objectivity, within the field of phylogenetics, been transformed at the onset of a molecular approach? If so, in what direction has this occurred?

2. The use of molecules as *characters* in evolutionary biology

The rise of the alternative schools of pheneticism and cladism and the ensuing *systematist wars* in the 1960s and 1970s challenged the entire edifice of evolutionary systematics to its methodological foundations (Hull, 1988; Vernon, 1988, 1993). The core of the attacks by the new schools was the idea that taxonomic groupings in classical evolutionary systematics mixed measures of similarity or *resemblance* with non-empirical hypotheses of phylogenetic relationships. Previous ideas or theories held by systematists were said to permeate the 'weighting' of different characters in the construction of evolutionary trees. Moreover, the critiques included an attack on the idea—defended by George G. Simpson and Ernst Mayr—that the construction of classifications required the intervention of subjectivity, of the intuitive judgment of the expert, and even artistic and pragmatic criteria (Hagen, 2001, 2003).

Just as pheneticists and cladists were raising their voices against the idiosyncratic methods of evolutionary taxonomists, a new scientific arena was being established. In the early 1960s molecular biologists, biochemists and biophysicists entered the field of evolutionary biology.[2] Ernst Mayr, Theodosius Dobzhansky and George G. Simpson responded critically to what they thought was an intervention of molecular biologists in a research field about which they knew almost nothing. The so-called architects of the Evolutionary Synthesis made a concerted effort to stress the complex relations of traits at the individual (organismal) level and the action of natural selection on biological populations, and resisted attempts to concentrate the efforts of research at the molecular level. We will not take up these issues here, as they have been already addressed elsewhere by historians of science (Dietrich, 1998; Hagen, 1999; Aronson, 2002; Suárez, 2007; Sommer, 2008).

As regards phylogenetic inferences and classification, classical biologists defended the view that the best characters to study were at the level of the individual organism, be they morphological or functional traits. But the quality (and quantity) of these characters was very much contested. The interpretation of fossils, and of morphological and functional evidence of all sorts required *judgment*, sometimes even *intuition* and *artistic* taste. Also, it required scientists to single out what counted as a character, and idiosyncratic intervention of the expert in the 'weighting of characters' was often inevitable (Hagen, 2001,2003). It was common, and remains so until this day in a few areas, for a given person to be recognized as *the* scientific authority of a particular taxonomic group.[3]

[1] Hagen (2003) has also addressed the role of objectivity and subjectivity in systematic research, and connected it to the introduction of computers and statistical tools in the field. What we want to stress in this paper is that the search for 'objective' tools permeates all the steps in phylogenetic reconstruction, and not only the construction of trees. For some systematists, having *objective* representations of phylogenetic relations means that taxonomies should show overall similarities among species (for example, Felsenstein, 2002); for others, it means the depiction of homologues (Williams & Ebach, 2005, is a recent example). Both of these positions re-enact differences that have been present throughout the long history of homology. However, these differences are of secondary importance in the context of the development of bioinformatics. On homology see Bock & Cardew (1999), Hall (1999), and Laubnicher (2002).

[2] The connection between this scientific arena and the struggle of evolutionary systematists against pheneticism and cladism is something that we do not address in this paper. Felsenstein (2004), Ch. 10, mentions some connections between the development of the molecular approach to evolution and that of the rival schools of systematics. Hagen (2001) argues that the analysis of very large amounts of data and the introduction of computers in systematic research in the 1960s connects the research of systematists, molecular evolutionists and population geneticists. Like Hagen, we want to emphasize the constraining force of automation and quantification on all parties involved; the introduction of these tools obscured more traditional debates (for instance, on homology) and intensified the differences between new contending parties (see below).

[3] Examples where certain taxonomists are considered to be the personal authority on a given biological group abound in the literature. Talking about the Snow Museum of Natural History in Lawrence, Kansas, David Hull says, for instance: 'The leading member of its staff and the biggest bee man in the world was (and continues to be) C. D. Michener' (Hull, 1988, p. 118). This raises, clearly, the question of the role of hierarchies, *the* social structure of systematics and the search for respectability (Vernon, 1993).

A couple of sentences extracted from a review of the state of human and primate classification by paleontologist George G. Simpson illustrate the critics' points of attack and the florid, subjective, and personalized language that characterized this area at the time:

> the significance of differences between any two specimens has almost invariably come to be enormously exaggerated by one authority or another in this field. Here the fault is not so much lack of taxonomic grammar as lack of taxonomic common sense or experience. Many fossil hominids have been described and named by workers with no other experience in taxonomy. They have inevitably lacked the sense of balance and the interpretative skill of zoologists who have worked extensively on larger groups of animals. It must, however, be sadly noted that even broadly equipped zoologists often seem to lose their judgment if they work on hominids. Here factors of prestige, of personal involvement, of emotional investment rarely fail to affect the fully human scientist, although they hardly trouble the workers on, say, angleworms or dung beetles. (Simpson, 1964, p. 7)

Simpson's quotation is quite explicit in acknowledging the role of subjectivity and individual judgment in the taxonomy of primates (certainly, as he declares, a special field of research), and is revealing of the methodological attitudes and procedures that would be the focus of the critiques of later systematists. Simpson argues that the taxonomist should be equipped with professional skills such as 'taxonomic common sense' or 'experience'. Many could agree that such skills are transmitted by disciplinary education, but Simpson also referred to a different, more emotional investment of the taxonomist. In particular, he admitted that the specialist was more prone to a loss of judgment when working on a particular group, namely, hominids. To avoid such biases, Simpson's recommendation was to cultivate a 'sense of balance and interpretative skill'.

More specifically, Simpson reflected on the relation between phylogenetic relations and classification: 'Resemblance provides important, but *not the only*, evidence of affinity. Classification can be made consistent with, even though not directly or fully expressive of, evolutionary affinity' (ibid., p. 9). Although other taxonomists might have agreed with that statement (pheneticists, for instance), Simpson did not clarify what the criteria or the rules for assigning classification and measuring affinity were. Even as he declared that statistical methods should be used on morphological traits in order to measure resemblance, nowhere in his text did he describe or use such methods. It is important to emphasize this absence given that Simpson, co-author of one of the first textbooks on the subject (Simpson & Roe, 1939; Simpson et al. 1960; see Hagen, 2003), was well aware of the importance of quantitative methods for the biologist—and in particular for the taxonomist. Thus, a member of the pheneticist school stated: 'Numerical taxonomy uses statistical methods to form groups whereas traditional (evolutionary) taxonomy only uses them to discriminate more precisely between groups already perceived' (quoted by Hull, 1988, p. 111).

Apparently unaware of the methodological debates affecting systematics—at least at the beginning—the new molecular systematists offered a battery of arguments that favored molecular characters as evidence for phylogenetic relations. Support for their arguments came from the discrete or quantized nature of amino acid substitutions in protein chains, whose molecular sequences were the only available ones at the time. In their view, this property of molecules (being linear or one-dimensional chains of individual residues) would easily enable measurement of similarities or differences, opening the door to *quantitative* comparisons between homologous peptide chains.[4] They thought that, eventually, this approach would lead to the direct comparison of DNA segments composed by long chains of nucleotide residues (something that looked far into the future at the time). In the words of Emile Zuckerkandl, one of the most prominent spokesmen of the new field, molecular characters were, in contrast to morphological characters, 'cleaner' or 'more direct' evidence of evolution (Zuckerkandl, 1964, p. 260).

Perhaps the most radical early advocate of the value of molecular over morphological characters was biochemist Emmanuel Margoliash. Margoliash's ultimate goal was to reconstruct the complete phylogeny of living beings using only *one* protein. To do so he chose Cytochrome c, a protein involved in electron transfer that has been present along the entire evolutionary history of organisms. He was also convinced that a correct interpretation of the similarities and differences of proteins could be achieved solely by means of *statistical* analysis, and argued that *homology* could be determined exclusively by statistical criteria (Margoliash, 1963, p. 677). Simpson reacted forcefully against the reduction of taxonomic criteria and concepts to statistical standards (Simpson, 1964; Aronson, 2002). To make matters worse, a profound misunderstanding pervaded the dialogue between classical and molecular evolutionists: while Simpson, Mayr and Dobzhansky thought of a protein molecule as *a single character*, and one that was 'not available' to the action of natural selection, the molecular biologists' view was that a single protein molecule contained as many character states (dozens or even hundreds) as the number of residues that composed it.

The conviction that molecular data were better suitable for a quantitative and non-idiosyncratic approach was reinforced by Walter Fitch's pioneering work in computer programming. Although other scientists (Edwards & Cavalli-Sforza, 1964, and Eck & Dayhoff, 1966) also used computers and pioneered the construction of analytical tools to reconstruct phylogenies, Fitch's approach was by far the most influential on the development of further methods.[5] Using Margoliash's data, consisting of the twenty known amino acid sequences of Cytochrome *c*, Fitch was able to construct a molecular phylogeny based on a measure of similarity that he defined as *minimum* (mutational) *distance* between two cytochromes (Fitch & Margoliash, 1967, 1968). The authors profited from 'a computer making a pairwise comparison of homologous amino acids' (Fitch & Margoliash, 1967, p. 280). Fitch's computational approach, along with many parallel developments in the life sciences, illustrates the entrance, in the mid 1960s, of the first generation of mini-computers for doing research in university laboratories (mostly IBMs of the 360 Series, and the common DEC PDP-8, but also the common LINC). As we will see, the introduction of these tools reconfigured the entire field of phylogenetics research.[6]

[4] We are not forgetting the attempts of Willi Hennig and the pheneticist school to obtain a measure of *overall similarity* between species based primarily on morphological characters. However, one of the most common critiques of the pheneticists' methodology centered on their idiosyncratic mixing and assignment of values to different kinds of traits (morphological, biogeographical, paleontological, and so on). The decline of 'morphological' pheneticism illustrates the fact that organismal characters are not as prone to statistical or quantitative manipulation as are molecular characters.

[5] On Cavalli-Sforza and Dayhoff, see Hagen (2001); a more detailed account of Dayhoff is given by Strasser (2006, 2008). While Edwards and Cavalli-Sforza used serological data, Eck and Dayhoff, like Fitch, used protein sequences data; the question is thus why Dayhoff has not received the proper credit or response to her earlier work. Bruno Strasser has pointed out to the fact that the National Biomedical Research Foundation, where she worked, was a *sui generis* private research institution, and that her gender might have played a role.

[6] Most probably, Fitch used an IBM machine of the 360 Series, bought by the University of Wisconsin (e-mail communication 6 June 2008). On the history of computers, see Ceruzzi (1999) and Shurkin (1985). The introduction of computers in biological research is a scarcely explored field, see Lenoir (1999), Hagen (2001), and November (2006).

While there had been previous evolutionary analyses of hemoglobin (Zuckerkandl & Pauling, 1962) and of Cytochrome *c* (Margoliash, 1963) that attempted a basic semi-quantitative analysis, and several other attempts to construct phytogenies using approximate distance measures of molecular data (such as the proportion of DNA hybridization in Bolton & Britten, 1963, or serum affinities in Goodman, 1960), the idea of a quantitative phylogenetic analysis of large sets of data came to fruition with the use of computers and the development of computer programs. As the number of molecular characters (protein and DNA sequences) available for comparison increased in the following decades, the development of computational tools became an integral part of the practices of phylogenetics. Very soon, however, it became clear that the introduction of the new tools did not put an end to the many methodological problems that continued to arise.

3. Problems underlying the construction of phylogenies

The general optimism of the early 1960s notwithstanding, some molecular evolutionists had already envisioned several difficulties that their field was to encounter in the following decades. These included the calculation of the number of 'actual' mutations that took place at a given nucleotide position in the DNA molecule (Zuckerkandl & Pauling, 1962; Fitch, 1966; Fitch & Margoliash, 1967), as well as 'the difficulties in interpretation at both extremes of the scale of comparison, namely in comparisons bearing on very closely or on very distantly related species' (Zuckerkandl, 1964, p. 255).

As we will show in this section, these methodological complications continue to present themselves today. What the early practitioners of molecular phylogenetics did not foresee was that the growing number of methodological problems would become one of the main subject matters of their field: procedural tribulations have stood in the way of reconstructing phylogenies in spite of the availability of molecular data. In prevalent accounts of phylogenetics, almost all differences between rival schools have been reduced to the development and use of statistical methods reflecting different philosophies of classification (parsimony *versus* likelihood) embedded in the available software utilities, and deeper disjunctions about homology, the role of similarity in systematics and, broadly, the difficulties posed by the reconstruction of history, have been relegated to second place. In this section we will scrutinise the procedures and decisions a molecular phylogeneticist needs to make when building a phytogeny, and we will direct our attention to the issues where the search for objectivity plays a distinctive role.

3.1. Selection of characters

Various types of molecular data can be treated as characters or traits: nucleotide and amino acid sequences, electrophoretic 'fingerprinting', antigenic serum affinities (immunological distance) or proportions of DNA hybridization between two species. The kind of information that can be extracted from the reconstructed phylogenies depends on the type of data used in its construction. Today, it is generally assumed that DNA, RNA or protein sequences are characters, and these include as many character-states as the number of residues or bases that compose each molecule. Molecular sequences (more than serological affinities and hybridization proportions) make the application of quantitative-statistical approaches possible and now constitute the most commonly used characters; in what follows we will refer only to these.[7]

The first step is to determine whether a chosen gene or its corresponding protein is better suited as evidence for the phylogeny of the species, or for the evolution of that particular gene or protein. Highly conserved genes, proteins or domains are thought to be useful for exploring the relationships between long lineages, whereas highly variable—and recent—molecular characters are more suitable for reconstructing phylogenies of closely related species or for discerning relationships between populations or between organisms. A combination of these traits is also common; highly conserved functional proteins or genes, in addition to one or several highly variable domains (or vice versa), can also be used. The phylogenetic relations obtained depend on the sequences selected, but scientists usually do not possess knowledge of the sequences prior to their phylogenetic investigations; on the contrary, knowledge about the rates of substitution of the molecules selected might be one of the results of their research.

