

Perspectives

Anecdotal, Historical and Critical Commentaries on Genetics

Statistical Methods for Evolutionary Trees

A. W. F. Edwards¹

Gonville and Caius College, Cambridge CB2 1TA, United Kingdom

ABSTRACT

In 1963 and 1964, L. L. Cavalli-Sforza and A. W. F. Edwards introduced novel methods for computing evolutionary trees from genetical data, initially for human populations from blood-group gene frequencies. The most important development was their introduction of statistical methods of estimation applied to stochastic models of evolution.

The comparative ease with which certain aspects of population analysis are made possible by blood-group studies has already revealed ethnographic distinctions with a detailed precision quite unattainable without their aid. Although inevitably some ethnographers have at first felt ill at ease with the new facts, and what is indeed much more important with the new *potentialities* of serological studies, yet from a wider point of view it must be evident that all that ethnography strives to do will be facilitated beyond precedent as our knowledge of the blood-groups is extended. The traces left on existing populations by such past events as race mixture, by the relative expansion or attenuation of numbers, or by evolutionary modification under environmental influences, have certainly not by existing means been successfully unravelled. In considering the use of blood-group evidence in clarifying the ethnic history of Man, the peculiar advantage should be emphasized that independent evidence is within reach from different parts of the human germ plasm, so that independent confirmation should often be obtainable from the evidence of different factors independently inherited.

R. A. FISHER (1956)

ANY historian of science who chronicles events in which he himself has taken part should issue a disclaimer. First, his own memory is fallible and may even be innocently selective. Thus, John Venn, the leading exponent of the genre of university and college biographical history (and incidentally the inventor of the Venn diagram), gave the date of his degree of doctor of science as 1883 instead of 1884 in his own entry in a volume that he himself compiled (VENN 1898). Second, in the development of scientific

methodology there is no new thing under the sun. Every “new” idea is like another node in a spreading network. When placed in its historical setting, not only will an idea be seen to be the intersection of many advancing lines, but more often than not the very idea will be found earlier in the network, born before its time and recognizable only with the advantage of hindsight. As the historian of science K. O. MAY (1975) warned:

On closer examination, however, the hope of finding a “first” comes to grief because of the historically dynamic character of ideas. If we describe a result with sufficient vagueness, there seems to be an endless sequence of those who had something within the vague specifications. Even plagiarists usually introduce innovations! If we specify the idea or result precisely, it turns out that exact duplications seldom occur, so that every mathematical event is a “first,” and the priority question becomes trivial.

In this account of a development in which I was personally involved, I describe it from the point of view of a historian of science, stressing the influence of social connections and the scientific atmosphere as much as the contributions of particular individuals.

In the 1960s, I collaborated with L. L. Cavalli-Sforza in his project to develop methods for constructing phylogenetic trees from genetical data using modern statistical methods and the then-new computers. In a relatively short period of intense activity at the University of Pavia, Italy, we developed three methods: (1) least-squares based on an “additive” tree; (2) minimum evolution, also called maximum parsimony; and (3) maximum likelihood. Although this was before DNA sequences

¹Author e-mail: awfe@cam.ac.uk

were available, these methods underlie modern work on phylogenetic analysis.¹

I preface the account with a description of how the statistician and geneticist R. A. Fisher (1890–1962) bequeathed an “intellectual milieu” (to use his own expression; FISHER 1959) that might almost have been created specifically for the task at hand. A detailed account of the development of the methods follows.

THE FISHERIAN BACKGROUND

As early as 1925, in his pioneering book *Statistical Methods for Research Workers*, FISHER had clearly stated what many have not understood (see EDWARDS 2003):

When a large number of individuals [of any kind of organism] are measured in respect of physical dimensions, weight, colour, density, etc., it is possible to describe with some accuracy the population of which our experience may be regarded as a sample. By this means it may be possible to distinguish it from other populations differing in their genetic origin