In the early 1960s molecular evolutionists depended for the most part on two different molecules, and the very few sequences available of them. Emile Zuckerkandl and Linus Pauling, for instance, relied on the sequences of globins and hemoglobins that had been worked out at Caltech and in other laboratories and medical institutions around the world (Morgan, 1998; Chadarevian, 1998; Suárez, 2007). Emanuel Margoliash, for his part, decided to sequence the Cytochrome *c* molecules of as many species as possible (Margoliash, 1963; see Fitch, 1988 for a personal account). As more sequences became available (ribonucleases, ATPases, SSU RNAr's, etc.), more details about the processes of molecular evolution became known and informed the development of methods. Today, the availability of enormous sequence databases (product of various large-scale genome sequencing projects), provides huge amounts of raw material, but also presents many difficult decisions to make.

The importance of making a good choice of character can be illustrated with Carl R. Woese's long research on the early evolution of life. In 1977, Woese and George E. Fox used the genes of the small subunit ribosomal RNA (SSU rRNA) with the intention of establishing a basic phylogeny for the history of life. According to them, this molecule possessed the characteristics that made it an excellent marker for this purpose:

> To determine relationships covering the entire spectrum of extant living systems, one optimally needs a molecule of appropriately broad distribution. None of the readily characterized proteins fits this requirement. However, ribosomal RNA isolated does. It is a component of all self-replicating systems; it is readily isolated; and its sequence changes but slowly with time—permitting the detection of relatedness among very distant species. (Woese & Fox, 1977, p. 5088)

The phylogeny that resulted from these sequences had an extraordinary impact on biologists' scenarios of the diversity of living systems. According to Woese and Fox (ibid.), the major division between organisms was no longer between Prokaryotes and Eukaryotes, but between three mayor domains: *Eukarya*, *Archaeobacteria* and *Eubacteria*. Apart from the obvious transformation of

[7] The reason sequences have become prevalent is a historical and epistemic question to be addressed. Their easy translation into digital data, the possibility of distinguishing between ancestral and derived character states, and the very large amounts of available data are part of the answer; we do not, however, deal with this here. Nor do we go into the details of the experimental practices and tools that underlie the gathering of DNA and protein sequences. This is not a trivial matter, however, since sequences or databases may have different degrees of reliability. By contrast, in the late 1980s there were several attempts to develop statistical methods for critically evaluating the strength of phylogenetic inferences obtained with DNA hybridization. Examples of this type of critical research include Templeton (1985, 1986), Felsenstein (1987), Marks et al. (1988), Sibley et al. (1990), and Dickerman (1991).

the biological landscape that this study brought about, it exemplifies why the results obtained strongly depend on the molecules chosen.[8]

When choosing a molecular character, the sequences to be compared should be *orthologous*,[9] and represent as many different species as possible. Once again, a methodological question arises. Since the mid 1960s statistical criteria have been used extensively to determine homology (Margoliash, 1963), and this practice has been criticized by molecular and organismal biologists alike.[10] Using the statistical criterion, for instance, may lead to the erroneous conclusion that sequences that are duplicated and subject to mutations will appear as homologous. Thus, some biologists have pointed out that molecular evidence makes little use of the distinction between analogy and homology, and that homology is (spuriously) established only after the Comparison has been made and the phylogeny has been inferred.[11] According to the more statistically inclined, the only 'observable' property is similarity, but similarity may well be the result of analogy or homology.[12]

3.2. Sequence alignments

Once a scientist has chosen a series of DNA or protein sequences, what follows is an alignment and comparison (these go together by the name of *sequence alignment*) between every single sequence of interest and the rest of the sequences. 'A sequence alignment is designed to exhibit the evolutionary correspondence between different sequences' (Thorne et al., 1991, p. 114). The data obtained in the comparison has two functions: to determine if the group of sequences is indeed homologous, and to quantify the extent of similarities or differences between them.

The statistical criterion of homology has dominated the field in spite of the early opposition of classical biologists. As Winter et al. observed: 'We propose that the word be taken to connote the occurrence of a degree of structural similarity among proteins greater than might be anticipated by chance alone [because] it would be impossible to conclude with *certainty* that two proteins are homologous' (Winter et al., 1968, p. 1433). Conventionally, a rule or criterion of thirty percent or higher degree of structural similarity (that is, having the same type of residues in the same position) between a pair of molecules has been taken to indicate homology. According to some, homology should be assumed only as a 'working hypothesis' (see Fitch, 1970). Nevertheless many molecular evolutionists understand homology as nothing more than a statistical criterion (see Goodman, 1960, 1996; Margoliash, 1963), and the statistical definition has prevailed despite Fitch's methodological reservations.

Having decided that a series of sequences are homologous, the alignment and comparison of molecules leads to even more methodological rules and decisions. In Fitch and Margoliash's paper a 'distance value' was given to every single mutation according to a table 'which gives the minimum number of nucleotides required to convert the coding from one amino acid to the other' (Fitch & Margoliash, 1967, p. 280). This table was a result of Fitch's previous work on the elucidation of the genetic code (Fitch, 1966) and it was an early instance of what would later become a more sensitive issue: the weighting of different types of substitutions (see below). At this point, however, it suffices to note that Fitch was well aware that 'not all differences counted the same'; he addressed this problem by making explicit the different values of nucleotide substitution for each of the amino acid differences.

Fitch and Margoliash's paper did not deal with the problem of aligning sequences of different length. However, ever since the publication of this pioneering paper, molecular phylogeneticsts have evaluated competing alignments according to a standard criterion that Joe Felsenstein and his colleagues made explicit: 'It will frequently be desired to find [the] alignment of two nucleic acid [or protein] sequences which maximises the number of positions that are identical' (Felsenstein et al. 1982, p. 133). How to achieve this goal continues to be one of the most debated issues in molecular phylogenetics.

The introduction of alignment software in the 1980s, made possible by larger computing capabilities, allowed more sophisticated answers but did not solve the problem. Moreover, as late as the 1990s computers were not always used to do alignments. Complaining of this, Thorne et al. argued: 'It is possible, and among some researchers, popular to align sequences by the eyeball. The eyeball technique is time-consuming, tedious, and irreproducible ... Computer-aided sequence alignment does not posses these disadvantages of the eyeball technique' (Thorne et al. 1991, p. 114). The implication is that relying on machines (*automating and mechanizing the procedure of alignment*) is more efficient and dependable than relying on human beings. The objectivity of sequence alignment is thus intertwined with the mechanization and automation of the alignment procedure. But the use of computers only displaces the methodological problem. Henikoff and Henikoff describe the current view on the different alignment methods:

> There are several different types of alignments: Global alignments of pairs of proteins related by common ancestry throughout their lengths, local alignments involving related segments of proteins, multiple alignments of members of protein families, and alignments made during data base search to detect homology. In each case, competing alignments are evaluated by using a scoring scheme for estimating similarity. (Henikoff & Henikoff, 1992, p. 10915)

The specialized literature also reports semi-global alignment methods. Each of these methods has been designed—and is believed to be more adequate—for different cases and with different purposes in mind. Thus, global alignment methods include all the characters in the compared sequences and are considered helpful when dealing with closely related sequences; however, they are unable to detect some important evolutionary events, such as *domain shuffling*.[13] Local alignment approaches, on the other hand, compare fractions of the sequences under study in order to find similar (or identical) regions even when they are not contiguous.[14] The

[8] See Sapp (2005) for a historical account of the impact of these events.

[9] Orthology is the relationship between any two characters that share a common ancestor and have diverged by descent, that is to say, are 'true' homologues (see Fitch, 1970, 1971, 2000).

[10] See Simpson (1964), Fitch (1970), Li & Graur (1991). For more on the concept of homology in molecular phylogenetics, see Zuckerkandl & Pauling (1965), Abouheif (1997), Fitch (2000), and note 1 above on homology.

[11] We owe this comment to Jon Marks.

[12] As pointed out before, the differences with respect to the concept and determination of homology (ancestry) have been reduced to differences about statistical methods (see Felsenstein, 2001).

[13] Domain shuffling refers to the 'construction' of a new protein by bringing together domains that were located in other positions. Global alignment methods usually perform the Needleman-Wunsch algorithm (Needleman & Wunsch, 1970), and can be implemented by the Needl or the Matcher programs in the EMBOSS package. According to Feng et al., 'This algorithm obtains a maximum score based on rewards for similarities or identities between two sequences and penalties for deletions in either sequence' (Feng et al., 1985, p. 112).

[14] These methods are implemented by the popular BLAST software, and in different versions of FASTA and the Water and Stretcher programs included in the EMBOSS package. Usually they are based on the dynamic-programming algorithm developed by Smith and Waterman in 1981.

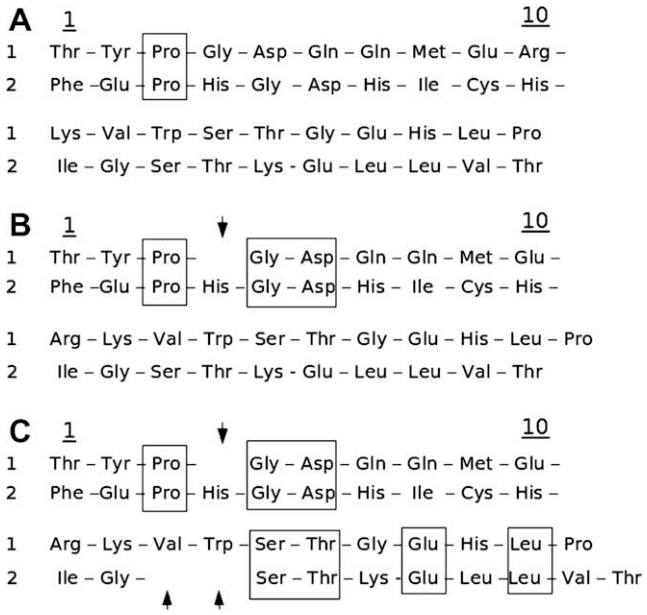

Fig. 1. (A) Alignment with no gaps added; only one pair of identical residues is found (framed). In (B) one gap was introduced in sequence 1 (shown by an arrow), and two more residues are then estimated to be identical (for a total of 3). In (C) a two-residues gap was introduced in sequence 2, yielding four extra identical residues, aligned (Modified from Jukes & Cantor, 1969, p. 98).

semi-global approach is a hybrid methodology suitable for very specific uses, that is, cases in which neither the global nor the local methods alone are appropriate (Brudno et al., 2003). This hybrid methodology allows scientists to find large identical or highly similar regions between two sequences, even when the regions are located on opposite sides of the molecules—a situation present in domain shuffling scenarios. Finally, it is also possible to make multiple sequence comparisons, which are extensions of the pair wise approaches and are used to identify similar regions between several sequences.[15]

The question is *which one* of these methods to use, since in many cases it cannot be asserted whether domain shuffling has taken place, either because there are no experimental results available or because those results were assigned different degrees of reliability. Moreover, this phenomenon might be detected only *after* the comparison between sequences has been made. The problem, then, is how to evaluate competing alignments methods, and then to evaluate the resulting alignments according to the standard criterion that maximizes the number of identical positions between sequences.

To meet the maximizing criterion the scientist may incorporate *gaps* or *indels*.[16] Fig. 1 shows the alignment of two random protein sequences and the effect of introducing gaps in their apparent similarity. As we can appreciate in the example, the introduction of three gaps (Fig. 1C) is the best way to maximize similarity, but this does not necessarily mean that the two proteins are homologous. In fact, homology is unlikely given that the sequences were chosen randomly by Jukes and Cantor (1969) in order to show the perils of introducing gaps.

Dealing with gaps is one of the most controversial tasks in molecular phylogenetics, and it has been the subject of some of the most heated discussions regarding the intervention of subjectivity (see for instance Felsenstein et al., 1982; Fitch & Smith, 1983; Thome et al., 1991). At one extreme of this practice, the scientist introduces as many gaps as she needs to make any pair of sequences look alike (as in Fig. 1). Such procedure is always in danger of leading to non-meaningful relations, as scientists are quick to notice. The following quotations by Jukes and Cantor illustrate this point:

> Many workers have attempted to improve the apparent homology between two protein chains by the insertions of gaps into one or both of the sequences. Often very large numbers of gaps have been proposed. The difficulty is that by judicious choice of gaps it is always possible to improve the homology between two protein chains. (Jukes & Cantor, 1969, p. 109)

And,

> The indiscriminate placement of a large number of gaps into protein sequences will almost always increase the apparent homology but it may lead to comparisons that are not statistically significant. (Ibid., p. 112)

In order to balance out the introduction of gaps, penalties are imposed. Gaps are weighted on *explicit* quantitative or semi-quantitative criteria or rules that take into account their number, position and/or length. However, it is all too common to read criticisms concerning a given criterion as subject to a particular bias. And moreover, the introduction of gaps is intertwined with the different methods used to measure similarity between sequences.

The similarity between two sequences (for instance, the substitution of a given hydrophobic residue by another hydrophobic residue in a protein sequence) or the difference between two DNA molecules can be translated into different measures of *distance* depending on different methods and criteria. According to Fitch and Smith (1983) there are basically two forms of comparing DNA sequences, both based on the dynamic programming algorithm of Needleman and Wunsch (1970).[17] The first method *minimizes a distance* measure among the sequences, and the second *maximizes a similarity* measure.[18] Because both methods allow the inclusion of gaps during the alignment, their weighting and the penalties assigned enter the picture and affect the measure of similarity.

It must be noted that in general, molecular phylogeneticists have tackled this problem by proposing *explicit* quantitative statistical criteria in an effort to distance themselves from methodological idiosyncrasies or reliance on authority and intuition, both of which they have used to criticize traditional evolutionary systematicists. Those critiques had focused on the fact that assigning a value of similarity or resemblance to morphological characters not only required the exercise of judgment and experience, but also implied personal preferences, and in some cases it was said to be grounded on previous ideas or hypotheses (as noted by Simpson, see quotation in Section 2). But the statistical approach offered by molecular phylogeneticists is not free of methodological con-

[15] A very popular implementation of this method is the family of the *Clustal* Software (Higgins & Sharp, 1988).

[16] Gaps are also known as *indels*, a term that refers to spaces added in an alignment to improve the number of matches. These spaces represent possible *in*sertions or *del*etions. Indels might have also informative properties, as argued by Podlaha et al. (2005) and by Schully & Hellberg (2006).

[17] Here Fitch and Smith refer to the Needleman & Wunsch (1970) and Sellers (1974) algorithms.

[18] As Smith, Waterman and Fitch noticed: 'Both are designed to produce an optimum measure between any two sequences as a function of the minimum number of changes required to convert one into the other . . . There are two major differences between the Needleman-Wunsch and the Sellers algorithms. The most obvious is that the Needleman-Wunsch algorithm results in alignments having a maximum similarity measure, while the Sellers algorithm results in alignments having a minimal distance or metric measure of dissimilarity. The second major difference between them is in their origin. The first was the result of a heuristic approach to an important biological problem, while the second was the result of a search for a rigorous mathematical solution for the problem' (Smith et al., 1981, p. 18).

cerns either, as new problems concerning the methods for weighting gaps surface:

> The weakness of the basic dynamic programming method and its subsequent modifications is the *lack of an objective procedure* to choose the relative weights of gaps and mismatches. The result of this weakness is that researchers are forced to use either of two flawed approaches to obtain an alignment between two sequences. One approach is to arbitrarily choose these weights and then obtain an alignment. If this alignment is *aesthetically pleasing* to the researcher, the process stops. Otherwise, the researcher continues to adjust the weights until an aesthetically pleasing alignment is obtained. Obviously, the *subjective* nature of this approach is not ideal. Another approach is to use the same set of weights for every pair wise alignment. This approach is less subjective than the former approach—only the initial choice of weights is subjective. (Thorne et al., 1991, p. 115; our emphasis).

It is clear that the methodological goal is to diminish the subjectivity of the weighting, or its counterpart, to increase its objectivity. In any case the recommendation is to minimize the acts of *choosing* where the scientist's will intervenes. The objectivity of sequence alignment has also been addressed by using more sophisticated mathematical tools, such as likelihood tests.[19] But even in this case the problem remains: the scientist still needs to choose between different types of substitution matrices when dealing with proteins. Yet another solution for minimizing the subjectivity of gap weighting is to use a *variety* of methods for doing so. The existence and use of all these methods point to the growing instrumentalism in the practice of this field. We will expound on this issue throughout the rest of the paper.

As we have already noted, methodological concerns underlie many important discussions in molecular phylogenetics. Aware of the difficulties of weighting gaps, Fitch and Temple Smith published in 1983 a list of ten recommendations, grouped either as "methodological" or "interpretative, that aimed to help the researcher in her task. Oddly enough, among their criteria they reintroduced the idea that alignments must have *biological meaning*. Implicit in this criterion is the reliance on trained judgement when interpreting biological facts (equally implicit is the recognition that statistical criteria are not enough). Fitch and Temple Smith (champions in the field of bioinformatics) also recommended a certain *familiarity* with the historical process under reconstruction, again appealing to the broader disciplinary experience of the scientist. In response, statistically inclined phylogeneticists claimed that applying more stringent statistical criteria could counteract methodological problems.

3.3. Substitution matrices

Nowhere is the deliberate and explicit inclusion of biological or qualitative considerations clearer than when comparing amino acid substitutions between pairs of protein sequences. Substitution matrices seek to convey the probability that a given amino acid might be substituted by some other amino acid. The resulting parameters are considered to indicate biological correlations due to the chemical properties of amino acids. In this case, a score is given to the substitution probability in order to evaluate the *significance* and *correctness* of the alignment. As Stephen Altschul noticed, 'Specifying an appropriate amino acid substitution matrix is central to protein comparison methods and much effort has been devoted to defining, analysing and refining such matrices' (Altschul, 1991, p. 555).

At the beginning of the 1970s, McLachlan had argued that 'The first step is to set up a measure of similarity for each pair of amino acids ... based on the observed frequencies of amino acid replacements in homologous proteins' (McLachlan, 1972, p. 419). McLachlan recommended taking into consideration not only the 'nude' numbers of differences and similarities between amino acid sequences, but also a set of criteria that once again introduced a measure of *weighting*, in this case of amino acid *properties* such as their being hydrophobic or hydrophilic, which are thought to have functional and thus evolutionary implications.

To the same end, Margaret Dayhoff and collaborators proposed in the 1978 edition of the *Atlas of protein sequence and structure* a substitution matrix that 'describe[s] the amino acid replacement probabilities between two sequences at various evolutionary distances'. They also proposed a scoring matrix (for those substitutions) that was 'more accurate ... and ... more sensitive in detecting distant relationships' (Dayhoff et al., 1978, p. 345). This substitution matrix, known as Percent Accepted Mutation (PAM), was based on the assumption that 'The observed behaviour of amino acids in the evolutionary process must consider the frequency of change of each amino acid to each other one [sic] and the propensity of each to remain unchanged' (Dayhoff et al., 1978, p. 345).

PAM matrices have been very popular since. Molecular phylogeneticists have praised their advantages and improved or revised their scope, although they have also noticed their failures and biases.[20] Some of the possible sources for error or biases are the different assumptions upon which PAM matrices are built: that proteins have an average amino acid composition, that mutation rates remain constant along the entire molecule, and that the probability of a mutation at a particular site is determined only by the residue at that position. As knowledge of the physical process of molecular evolution has increased, such suppositions have been questioned.

Moreover, PAM matrices are also affected by the kind of evidence used for their construction. Dayhoff, Schwartz and Orcutt used *closely* related proteins in their study (Dayhoff et al., 1978, p. 345). According to Henikoff and Henikoff, these sequences were '85% identical. However, the most common task involving substitution matrices is the detection of much more distant relationships, which are only inferred from substitution rates in the Dayhoff model' (Henikoff & Henikoff, 1992, p. 10915). In 1992, Jones et al. updated Dayhoff s matrix. They claimed that even though the main features of the 1978 matrix still held, their renovation uncovered important aspects that were not taken into account in the original matrix (Jones et al., 1992, p. 281).

Despite the overwhelming presence of DNA sequence-based phytogenies during the 1990s, the use of proteins was revitalized when Tobias Müller and Martin Vingron elaborated a second extension to Dayhoff s matrices. They used an Evolutionary Markov Model,[21] which under certain conditions 'allows a meaningful

[19] We will describe the maximum likelihood approach with more detail when referring to the construction of trees below. In the case of likelihood methods the debate on *subjectivity* is not focused on the mathematical tool *per se*, but on the scientist's conviction that this is the best way to assess similarity, and on the assumptions about DNA evolution that are built into the models tested. For a discussion of the statistical characteristics and possible inconsistencies of this methodology, see Sober (1988a,b).

[20] See Fitch & Smith (1983), Feng et al. (1985), Atschul (1991), Henikoff & Henikoff (1992), and Jones et al. (1992).

[21] A Markov Model (or process) is a mathematical model for an evolutionary random process with no memory. This means that the current state is known but it is independent of its past, and its future can only be presented as transition probabilities to any of the new outcomes (including remaining in the current state). Evolutionary models involving base changes or amino acid changes in a sequence are modeled as Markov processes because they can be conceptualized in the following way: for example, if at a given in site in a DNA sequence we find a Thymine residue, it has P probability of changing to Adenine, Q probability of changing to Guanine, R probability of changing to Cytosine, and S probability of remaining as Thymine on the next simulation (or mutation) round. In any case, the specific nucleotide that is present before the analysis is made is of no relevance to the modeling process.

description of the amino acid replacement process' (Müller & Vingron, 2000, p. 764).[22] Later, Müller et al. (2002) proposed another method for estimating amino acid substitution models. In this case they developed a maximum likelihood method and compared it with both their 2000 Markov chain method, and with Dayhoff's 1978 approach.[23] They recognized that the maximum likelihood approach 'perform[ed] best for *small* data sets, whereas for *larger* data sets—where maximum likelihood becomes computationally unfeasible—, the resolvent [2000] method is a good alternative' (Müller et al., 2002, p. 12; our emphasis). Both methods were said to outperform Dayhoff's. But it was again up to the researchers' judgment and experience to decide which matrix to use in each case.

Henikoff and Henikoff had developed a different approach to the construction of substitution matrices ten years before; they called it BLOSUM. They derived their substitution matrices 'from about 2000 blocks of aligned sequence segments characterizing more than 500 groups of related proteins' (Henikoff & Henikoff, 1992, p. l0915). The two kinds of matrices (PAM and BLOSUM) are somehow compatible, or at least comparable (ibid., p. 10917), and this again introduces the need to choose between them or to use both methods in order to have more robust results.

The weighting of characters (in this case, of the differences between amino acids according to one or more matrices) resurfaces as a point of conflict. It is agreed that there should be some kind of measure for difference or similitude when comparing amino acid sequences, but determination of the proper measure is no minor task. In addition, the degeneracy of the genetic code may increase the probability of some amino acids remaining unchanged (even in the case of genetic mutation) while others will be more prone to change in strictly molecular terms, and these matters should also be taken into consideration.

In sum, we have two major positions on the methods of molecular phylogenetics. Those who favor the *biologically* sensitive approach assert that 'Although there is no certainty that weighting will help in any particular situation, it is seldom a handicap. The best weighting schemes take account of both genetic likelihood and structural similarities between amino acids' (Feng et al., 1985, pp. 123–124).[24] Instead, authors who emphasize the importance of *statistical* and formal criteria declare that 'All of our assumptions are questionable from a biological point of view. However, from the perspective of data analysis it is obvious that one needs to simplify to make model fitting practical. The challenge is to reflect as much of the reality as possible' (Müller et al, 2002, p. 12). It is clear, then, that molecular phylogeneticists are well aware of the conundrum of developing more objective methods, and while each side has its preferences (the tension being primarily between biological sensitivity or statistical footing), both seek to make their own criteria as explicit as possible.

3.4. The degeneracy of the genetic code and the multiple hit problem

Even before the genetic code was completely elucidated there were a number of reflections on its redundancy or *degeneracy*, and its effect on the evolutionary process. Zuckerkandl and Pauling (1965), for instance, noted that:

> The evaluation of the amount of differences between two organisms as derived from sequences in structural genes or in their polypeptide translation is likely to lead to quantities different from those obtained on the basis of observations made at any other, higher level of biological integration. On the one hand some differences in the structural genes will not be reflected elsewhere in the organism, and on the other hand some differences noted by the organismal biologist may not be reflected in structural genes. The first proposition should hold on account of the degeneracy of the genetic code. (Zuckerkandl & Pauling, 1965, pp. 99–100)

By 1966, Fitch had already concluded that the degeneracy of the code represents an evolutionary advantage when it comes to *buffering* the effects of mutations, and he reflected thoroughly on the increased probability of some mutations occurring, which allowed for more variation at the genetic level. Moreover, in the molecular tree constructed by Fitch and Margoliash (1967) they took into account not the amino acid sequences but the *minimum changes* in nucleotide composition expected from the recently worked out genetic code.

Very soon, however, a new difficulty concerning the comparison of molecules and the chemical structure of DNA emerged, namely, the quantitative determination of how many point mutations had occurred at a given nucleotide position during the evolutionary history of a molecule. It is usually very difficult to estimate whether a 'hidden' substitution has taken place, either because it has been reversed or, in an even more general case, because it is not possible to assess the number of possible intermediate point mutations that have taken place at that position. This is known as the *multiple substitutions* or *multiple hit problem*. Molecular evolutionists explain it thus:

> when the degree of divergence between two nucleotide sequences is small, the chance for more than one substitution to have occurred at any site is negligible, and if the number of divergence is substantial, then the observed number of differences is likely to be smaller than the actual number of substitutions due to *multiple substitutions* or *multiple hits* at the same site. (Li & Graur, 1991, p. 48)

Again, molecular phylogeneticists have devised a number of statistical methods and criteria to estimate the rates of synonymous and non-synonymous nucleotide substitutions in a given sequence, based on parameters such as the so-called transition[25] *versus* transversion[26] ratios, and the number of degenerate positions for a set of codons. Such methods, needless to say, do not remain unproblematic and some models are said to be better than others.[27] In fact, the number of alternative methods and the growing number of criteria to choose among them is an indication of the difficulties underlying the issue (for more details see Dietrich, 2008).

3.5. (Re)constructing the tree

Philosophers of biology have paid considerable attention to the methodological debates related to the construction and selection of trees once the comparison of traits has been made.[28] Fitch and

[22] Their approach consisted in a series of iterations that went back and forth 'between estimating the evolutionary distances between the sequences in an alignment and updating the current rate matrix' (Müller & Vingron, 2000, p. 765). In contrast to Dayhoff, they were able to estimate their matrices using alignments with various degrees of divergence.

[23] The model proposed by Dayhoff is also a Markov chain that acts independently at each site of the studied protein. See Müller & Vingron (2000) and Müller et al. (2002).

[24] If a hydrophobic amino acid is replaced by another hydrophobic residue, it is generally agreed that there are more chances that the chemical properties, and thus the biochemical and evolutionary functions, will be preserved at that site of the protein.

[25] The substitution of a purine (adenine or guanine) for another purine, or a pyrimidine (thymine or cytosine) for another pyrimidine.

[26] The substitution of a pyrimidine for a purine or *vice versa*.

[27] For instance, see the discussion of this issue included in Jukes & Cantor (1966), Kimura (1980), Li et al. (1985), Nei & Gojobori (1986), and Li (1993), to cite just a few.

[28] These include Schejter & Agassi (1981), Sober (1988a,b, 2006), Haber (2005), Viktor (2006), Fitzhugh (2006), and Rieppel (2006).

Margoliash's tree was conceptualized as 'a graphical representation of the order in which the subsets [of characters] were joined' (Fitch & Margoliash, 1967, p. 280). The results of the alignment, in the form of a distance matrix, were used to reconstruct the ancestral relations between the sequences compared. Nowadays this task is performed by highly specialized software that aims to reconstruct a given phylogeny from these data (usually in the form of a distance or similarity matrix) by using one or several algorithms.[29] This adds a major and different concern: for most practitioners the software packages are literally black boxes, and the automation of procedures obscures the methodological decisions implied in those packages. An unintended consequence of this fact is that today's practitioners tend to move across the methodological boundaries once imposed by rivaling conceptions, simply by taking advantage of the different statistical tools available to them. The tools at hand prevail over methodological commitments. We will come back to both issues in our concluding remarks.

The methods for constructing phylogenies are classified into three major groups: (1) Distance Based Methods (DBM), (2) Maximum Parsimony Methods (MPM), and (3) Maximum Likelihood Methods (MLM). The last two are generally grouped together as character based methods. There are also Bayesian inference methods, but since these are included in the likelihood methods, we will not treat them separately.

Distance based methods (DBM) count the differences between two sequences and transform them into a distance matrix. The matrix is then used to build a tree by grouping sequences that have the shortest distance and adding others as the distances grow larger. It is a simple method, and it does not demand enormous computational capabilities. The method described by Fitch and Margoliash in 1967 belongs to this category. In this case the tree's topology was assumed to be 'true,' since it was consistent (convergent) with the taxonomic knowledge of the time.

Distance methods have been criticized because they assume that the rate of evolution is constant for every branch of the tree; and more importantly, because they do not take into account the possibility of multiple hits in the same nucleotide position.[30] However, different corrections and adjustments can be made and several 'models of evolution' can be implemented under this methodology. Most of the software used nowadays to reconstruct phylogenetic trees using distance methods is capable of adjusting for multiple hits.[31] A different critique against distance methods is that some information is lost due to the fact that the only parameter considered is the number of differences per site. Practitioners agree that tree topology should be backed up by some kind of statistical analysis and, typically, the *bootstrap* approach is used (see below).

Maximum Parsimony Methods (MPM) assume that the most parsimonious evolutionary tree 'might be expected to have a high degree of correspondence to the true phytogeny. Its justification lies in the most efficient use of information in the most efficient way and *does not presuppose that evolution follows a most parsimonious course*' (Fitch, 1971, p. 406; our emphasis).[32] Critics are quick to point out that more than one 'most parsimonious tree' is often obtained from the same dataset. Indeed, Maximum Parsimony Methods frequently yield numerous trees—sometimes thousands of them—with the same score (that is, they are equally parsimonious). It becomes necessary to choose one among the several trees, in which case only groupings that are strictly *convergent* are considered to be supported by the data.[33]

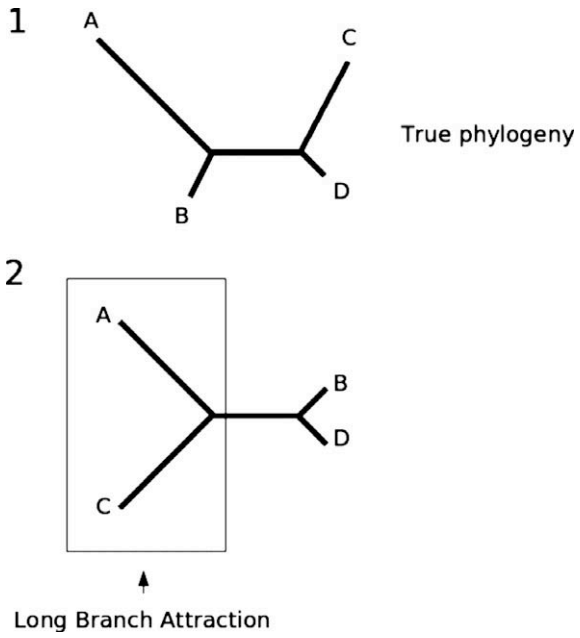

Fig. 2. Long Branch Attraction (LBA). Species A and B, and species C and D are phylogenetically related, but given the high substitution rates in A and C, they are rendered as closely related. That is, the two long branches attract each other, while the two short ones are excluded—another artifact known as Short Branch Exclusion (SBE) (modified from Felsenstein, 1978, and Stiller & Harrell, 2005).

A second problem of parsimony methods (but one that is not restricted to these) is the so-called *Long Branch Attraction* (LBA) phenomenon, depicted in Figure 2. The LBA phenomenon is an artifact that is produced when rapidly evolving lineages are analyzed. The observed results support the (equivocal) hypothesis that those lineages are closely related, regardless of their actual evolutionary history. This situation is a by-product of the multiple hit problem, since in some cases DNA substitution rates are so high that the probability that two lineages may have the same nucleotide in the same place increases considerably. Parsimony methods erroneously construct trees that lead to conclude that a given character evolved only once in a common ancestor. Thus, in Fig. 2, species A and C are interpreted as having a common ancestor insofar as it is not possible to know the number of multiple mutation events that have separated them during evolution. This problem can be solved either by using other methods, such as likelihood, or by adding new species to the comparison in order to increase the 'resolution' of those branches. But identification of such artifacts requires expertise, a procedure that is difficult to account for within an impartial and explicitly quantitative methodology.

As mentioned, some of the most notorious methodological debates in molecular evolution have concerned the use of the parsimony criterion for constructing and choosing between phylogenies (see Sober, 1988a, b). Nevertheless, scientists and philosophers alike have provided important arguments in favor of its application. Parsimonious phylogenies probably constitute the

[29] Two programs that are commonly used to infer phylogenies are PAUP and Phylip. They perform different algorithms and are based on different methodologies; they also make different mathematical and philosophical assumptions (see below).

[30] This is due to the fact that they deal not with the actual characters, but with a measure of their similarity (or difference). Thus, in principle it is not easy to assign a difference value to something that is already non-equal to the original character state.

[31] There are a number of methods that use distance measures to reconstruct trees, among them the Unweighted Pair Group Method with Arithmetic Mean (UPGMA), Neighbour Joining (NJ), Fitch-Margoliash (FM) and Minimum Evolution (ME). The use of UPGMA is not recommended according to the specialized literature.

[32] In this context, parsimony refers to the idea that 'the best supported phylogenetic hypothesis is the one that requires fewest evolutionary changes' (Sober, 1988a, p. 31).

[33] See Fitzhugh (2006) for a detailed proposal of how to take into account the Hempelian *requirement of total evidence*.

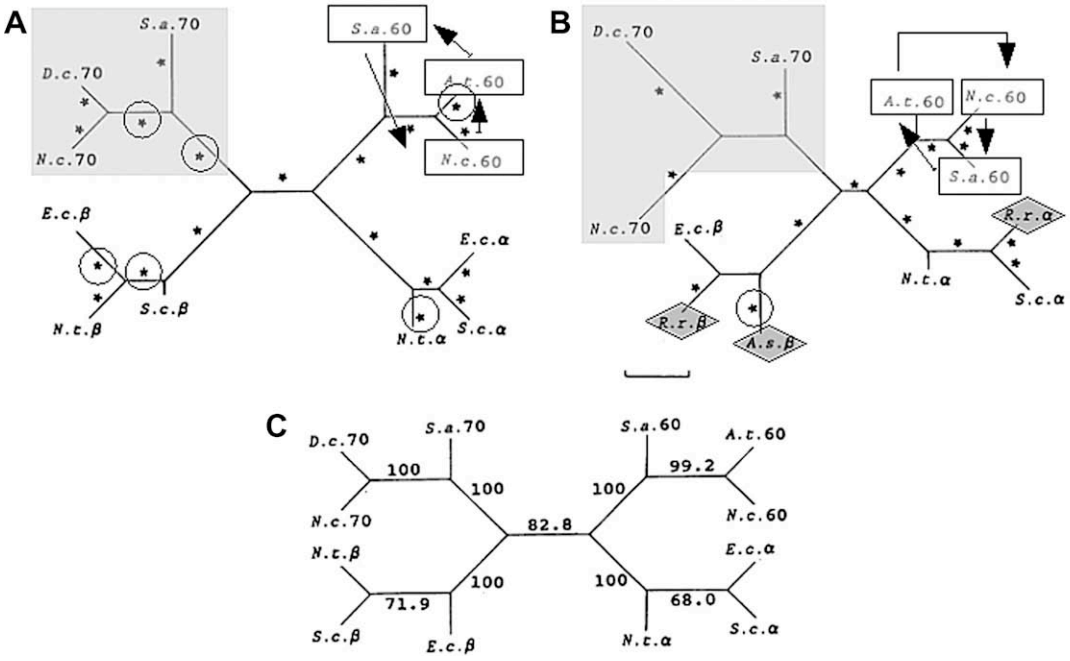

Fig. 3. Tree (A) was obtained using a likelihood algorithm, while (B) was constructed using a parsimonious method. The general topology of both trees looks very similar, but closer scrutiny reveals that they only have the same order and topology as clade 1 (gray area), even though the methods do not render all branches (marked with circles). Clade 2 (marked with squares) has the same species on both trees, but the order of the elements is different, that is, it is rotated (see arrows) and, again, not all the branches are equally supported. Clades 3 and 4 were constructed using different molecules in each clade (gray diamonds), however some consistency is found in the molecules 'shared' by those clades, even though they are not equally supported by the data. (C) gives the Bootstrap values for each of the branches of tree A; Clades 3 and 4 do not have the best bootstrap values, casting some doubts on how well supported they are by the data.

most widely extended approach, despite the arrival of a more statistically oriented pheneticism. The reasons for this are not obscure, and they have to do with biologists' long-standing concern with ancestral relations or homology, as opposed to mere resemblance.[34] To the more statistical minds, however, this methodology continues to be problematic (see Felsenstein, 2001).

The maximum likelihood approach (MLM) is more in accordance with the original empiricist tenets of pheneticism, and it competes with parsimony methods. The approach includes statistical methods that assess the probability that a set of data might be explained by an evolutionary model that is determined *a priori*. This model might state, for example, that all possible substitutions between the four nucleotides are equally probable. The likelihood of a tree will obtain from the sum of the probabilities of each possible substitution reconstruction under a particular substitution model, and these likelihoods will then be multiplied to give an overall likelihood of the tree. In general, not all the branches of a tree are equally likely, but in a 'good' tree most of the branches will have high scores of likelihood. It is also possible that a 'good' tree under a certain model is actually a 'bad' one under a different model; this occurs because the evolutionary models used in each case are *ad hoc* to a certain extent.[35]

Maximum Likelihood Methods use many computational resources due to the very large amount of calculations they require. This used to be a disadvantage, but is becoming methodologically less important given the growing computing capabilities of modern biological laboratories, particularly since the 1990s. There are still situations, however, in which the use of computational resources is impractical (for an example see Müller & Vingron, 2000). Regardless of those circumstances, the defenders of this approach argue that it 'outperform [s] M[inimum] E[volution] and M[aximum] P[arsimony] when the data analysis proceeds according to the same model that generates the data' (Brinkman & Leipe, 2001, p. 344). Supporting this claim is the belief that a *blind* statistical approach is better suited for phylogenetic reconstruction than a biologically sensitive one. In this case, *election* of a tree is not required since the maximum likelihood score produces the best tree. Also, the parameters of the score are given in advance as a set of explicit rules and statistical criteria that detractors still consider to be unwarranted assumptions about the evolutionary process.

Frequently, discrepancies between reconstructed trees point to the inadequacy of one phylogenetic inference method or another. Figure 3, taken from Gogarten et al. (1989), is a good example of the differences that may result from using different methods to reconstruct phylogenetic trees.

From a critical standpoint, the algorithms used to build trees reflect oversimplified assumptions about within-molecule or between-lineage variation in rates of sequence change, and are insensitive to mutational saturation. As we mentioned before, however, it is usually the case that there is more than one possible tree for a particular set of data, or that a 'quality value' needs to be assigned to the tree(s) obtained. To do so, a number of methods exist, the most popular being *Bootstrap*—a re-sampling of the tree. Bootstrap is a statistical tool suitable for resampling trees constructed by any of the three methods. It was developed by Efron (1979), and Felsenstein (1985) was the first to use it for phyloge-

[34] See Sober (1988a) for a more detailed treatment of this problem, and Hall (1994) and Bock (1994) for thorough historical and recent treatments of the concept and the problem of homology.

[35] Maximum likelihood methods require postulating a prior evolutionary model to which the test assigns a fitting probability; that is the reason why models may be a better or worse fit for the data.

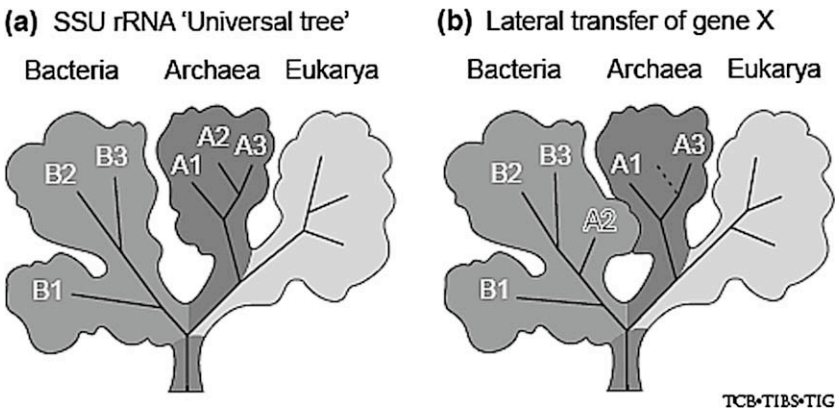

Fig. 4. A hypothetical phylogeny where a Lateral Gene Transfer event has taken place is compared with the 'Universal Tree' or the 'Tree of Life' (from Doolittle, 1999, p. M6).

netic analysis.[36] *Parametrical Bootstrap* is a variation that uses simulated replicates (rather than pseudo replicates as in *non-parametrical bootstrapping*) to evaluate the tree in a way similar to non-parametrical bootstrapping. In any case, the decision to use one procedure or another is not free of judgment, experience, preferences or interpretation.

Disagreements about the best tree choice are sometimes interpreted as something more than methodological differences. For instance, the quality of the data (gene sequences) used as characters may *systematically* affect the results. But as Rieppel and Kearney (2007) have pointed out, very little research on the quality of evidence (data bases) has been conducted.

Something that all too often is forgotten by molecular systematists is that the two main inferential approaches (parsimony and likelihood) provide sets of rules *for evaluation of the results*, not rules for their *acceptance*. In the case of parsimony, 'the most you should conclude is that the most parsimonious tree is *better supported* than the others, you are not obliged to believe that the most parsimonious tree is true' (Sober, 2006, p. 44). In the case of likelihood, if a first hypothesis confers a higher probability on the data than a second hypothesis, it does not necessarily follow that the first one is true; it does not even follow that it has a higher probability of being true, it is only best supported by the data (ibid.).

Many systematists continue to argue that accepting or choosing a tree requires more than statistical tests and results. To make this point even clearer, it is important that we quote a recommendation included in a widely used textbook on bioinformatics, concerning the search for a good phylogenetic tree:

> One of the best ways to economize the search effort is to *prune* the data set. For example, it might be *apparent* from the data alone or from preliminary searching that a particular cluster of five terminals is unresolvable, that the arrangement of these terminals does not impact the remainder of the topology, and/or that resolution of these terminals is not the objective of the analysis. *Removing* four of the terminals from the analysis *simplifies* the search by several orders of magnitude. Every analysis is unique. The elements that influence the choice of optimal search strategy (amount of data, structure of data, amount of time, hardware, objective of analysis) are *too variable to suggest a foolproof recipe*. Thus, *researchers must be familiar with their data*; they must also have specific objectives in mind, understanding the various search procedures as well as capabilities of their hardware and software. (Brinkman & Leipe, 2001, pp. 345–346; our emphasis)

The language employed in this textbook differs widely from the one used by G. G. Simpson in his writings on primate taxonomy. It alludes to amounts of data and time, and to hardware capabilities, instead of interpretative skill or emotional investment. Nevertheless, just like Simpson's text, the paragraph lays bare the decisions, experience and goals of the individual scientist (the objectives of analysis), all of which cannot be reduced to the practice of feeding a computer machine with data. In the end, molecular phylogeneticists are fully aware that all three methodologies have advantages and disadvantages, and a recent examination of the matter bears witness to the instrumentalism that characterizes the actual practices of the field (Rieppel, 2006).

3.6. Horizontal genetic transfer: questioning the tree itself

A different kind of problem arises when Lateral or Horizontal Gene Transfer (LGT/HGT) is taken into consideration because knowledge of the physical processes of molecular evolution is crucial to account for the different positions. In addition to this knowledge, methodological considerations also play an important role.

Microbial phylogenetics constitutes a field in which molecular data have become increasingly useful 'because the morphological evidence of biologists has proved to be inadequate to the task of organizing the major groups of bacteria and because the fossil record is difficult to interpret' (Schwartz & Dayhoff, 1978, p. 396). According to Doolittle, bacteria 'offer relatively little in the way of complex morphology and behavior' (Doolittle, 1999a, p. 2124), which is why sequence data have become essential.

One of the first attempts to reconstruct microbial phylogenies using molecular data (DNA and protein sequences of ferredoxines, *c* type cytochromes, and 5S ribosomal RNAs) was possible only once 'enough sequence information [had] become available from diverse types of bacteria and blue-green algae, and from cytoplasm and organelles of eukaryotes' and it allowed scientists 'to attempt the construction of a biologically comprehensive evolutionary tree' (Schwartz & Dayhoff, 1978, p. 395). It is worth mentioning that the results obtained by Schwartz and Dayhoff were taken to corroborate the theory of the endosymbiotic origin of eukaryotic organisms (Sagan, 1967; Margulis, 1981).

[36] The method generates several new data sets from the original one; it then reproduces the whole process and computes the number of times that a particular branch appeared in the tree (Bootstrap value). The new datasets are built by randomly sampling columns of characters *with* replacement, which means that it is possible to have the same position in the new dataset more than once, as well as to have a new dataset as large as the original one, but consisting of only one position repeated several times. There is still discussion on what this method is really measuring; Felsenstein (1985) suggested that it was a measure of repeatability, and on later interpretations it has been assumed to be a measure of accuracy (the probability that the true phylogeny is recovered).

Together with the work of Woese and Fox (1977) that we mentioned above, the field of bacterial phylogenetics reconfigured the way in which biologists today understand the early history of life and the distinctive turning points in the evolutionary history of Eukarya, Bacteria and Archaea. The transformations of biological thinking effected by the recognition of the three-domain division of organisms, together with the explanatory role played by endosymbiosis, are enormous and have only recently begun to be addressed by students of science.[37] Paradoxically, however, these discoveries also brought about some unexpected facts that threatened the entire enterprise of constructing the universal tree of life, namely, the realization of the importance of HGT/LGT (as shown in Fig. 4).

Schwartz and Dayhoff had noticed the possibility of HGT as early as 1978, although they quickly discarded it.[38] Nevertheless, as contradictory and conflicting data began to accumulate during the 1990s, and as more examples of non-matching phylogenies emerged, the idea that horizontal gene transfer was more common than previously expected came forward as a possible explanation for the observed anomalies. As a result, a new picture of microbial phylogenetics which challenged Woese's tree of life emerged (Hilario & Gogarten, 1993; Gogarten et al., 2002), with Doolittle as its main champion:

> If lateral gene transfer can affect all genes, and has affected some substantial fraction of genes over the past 3.8 billion years (since the origin of life), then much of what molecular phylogeneticists have hoped to accomplish is at risk, especially in the area of prokaryote evolution ... efforts to deduce the genetic make up of the last common ancestor of all extant life now appear misguided. There is no guarantee that a gene currently represented in some Bacteria, Archaea and Eukarya was present in their common ancestor—it could have arisen more recently in one domain and spread to the others. (Doolittle, 1999a, p. M6)

Even more radically, Doolittle points out: 'Molecular phylogeneticists will have failed to find the "true tree", not because their methods are inadequate or because they have chosen the wrong genes, but because the history of life cannot properly be represented as a tree' (Doolittle, 1999b, p. 2124). Underlying this cloudy scenario is a profound questioning of the pertinence of genetic molecular data for the purposes of reconstructing evolutionary patterns:

> it must be admitted (i) that it is not logical to equate gene phylogeny and organismal phylogeny and (ii) that, unless organisms are construed as either less or more than the sum of their genes, there is no unique organismal phylogeny. Thus there is a problem with the very conceptual basis of phylogenetic classification, (ibid., p. 2125)

In the place of Woese's 'Universal Tree of Life,' Gogarten and Doolittle have proposed a 'web of life' and, more recently, a 'web of genomes,' depicted in Figure 5.

On Woese's view; 'these excesses of interpretation result principally from a failure to take into account sufficiently [or to understand] the dynamics of HGT' (Woese, 2005, p. 105). According to him much of HGT is 'inconsequential' and has 'few if any phenotypic consequences' insofar as gene transfers have 'transient' residence or transient function or significance; that is, these genetic elements occur only in 'patches' and they persist only as long as the niche they inhabit persists. A more rare and permanent HGT, on the other hand, introduces biological 'novelty' and is the one process critical to cellular evolution (ibid.). Even though Woese's assertions rely on general biological facts, the lack of precise data concerning the relative occurrence of 'inconsequential' or 'transient' HGT suggests that his own opinion might be seen as a matter of judgment or even 'excessive interpretation'.

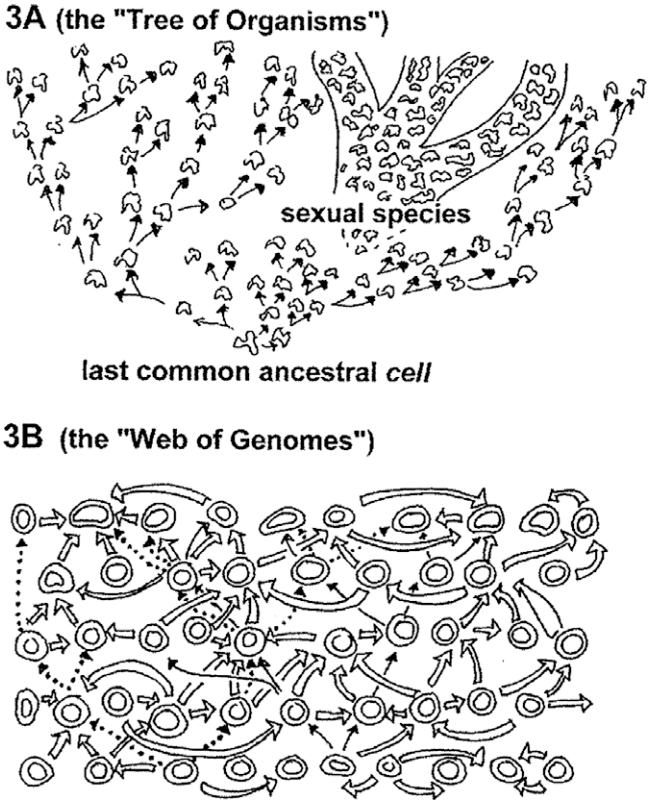

Fig. 5. The 'Tree of Life' *versus* the 'Web of Genomes' (from Doolittle, 2005, p. 131). Reproduced by Permission of Oxford University Press.

Still others challenge what they consider to be a *religious* reverence to Woese's Tree of Life. For instance, microbial phylogeneticist William Martin has recently pointed to the fact that there are as many methodological concerns with the SSU rRNA trees, as with other phylogenetic trees (Martin, 2005). He goes on to argue that molecular evolutionists should substantiate their reconstructions with more diverse biological and geological evidence. Moreover, he claims, attention should be given to the information provided by studies on the early atmosphere and geological conditions of the Earth in order to understand the pattern of evolution and the emergence of diversity.

O'Malley and Boucher (2005) have published a detailed historical account of this debate. What we would like to emphasize here are its methodological overtones. Doolittle expresses the matter thus:

[37] Sapp (2005, 2007); O'Malley & Dupré (2007).

[38] Horizontal Gene Transfer was discarded even though it was a well-known phenomenon characterized in the experiments of Avery, MacLeod & McCarthy (1944) on the transformation of pneumococci bacteria using exogenous DNA. To quote Schwartz and Dayhoff: 'we assume that the major types of bacteria have conserved the integrity of the groups of the sequences performing basic metabolic functions; we also assume that the substitution of a new sequence for one already functioning in a group through genetic transfer is sufficiently rare to be discounted. Frequent transfer between closely related species should not impair our ability to deduce the course of evolution of the major bacterial types. Only sequences that were transferred will lead to conflicting evolutionary histories for the species involved; sequences from any of the close species would be equally useful in deducing the evolutionary position of the bacterial type' (Schwartz & Dayhoff, 1978, p. 396).

> If we had a continuous videotape of all such [evolutionary] events in the last 4 billion years, we could reconstruct this tree in an unambiguous fashion. We do not have such a video, however; all we have are genome sequences and LGT, [which] means that there is no unique pattern of relationships between genomes. (O'Malley & Boucher, 2005, p. 131)

In the same vein, Martin writes: '[I]n an ideal scientific world, where everything is simple and straightforward, the analysis of genome sequences would have fully uncovered the basic backbone of life's history by now' (Martin, 2005, p. 134).

The methodological challenges then include how to weigh the relative importance of a phenomenon that is well in the past, when environmental and biological conditions were very different from today, as well as how to interpret some of its consequences: inconsequential or not? Weighting the relative occurrence of LGT (or HGT), however, is not the same as weighting the introduction of a methodological decision within the accepted framework of reconstructing the tree of life. Depending on the relative weight assigned to the phenomenon, the whole enterprise of comparing *individual* genes (or individual proteins) may be considered fatally flawed and the tree of life a chimera. The latest approaches in molecular phylogenetics respond to this challenge either by comparing *whole genomes*[39] or by excluding frequently transferred genes from the analysis; the latter involves, again, the weighting of evidence and the judgment of experts. Still, the reconstruction of early life depends on very different types of data (biological, chemical and geological) in the manner urged by Martin (2005). For some evolutionists, only when abandoning the restriction of using exclusively molecular data does the goal of constructing the tree of life seem possible.

4. The search for objectivity—or the avoidance of subjectivity

The plurality of meanings associated with objectivity has been the focus of both philosophical and historical investigation. Philosophers of science distinguish two modes or aspects of objectivity: one ontological, concerned with the way *in which things really are*; the other epistemological, concerned with the proper investigative attitudes *(impersonal* rather than idiosyncratic, or *public* instead of private), and the methods for reliable scientific research and representation (see, for instance Megill, 1994; Longino, 1990; Lloyd, 1995; Rescher, 1997).

In the case of molecular phylogenetics, this dichotomous type of analysis has important limitations. First, methodological debates of molecular phylogenetics incorporate both ontological and epistemological aspects. The search for what phylogeneticists call the 'the tree of life' (even if instrumentally understood as 'the best explanation for a set of data' for a given hypothesis), as well as the debates concerning the impact of HGT on the enterprise as a whole, betray the purpose of depicting the evolutionary process itself, that is, the actual patterns of speciation. To deal with this state of affairs, methods of inference become increasingly sophisticated as more details on the structure and dynamics of genes and genomes are established, and phenomena like the rates of evolution of different branches, or the incidence of multiple mutations in certain gene (or protein) domains are measured (not without contestation). As Sober (1988a) has remarked, knowledge of the process (which belongs to the ontological realm) affects the development of the methods of inference at the epistemological level.

The second limitation is even more important in our view, since it points to a range of problems infrequently addressed in a contextualized manner by philosophers of science. The development of more objective methods of phylogenetic inference is driven not only by a growing knowledge of the process, but also by a set of intertwined practices of *quantification* and *automation*.[40] We call this complex array of practices that constantly fret over the objectivity of the methods and the subjectivity of the interpretations, and are increasingly dependent on bioinformatics, *methodological anxiety*. The consequences of these practices on the dynamic of the field become apparent and are better addressed by a historical or situated perspective like the one we present in Section Three, instead of Bayesian or analytical philosophical approaches that recreate the never-ending debates among scientists (for instance, Haber, 2005, and Fitzhugh, 2006).

Over the last few decades, historians and sociologists of science have developed a new interest in the formation of basic categories of science (such as truth, fact, objectivity and proof).[41] The diversity of meanings attached to the notion of objectivity (that range from consideration of detached and impersonal attitudes, to a property that is ascribed to representations) does not, however, do justice to the very different contexts in which the concept has arisen, nor to the many layers composing the idea of objectivity and its correlate, subjectivity (Daston & Galison, 1992). Despite the usefulness of broad historical studies and the fact that these have managed to settle some issues in science studies, the case of molecular phylogenetics illustrates the limitations of general classifications of objectivity (for instance, Daston & Galison, 2007).

As already mentioned, in this case the analysis of objectivity is complicated by the fact that quantification and automation are intimately linked. Quantification has been said to play a role in the emergence of practices of objectivity. The 'view from nowhere' (Nagel, 1986), associated with certain forms of objectivity, forces a shift from the private to the public dimension of science in order to escape from local or *tacit* forms of knowledge. It was the appeal to familiarity, judgment, intuition and individual experience—forms of tacit knowledge—that was largely criticized either by cladists, pheneticists or molecular evolutionists in the 1960s (Hagen, 2003). The shift towards a more public dimension of knowledge requires quantification and the use of technologies that made possible both quantification and the communicability of methods, criteria and decisions (Porter, 1995). This is particularly true when a contested field or a polemic issue is at play, and when cognitive authorities are challenged (as it was the case of systematics in the early 1960s). It is only when rules and criteria are restrictive enough, and are made explicit by measurements and by statistical means of inference applied to quantitative data, that the goal of objectivity seems attainable. As we have seen, molecular phylogeneticists have taken this course of action repeatedly.

In contrast to other types of biological evidence, DNA and protein sequences are particularly prone to a quantitative analysis. They yield very large amounts of discrete (digitalized) raw data that can be analyzed with statistical tools. In previous works one of us has scrutinized the differences between these and other types

[39] Huynen et al. argue that 'the rate of this process [Lateral Gene Transfer] is not so high as to preclude a phylogenetic view of genome evolution. Genome phylogeny based on gene content disregards the evolutionary history of genes' (Huynen et al., 1999, p. 1443a).

[40] This point has been raised independently by Hagen (2001, 2003). However, while he focuses on the introduction of computers and statistical methods in systematics, illustrated in the confrontation between G. G. Simpson and Sokal & Sneath (1963), our focus is in the development of molecular phylogenetics, and on the question of objectivity/subjectivity in each of the steps a scientists takes in the construction of phylogenies.

[41] Well known examples of this trend include the socio-historical accounts of *truth* by Shapin (1995), of the concept of *fact* by Daston & Park (1998), of objectivity by Daston and Galison (1992, 2007) and Porter (1995), and of *proof* by MacKenzie (2001).

of molecular data, such as protein fingerprinting (Suárez, 2007) or DNA hybridization (Suárez, 2001) with respect to quantification practices. The discrete nature of amino acid and nucleotide residues and their linear structure/representation offer an apparently simple way to do one-by-one comparisons of the molecular traits of distinct species. Speaking the language of quantitative and statistical analysis[42] has been one of the aims of molecular biologists alike, since the early days of the field. In the field of phylogenetics they have quantified many things: data (including amounts of similarity or distance), criteria and even hardware capabilities. The automation of sequencing and the development of protein and gene databases around the world have also contributed with an ever-growing source of data to which quantitative analysis can be applied. In spite of this apparent consensus on how to handle the data, quantification practices have been the cause of many tribulations; scientists are compelled to specify their criteria, to develop even more sophisticated software, and to make use of tools of statistical inference in order to deal with the overwhelming numbers and with the myriad methodological decisions that need to be taken in order to construct a phylogenetic tree.

Computers are today's primary tools for doing phylogenetic research. The number of calculations needed to produce state-of-the-art phylogenetic trees is amazingly high. Computers and bioinformatics are used not only to calculate similarities or distances, but also to incorporate quantitative and semi-quantitative criteria and to perform the statistical inferences that generate representations such as cladograms and phylogenetic trees. They are also involved in many of the crucial steps of comparison; for instance, they participate in the alignment of sequences. The operations are tedious to perform, the possibility of generating mistakes increases with the number of operations, and the limits in the computing capabilities of human beings are an obvious obstacle, as molecular phylogeneticists are quick to recognize:[43]

> A ... problem is the large number of comparisons which can be made. First is the difficulty of comparing all of the possible alignments of two fairly dissimilar sequences in an attempt to look for any traces of homology. Even more serious is the fact that the number of possible comparisons between two distinct protein sequences increases as the square of the number of proteins of known sequence. Thus the advantages of using a modern digital computer to compare amino acid sequences are overwhelming. Computer comparisons are rapid and thorough. They also permit semi-quantitative criteria to be developed which will allow the significance of a suspected homology to be estimated. (Jukes & Cantor, 1969, p. 97)

Moreover, *quantification does not necessarily guarantee objectivity*, even as disentangling quantification and objectivity at the level of practice is an almost impossible endeavor. In the case of phylogenetics, statistical methods of inference and the reliance on *mechanical procedures* have played perhaps a more significant role in achieving impartiality than traditionally conceived practices of objectivity.

In this sense, along with quantification comes the need for automation, which in this field means relying on digital computers and software packages. As Jukes and Cantor point out, computers are used to calculate, but also to estimate the *significance of homologies*. Given that numbers (like letters, images or texts in general) do not speak for themselves, methods of statistical inference are deployed in the analysis and *interpretation* of huge amounts of molecular data, and even in the recent development of methods to choose between alternative methods. Though statistical tests do not provide rules for *acceptance* of the results (as we saw in Section 3), computers and statistical programs are used by a 'detached' phylogeneticist who aims at producing a tree by mechanical means, that is, through a set of rules (of inference) that a machine carries out.

The connection of automation with judgment and rational decision has been evident since the early reflections on its impact on business and scientific research. Automation has been equaled to 'the mechanization of judgment' and to 'machine control by non-human means' (Buckingham, 1961, pp. 14–15).[44] When connected to the use of computers, the study of automation has lead to 'a new fascination with the possibility of super-rationalism' (ibid., p. 22). Interestingly enough, in fields such as molecular phylogenetics this fascination has presented itself as an enduring anxiety to produce better computing methods.

The use of computers, along with the tools for statistical inference embedded in the software packages of phylogenetic reconstruction, constitutes very particular practices of objectivity. Quantification allows for the automation (mechanization) of measuring and the performance of statistical inferences; it also leads to representations that convey (measurable) phylogenetic distances among species, even if these are contested at each and every step of their construction. *Objectivity is thus built into the methods and representations of molecular phylogenetics through a concerted set of practices that emphasize quantification and automation.*

A last reflection on the general implications of these practices is in order. Novel tools tend to transform work and modify research requirements, and new requirements entail a new organization of work. All of this, in turn, makes new demands on the individuals involved if they are to perform the job adequately. What this means for scientific research in the field of molecular phylogenetics is not a trivial matter. Historically, an awareness of methodological issues has characterized systematists as a scientific community. This procedural sensibility shows no sign of having decreased with the advent of bioinformatics. However, we should be reminded that the structure of communities is never homogenous, and that not every practitioner takes part in the methodological debates that usually involve the most prominent scientists of the field.

As several authors have previously pointed out, the use of computers and the Internet for biological research does not necessarily lead to self-empowerment of users or to self-regulation of this practice. The Internet has greatly affected the process and criteria of standardization and control. Once the user (the scientist) develops a dependent relationship with computers and bioinformatic tools, she has few choices left at hand. Many practitioners of molecular phylogenetics seem to be 'trapped in the net' (Rochlin, 1997), victims of the sophisticated statistical packages that mechanize the output of knowledge based on virtual information and online access to databases—black boxes, in a literal sense. Rochlin's impression echoes wider concerns about the use of computers and the Internet:

[42] Kay's classic study (1993) and Holmes's most recent work (2006) are two instances where the scope of the quantitative-statistical drive of early molecular biologists—which owes much to their background in physics—is revealed. This drive has been operational in defining the ethos of their field.

[43] On the physical limits to computability, see Cherniak (1990). The messiness of natural systems, the historical and social nature of science, and the limits of human computing capabilities are—in our view—important arguments countering a philosophy of science that appeals to criteria such as the 'requirement of total evidence' (RTE) as the means to solving the debate between parsimony and likelihood approaches (for instance, Fitzhugh, 2006).

[44] In general, automation is defined as 'any continuous and integrated operation of a production system that uses electronic or other equipment to regulate and coordinate the quantity and quality of production' (Buckingham, 1961, p. 15).

> in practice, what it creates is an asymmetric dependency relationship ... where the user has little choice other than to accept it, and stay current with the latest version of software, or to reject it and drift down the irreversible path of obsolescence. (Ibid., p.25)

These considerations of the character of scientific work acquire an increased importance as biology is being materially and conceptually transformed by bioinformatics; they should therefore be given their due place in philosophical accounts of current biological research.

5. Concluding remarks: reflections on the nature of history

In the mid 1960s molecular evolutionists argued that molecular evidence was *cleaner* or *more direct* than the morphological evidence on which classical systematists had previously relied. Their voice was raised independently but simultaneously with those that challenged traditional evolutionary systematics, accusing it of relying heavily on authority and judgment, and of building previous hypotheses into the (re)constructed phylogenies. Informational molecules (protein and DNA) were ratified as the primary evidence of evolution, as *documents on the history* of living beings (Suárez, 2007), *not only* because the entire theoretical apparatus of evolutionary biology was built upon the explanatory primacy of genes, but also because molecular evidence was better suited for quantitative and statistical analysis than any other type of evidence.[45]

As Theodore Porter has demonstrated, measurement and statistics have been crucial in 'transforming local experimental skills into public knowledge' (Porter, 1992, p. 633); our contention is that they also transformed the *comparative skills* of molecular phylogenetics in the same way. This transformation was possible when considerable amounts of molecular data opened the door to the uses of statistical methods. Gould has written that 'the sciences of organic diversity do not usually seek identity in repeated experiment, but work by comparing the similarities among objects of nature as given. Kind, extent and amount of similarity provide the primary data of historical science' (Gould, 1986, p. 66). Classificatory or comparative skills, and their experimental correlates, have always depended on the kind and amount of data, but also on *localized* experience and judgment: the ability to observe and to focus on certain characteristics while being oblivious to others, or the expertise and familiarity with one's subject that make a systematist particularly good at what she does. But until recently, the comparative sciences faced a specific challenge: the complexity of organisms and their characters (how to individuate them?), and the difficulties of ascribing quantitative measures of similarity and difference. As we have seen, the comparative skills needed to single out, weight and ultimately compare two different sets of characters are not easily communicable or transmissible, and very often in the field of systematics are synonymous with the cognitive authority of individuals.

Molecular phylogeneticists have fostered the quantification of similarities and differences, the use of explicit quantitative and semi-quantitative criteria and of statistical methods of inference, and a wide array of techniques of self-restraint that include relying on machines (computers) for their research practices. But this is not the whole story. *Advances in making explicit the methodological criteria of molecular phylogenetics owe as much to the molecular and bioinformatic 'revolutions' triggered by quantification and automation as to the demands for clear criteria made by cladism and pheneticism.* Both schools have deep roots in the comparative organism-centered traditions of classical biology, which is why two further reflections on the nature of historical reconstructions deserve attention.

First, as illustrated in the debate concerning HGT and the 'tree of life', molecular phylogenetics has come—as have many historical disciplines—to value different sources of data, expertise and methods. The recognition among molecular phylogeneticists that morphological, geological, biogeographical, metabolic and other organismic (including genomic) data are important when assessing the biological relevance of a proposed classification (which aims to depict an historical pattern of speciation), points to the general value of robustness or 'attainable consilience' as a methodological desideratum (see Gould, 1986). A particularly interesting example of the value of consilience at work in phylogenetics practices is the very pragmatic attitude of using alternative methods for reconstructing phylogenies. This attitude may even be facilitated by the automation of procedures: a scientist does not make a major investment if she decides to 'run' her molecular data on different software packages, regardless of the methodological commitments of each of these tools. For instance, Figure 3 (above) shows how ancient duplicates of ATPase genes were used to root the tree of life and infer the origin of eukaryotes; the authors of the paper in which this figure appears use both maximum likelihood and maximum parsimony methods to assess the robustness of their conclusions.

This example takes us to the second issue. The practices of objectivity in molecular phylogenetics can be seen as part of a major trend towards ensuring the communicability of knowledge in a global scientific system. But molecular trees also need to make sense biologically—that is, locally—speaking: they are required to be compatible with the evidence of more traditional biological fields because the evolution of organisms is, to utter a truism, a contingent historical process, dependent on many (local) variables. The demand to come with biologically meaningful explanations—as opposed to mere statistical ones—has increased in the last years, as concerns about the relation between development and evolution have begun to transform the landscape of contemporary evolutionary biology.[46] This issue brings out a tension between 'biologically' and 'statistically' inclined systematists, and their uses of biological and/or statistical criteria. The former are more inclined to recognize the value of judgment and familiarity with biological groups; they also tend to incorporate different sources and kinds of data, familiar as they are with the difficulties of reconstructing the history of life. The latter struggle to sharpen the extent to which their criteria and methods are explicit, in their aim to provide robust results.

At the heart of this tension lies the ambiguous nature of all historical sciences. From Darwin's work onwards, the ordering of species is taken to reflect a historical pathway. But as Gould persuasively questioned: 'How can history be incorporated into science?' (ibid., p. 61). Without repeating Gould's account of the incorporation of history into science with Darwinian methods,

[45] This does not mean that molecular traits (in particular, DNA and protein sequences) are the only quantifiable characters. As the rise of pheneticism in the 1950s shows, there were many attempts to quantify morphological characters. S. J. Gould goes along the same lines in his insistence that morphological sciences had come to age in the early 1970s thanks to the development of 'a panoply of new machinery and techniques' that 'extended the bounds of both perception ... and analysis'. Curiously enough, Gould continues: 'But no machine can match the finest computer of them all', namely, 'evolutionary theory' (Gould, 1973, p. 401). Nevertheless, the question of why DNA and protein sequences lend themselves so easily to statistical treatment by computers needs to be addressed (see n. 7 above).

[46] For an illustration of this, see the work of evolutionist Neil Shubin (Shubin et al., 1997, 2006).

we can highlight his awareness that 'history is the domain of narrative—unique, unrepeatable, unobservable, large-scale, singular events' (ibid.). The answer he gave to this conundrum twenty years ago foresaw today's tension:

> In principle, the recovery of homology only requires a source of information with two properties: sufficiently numerous and sufficiently independent items to preclude, on grounds of mathematical probability alone, any independent origin in two separate lineages. (Ibid., p. 68)[47]

Molecular evidence meets both of these requirements. Also, as early molecular evolutionists asserted, molecular evidence is believed to be cleaner or more direct than morphological evidence. But, as we saw in Section 3, the availability of huge amounts of data and statistical methods has not solved the conundrum.

Perhaps we need a shift in perspective. 'Objectivity' may not have been gained, but has the ideal of objectivity in systematics been transformed with the advent of the molecular approach? Certainly, although many of the core problems of traditional phylogenetics remain pretty much the same, the introduction of new objects of research (sequences) and tools (computers and statistical tools) has definitely reshaped the ideals and practices of objectivity in this field. More importantly, the practices associated with these tools have transformed our knowledge of the biological processes and patterns of evolution, and they have given birth to new research problems.

What direction has the ideal of objectivity taken? Insofar as the goal of objectivity and the avoidance of subjectivity are two sides of the same coin, the answer to this question can be sought in the particular forms of subjectivity that molecular systematists are keeping away from their work. Our analysis of the methods and practices of molecular phylogenetics shows that issues of prestige or authority do not emerge as dangerous sources of subjectivity for the contemporary practitioner. Judgment and expertise, although frequently referred to, lack the idiosyncratic connotation that they had for traditional taxonomists dealing with primates in the early 1960s. What seems to be common to the practices of different kinds of molecular systematists is the calculated avoidance of building previous hypotheses about phylogenetic relations into the trees they reconstruct, and of allowing prior assumptions about evolutionary patterns to affect measures and analyses of resemblance. Nevertheless, as we have shown here, molecular systematists of every provenance (those inclined to parsimony as well as those who prefer likelihood approaches) make biological and/or methodological assumptions about the nature of the evolutionary process.[48]

This, however, is a different methodological situation to the one prevailing in the 1960s. Some of the original problems have been displaced, while others—the replacement of human experience and judgment with automation and the consequent loss of control and self-regulation—are new. In conclusion, molecular phylogeneticists, whether they are biologically or statistically inclined, have created a new set of practices that cannot be understood in terms of impersonal or mechanical objectivity alone. This set of practices is better portrayed as pursuing a mixture of different ideals of objectivity that integrate traditional biological practices (such as the reliance on convergence or robustness of data and methods), but also integrate new procedures associated to quantification and the rise of bioinformatics. The practices of objectivity have been retooled.

Acknowledgements

Edna Suárez-Díaz would like to thank the generous support of the Max Planck Institute for the History of Science in Berlin, the Programa de Apoyos para la Superación Académica (PASPA) of the Universidad Nacional Autónoma de México for the granting of a scholarship for a sabbatical year (2005–2006), and the Universidad Nacional Autónoma de México for making possible a research stay from 2007 to 2008. Víctor H. Anaya-Muñoz has a Deutscher Akademischer Austausch Dienst (German Academic Exchange Service) fellowship for doctoral studies at the Institute for Theoretical Biology of the Humboldt Universität in Berlin. Both authors thank Dr. Hans-Jörg Rheinberger and participants of MPIWG's Department III Colloquium for their valuable comments on a previous version of this paper. We also thank Lorraine Daston, Maureen O'Malley, Sergio Martínez, Hannes Luz, Skuli Sigurdsson, Rasmus Winther and an anonymous referee for their valuable and useful suggestions. Special thanks are due to Dr. Lorraine Daston for having generously provided early access to the book manuscript she recently published in collaboration with Peter Galison (2007).

References

Abouheif, E. (1997). Developmental genetics and homology: A hierarchical approach. *Trends in Ecology and Evolution, 12*, 405–408.
Altschul, S. F. (1991). Amino acid substitution matrices from information theoretic perspective. *Journal of Molecular Biology, 219*, 555–565.
Aronson, J. D. (2002). 'Molecules and monkeys': George Gaylord Simpson and the challenge of molecular evolution. *History and Philosophy of the Life Sciences, 24*, 441–465.
Avery, O. T., MacLeod, C. M., & McCarthy, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *The Journal of Experimental Medicine, 19*(2), 137–158.
Bock, G. R., & Cardew, G. (Eds.). (1999). *Homology. Novartis Foundation*. Chichester: J. Wiley and Sons.
Bolton, E., & Britten, R. (1963). *Yearly Report, Department of Terrestrial Magnetism, Carnegie Institution of Washington*, *62*, 303–327.
Brinkman, F. S. L., & Leipe, D. D. (2001). Phylogenetic analysis. In A. D. Baxenasis & B. F. F. Ouellete (Eds.), *Bioinformatics: A practical guide to the analysis of genes and proteins* (2nd ed.). New York: J. Wiley & Sons.
Brudno, M., Malde, S., Poliakov, A., Do, C. B., Couronne, O., Dubchak, I., & Batzoglou, S. (2003). Global alignment: finding rearrangements during alignment. *Bioinformatics, 19*(Suppl. 1), i54–i62.
Buckingham, W. (1961). *Automation: Its impact on business and people*. New York: The New American Library.
Ceruzzi, P. E. (1999). *A history of modern computing* (2nd ed.). Cambridge, MA: Massachusetts Institute of Technology Press.
Cherniak, C. (1990). *Minimum rationality*. Cambridge, MA: Massachusetts Institute of Technology Press.
Chadarevian, S. de (1998). Following molecules: Hemoglobin between the clinic and the laboratory. In S. de Chadarevian & H. Kamminga (Eds.), *Molecularizing biology and medicine: New practices and alliances 1910–1970s* (pp. 171–220). Amsterdam: Harwood Academic Publishers.
Daston, L., & Galison, P. (1992). The image of objectivity. *Representations, 40*, 81–128.
Daston, L., & Galison, P. (2007). *Objectivity*. Brooklyn, NY: Zone Books.
Daston, L., & Park, K. (1998). *Wonders and the order of nature, 1150–1750*. New York: Zone Books.
Dayhoff, M. O., Schwartz, R. M., & Orcutt, B. C. (1978). A model of evolutionary change in proteins. In M. O. Dayhoff & R. M. Schwartz (Eds.). *Atlas of protein sequence and structure* (Vol. 5, pp. 345–352). Washington, D.C.: National Biomedical Research Foundation.
Dickerman, A. W. (1991). Among-run artifacts in DNA hybridization. *Systematic Zoology, 40*(4), 494–499.
Dietrich, M. R. (1998). Paradox and persuasion: Negotiating the place of molecular evolution within evolutionary biology. *Journal of the History of Biology, 31*, 85–111.
Dietrich, M. R. (2008). Modeling molecular evolution: DNA and the rise of nucleotide substitution models. Paper presented at the Making Sequences Matter Workshop, Yale University (20–21 July), organized by B. J. Strasser and M. Sommers.

[47] It is noteworthy that Gould considered the best example of this to be not the most sophisticated techniques of DNA sequencing, but 'rather the "crudest" brute-force matching of all single-copy DNA' (Gould, 1989, p. 68)—DNA hybridization, a tool that reconfigured the field of ornithology in the late 1980s (see Sibley & Ahlquist, 1990).

[48] To illuminate the sterile debate on the relative merits of likelihood versus parsimony, the corresponding assumptions that each group makes should be considered contextually rather than in isolation, and their relative merits should be evaluated in particular cases, as philosopher Elliot Sober has recently proposed (Sober, 2006; see also Rieppel, 2006).

Doolittle, W. F. (1999a). Lateral genomics. *Trends in Biochemical Sciences, 24*(12), M5–M8.
Doolittle, W. F. (1999b). Phylogenetic classification and the universal tree. *Science, 284*(5423), 2124–21297.
Doolittle, W. F. (2005). If the tree of life fell, would it make a sound? In J. Sapp (Ed.), *Microbial phylogeny and evolution: Concepts and controversies* (pp. 119–133). Oxford: Oxford University Press.
Eck, R. V., & Dayhoff, M. O. (1966). Evolution of structure of ferrodoxin based on living relics of primitive amino acid sequences. *Science, 152*, 363–366.
Edwards, A. W. F., & Cavalli-Sforza, L. L. (1964). Reconstruction of evolutionary trees. In V. H. Heywood & J. McNeill (Eds.), *Phenetic and phylogenetic classification* (pp. 67–76). London: Systematics Association.
Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7*(1), 1–26.
Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology, 27*(4), 401–410.
Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the Bootstrap. *Evolution, 39*(4), 783–791.
Felsenstein, J. (1987). Estimation of hominoid phylogeny from a DNA hybridization data set. *Journal of Molecular Evolution, 3*, 123–131.
Felsenstein, J. (2001). The troubled growth of statistical phylogenetics. *Systematic Biology, 50*(4), 465–467.
Felsenstein, J. (2004). *Inferring phylogenies*. Sunderland, MA: Sinauer Press.
Felsenstein, J., Sawyer, S., & Kochin, R. (1982). An efficient method for matching nucleic acid sequences. *Nucleic Acids Research, 10*(1), 133–139.
Feng, D. F., Johnson, M. S., & Doolittle, R. F. (1985). Aligning amino acid sequences: Comparison of commonly used methods. *Journal of Molecular Biology, 21*, 112–125.
Fitch, W. M. (1966). The relation between frequencies of amino acids and ordered trinucleotides. *Journal of Molecular Biology, 16*, 1–8.
Fitch, W. M. (1970). Distinguishing homologous from analogous proteins. *Systematic Zoology, 19*, 99–113.
Fitch, W. M. (1971). Toward defining the course of evolution: Minimum change for a specific tree topology. *Systematic Zoology, 20*(4), 406–416.
Fitch, W. M. (1988). This week's citation classic. *Current Contents, 27*, 16.
Fitch, W. M. (2000). Homology: A personal view on some of the problems. *Trends in Genetics, 16*(5), 227–231.
Fitch, W. M., & Margohash, E. (1967). Construction of phylogenetic trees. *Science, 155*(3760), 279–284.
Fitch, W. M., & Smith, T. F. (1983). Optimal sequence alignments. *Proceedings of the National Academy of Sciences, 80*, 1382–1386.
Fitzhugh, K. (2006). The 'requirement of total evidence' and its role in phylogenetic systematics. *Biology and Philosophy, 21*, 309–351.
Gogarten, J. P., Doolittle, W. F., & Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Molecular Biology and Evolution, 19*(12), 2226–2238.
Gogarten, J. P., Kibak, H., Dittrich, P., Tainz, L., Bowman, E. J., Bowman, B. J., Manolson, M. F., Poole, R. J., Date, T., Oshima, T., Konishi, J., Denda, K., & Yoshida, M. (1989). Evolution of the vacuolar H+ -ATPase: Implications for the origin of eukariotes. *Proceedings of the National Academy of Sciences, 89*(September), 6661–6665.
Goodman, M. (1960). On the emergence of intraspecific differences in the protein antigens of human beings. *The American Naturalist, 94*(875), 153–166.
Goodman, M. (1996). Epilogue: A personal account of the origins of a new paradigm. *Molecular Phylogenetics and Evolution, 5*(1), 269–285.
Gould, S. J. (1973). The shape of things to come. *Systematic Zoology, 22*(4), 401–404.
Gould, S. J. (1986). Evolution and the triumph of homology, or why history matters. *American Scientist, 74*, 60–69.
Haber, M. H. (2005). On probability and systematics: Possibility and phylogenetic inference. *Systematic Biology, 54*(5), 831–841.
Hagen, J. (1999). Naturalists, molecular biologists, and the challenges of molecular evolution. *Journal of the History of Biology, 32*, 321–341.
Hagen, J. (2001). The introduction of computers into systematic research in the United States during the 1960s. *Studies in History and Philosophy of Biological and Biomedical Sciences, 32*, 291–314.
Hagen, J. (2003). The statistical frame of mind in systematic biology from quantitative zoology to biometry. *Journal of the History of Biology, 36*, 353–384.
Hall, B. K. (Ed.). (1994). *Homology: The hierarchical basis of comparative biology*. London: Academic Press.
Henikoff, S., & Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences, 89*, 10915–10919.
Higgins, D. G., & Sharp, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene, 73*(1), 237–244.
Hilario, E., & Gogarten, P. (1993). Horizontal transfer of ATPase genes: The tree of life becomes a net of life. *BioSystems, 31*, 111–119.
Holmes, F.L. (2006). *Reconceiving the gene: Seymour Benzer's adventures in phage genetics* (W. C. Summers, Ed.). New York: Yale University Press.
Hull, D. L. (1988). *Science as a process*. Chicago: University of Chicago Press.
Huynen, M., Berend, S., & Bork, P. (1999). Lateral gene transfer, genome surveys, and the phylogeny of prokaryotes. *Science, 286*, 1443a.
Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applied Biosciences, 8*, 275–282.
Jukes, T. H., & Cantor, C. R. (1969). Evolution of protein molecules. In H. N. Munro (Ed.). *Mammalian protein metabolism* (Vol. 3, pp. 21–132). New York & London: Academic Press.
Kay, L. E. (1993). *The molecular vision of life: Caltech, the Rockefeller Foundation and the rise of the new biology*. New York: Oxford University Press.
Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution, 16*, 111–120.
Laubnicher, M. D. (2000). Homology in development and the development of the homology concept. *American Zoologist, 40*(5), 777–788.
Lenoir, T. (1999). Shaping bionieatcine as an information science. In M. E. Brouden, T. B. Hahn, & R. V. Williams (Eds.), *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems* (pp. 27–45). Medford, NJ: Information Today, Inc.
Li, W.-H. (1993). Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *Journal of Molecular Evolution, 36*, 96–99.
Li, W.-H., & Graur, D. (1991). *Fundamentals of molecular evolution*. Massachusetts: Sinauer Press.
Li, W.-H., Wu, C.-I., & Luo, C. C. (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular Biology and Evolution, 2*(2), 150–174.
Lloyd, E. A. (1995). Objectivity and the double standard for feminist epistemologies. *Synthese, 104*, 351–381.
Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton: Princeton University Press.
Margoliash, E. (1963). Primary structure and evolution of cytochrome c. *Proceedings of the National Academy of Sciences, 50*(4), 672–679.
Margulis, L. (1981). *Symbiosis in cell evolution: Microbial communities in the Archean and Proterozoic eons*. New York: W. H. Freeman and Company.
MacKenzie, D. (2001). *Mechanizing proof: Computing, risk and trust*. Cambridge, MA: Massachusetts Institute of Technology Press.
Marks, J., Schmid, C. W., & Sarich, V. M. (1988). DNA hybridization as a guide to phylogeny: Relations of the Hominoidea. *Journal of Human Evolution, 17*, 769–786.
Martin, W. (2005). The missing link between hydrogenosomes and mitochondria. *Trends in Microbiology, 13*(10), 457–459.
McLachlan, A. D. (1972). Repeating sequences and gene duplication in proteins. *Journal of Molecular Biology, 64*, 17–437.
Megill, A. (Ed.). (1994). *Rethinking objectivity*. Durham, NC & London: Duke University Press.
Morgan, G. J. (1998). Emile Zuckerkandl, Linus Pauling and the molecular evolutionary clock, 1959–1965. *Journal of the History of Biology, 31*, 155–178.
Müller, T., Spang, R., & Vingron, M. (2002). Estimating amino acid substitution models: A comparison of Dayhoff's estimator, the resolvent approach and a maximum likelihood method. *Molecular Biology and Evolution, 19*(1), 8–13.
Müller, T., & Vingron, M. (2000). Modeling amino acid replacement. *Journal of Computational Biology, 7*(6), 761–776.
Nagel, T. (1986). *The view from nowhere*. New York: Oxford University Press.
Needleman, S., & Wunsch, C. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology, 48*(3), 443–453.
Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution, 3*(5), 418–426.
November, J. (2006). *Digitalizing life: The introduction of computers to biology and medicine*. Ph.D. thesis, Princeton University.
O'Malley, M. A., & Boucher, Y. (2005). Paradigm change in evolutionary microbiology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 183–208.
O'Malley, M., & Dupré, J. (2007). Towards a philosophy of microbiology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 38*, 775–779.
Podlaha, O., Webb, D. M., Tucker, P. K., & Zhang, J. (2005). Positive selection for indel substitutions in the rodent sperm protein Catsper 1. *Molecular Biology and Evolution, 22*(9), 1845–1852.
Porter, T. M. (1992). Quantification and the accounting ideal in science. *Social Studies of Science, 22*(4), 633–651.
Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton: Princeton University Press.
Rescher, N. (1997). *Objectivity: The obligations of impersonal reason*. Notre Dame: University of Notrè Dame Press.
Rieppel, O. (2006). Parsimony, likelihood and instrumentalism in systematics. *Philosophy and Biology, 22*(1), 141–144 (Book review).
Rieppel, O., & Kearney, M. (2007). The poverty of taxonomic characters. *Biology and Philosophy, 22*(1), 95–113.
Rochlin, G. I. (1997). *Trapped in the net: The unanticipated consequences of computerization*. Princeton: Princeton University Press.
Sagan, L. (1967). On the origin of mitosing cells. *Journal of Theoretical Biology, 14*(3), S.255–S.274.
Sapp, J. (Ed.). (2005). *Microbial phylogeny and evolution: Concepts and controversies*. Oxford: Oxford University Press.
Sapp, J. (2007). The structure of microbial evolutionary theory. *Studies in History and Philosophy of Biological and Biomedical Sciences, 38*, 780–795.
Schejter, A., & Agassi, J. (1981). Molecular phylogenetics: Biological parsimony and methodological extravagance. In J. Agassi & R. S. Cohen (Eds.), *Scientific philosophy today* (pp. 333–356). Dordrecht: D. Reidel Publishing Company.
Schully, S. D., & Helberg, M. E. (2006). Positive selection on nucleotide substitutions and indels in accessory gland proteins of the Drosophila pseudoobscura subgroup. *Journal of Molecular Evolution, 62*, 793–802.
Schwartz, R. M., & Dayhoff, M. O. (1978). Origins of prokaryotes, eukaryotes, mitochondria and chloroplasts. *Science, 199*(4327), 395–403.

Sellers, P. H. (1974). On the theory and computation of evolutionary distances. *Journal of Applied Mathematics (Siam), 26*, 787–793.
Shapin, S. (1995). *A social history of truth: Civility and science in seventeenth century England*. Chicago: University of Chicago Press.
Shubin, N. H., Tabin, C., & Carroll, S. (1997). Fossils, genes and the evolution of limbs. *Nature, 388*, 348–639.
Shubin, N. H., Daeschler, D. E., & Jenkins, F. S. (2006). The pectoral fin of Tiktaalik roseae and the origin of the tetrapod limb. *Nature, 440*, 764–771.
Shurkin, J. (1985). *Engines of the mind*. New York: Washington Square Press.
Sibley, C. G., & Ahlquist, J. E. (1990). *Phylogeny and classification of birds: A study in molecular evolution*. New Haven: Yale University Press.
Sibley, C. G., Comstock, J. A., & Ahlquist, J. E. (1990). DNA hybridization evidence on hominoid phylogeny: Results from an expanded data set. *Journal of Molecular Evolution, 30*, 202–236.
Simpson, G. G. (1964). The meaning of taxonomic statements. In S. L. Washburn (Ed.), *Classification and human evolution* (pp. 1–31). London: Methuen & Co.
Simpson, G. G., & Roe, A. (1939). *Quantitative zoology: Numerical concepts and methods in the study of recent and fossil animals*. New York: McGraw-Hill.
Simpson, G. G., Roe, A., & Lewontin, R. C. (1960). *Quantitative zoology* (Rev. ed.). New York: Harcourt, Brace & Co.
Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology, 147*, 195–197.
Smith, T. F., Waterman, M. S., & Fitch, W. M. (1981). Comparative biosequence metrics. *Journal of Molecular Evolution, 18*, 38–46.
Sober, E. (1988a). *Reconstructing the past: Parsimony, evolution and inference*. Cambridge, MA: Massachusetts Institute of Technology Press.
Sober, E. (1988b). Likelihood and convergence. *Philosophy of Science, 55*(2), 228–237.
Sober, E. (2006). Parsimony and its presuppositions. In V. A. Albert (Ed.), *Parsimony, phylogeny and genomics* (pp. 43–56). Oxford: Oxford University Press.
Sokal, R. R., & Sneath, P. H. A. (1963). *Principles of numerical taxonomy*. San Francisco: W.H. Freeman.
Sommer, M. (2008, Forthcoming). History in the gene: Negotiations between molecular and organismal anthropology. *Journal of the History of Biology*, 41(3). (Available at doi:10.1007/s10739-008-9150-3)
Stiller, J. W., & Harrell, L. (2005). The largest subunit of RNA polymerase II from Glaucocystophyta: Functional constraint and short branch exclusion in deep eukariotic phylogeny. *BioMed Central Evolutionary Biology, 5*(71). doi:10.1186/1471-2148-5-71.
Strasser, B. J. (2006). Collecting and experimenting: The moral economies of biological research, 1960—1980s. In H. J. Rheinberger, & S. de Chadarevian (Eds.), *History and epistemology of molecular biology and beyond: Problems and perspectives* (pp. 105–123). Preprint, 310. Berlin: Max Planck Institute for the History of Science.
Strasser, B. J. (2008). Collecting and experimenting: The moral economies of biological research, 1960s–1980s. Paper presented at the Making Sequences Matter Workshop, Yale University (21 June), organized by B. Strasser and M. Sommers.
Suárez, E. (2001). Satellite-DNA: A case-study for the evolution of experimental techniques. *Studies in History and Philosophy of Biological and Biomedical Sciences, 38*(1), 31–57.
Suárez, E. (2007). The rhetoric of informational molecules: Authority and promises in the early days of molecular evolution. *Science in Context, 20*(4), 649–677.
Templeton, A. R. (1985). The phytogeny of the hominoid primates: A statistical analysis of the DNA-DNA hybridization data. *Molecular Biology and Evolution, 2*, 420–433.
Templeton, A. R. (1986). Further comments on the statistical analysis of DNA-DNA hybridization data. *Molecular Biology and Evolution, 3*, 290–295.
Thorne, J. L., Kishino, H., & Felsenstein, J. (1991). An evolutionary model for maximum likelihood alignment of DNA sequences. *Journal of Molecular Biology, 33*, 114–124.
Vernon, K. (1988). The founding of numerical taxonomy. *British Journal for the History of Science, 21*, 143–159.
Vernon, K. (1993). Desperately seeking status: Evolutionary systematics and the taxonomists' search for respectability. *British Journal for the History of Science, 26*, 207–227.
Williams, D. M., & Ebach, M. C. (2005). Drowning by numbers: Rereading Nelson's 'Nullius in verba'. *The Botanical Review, 71*(4), 355–387.
Winter, P. W., Walsh, K. A., & Neurat, H. (1968). Homology applied to proteins. *Science, 168*(3861), 1433.
Woese, C. R. (2005). Evolving biological organization. In J. Sapp (Ed.), *Microbial phylogeny and evolution: Concepts and controversies* (pp. 99–118). Oxford: Oxford University Press.
Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences, 74*(11), 5088–5090.
Zuckerkandl, E. (1964). Perspectives in molecular anthropology. In S. L. Washburn (Ed.), *Classification and human evolution* (pp. 243–272). London: Methuen & Co.
Zuckerkandl, E., & Pauling, L. (1962). Molecular disease, evolution and genic heterogeneity. In M. Kasha & B. Pullman (Eds.), *Horizons in biochemistry: Albert Szent-Györgyi dedicatory volume* (pp. 189–225). New York: Academic Press.
Zuckerkandl, E., & Pauling, L. (1965). Evolutionary divergence and convergence in proteins. In V. Bryson & H. Vogel (Eds.), *Evolving genes and proteins* (pp. 97–166). New York: Academic Press.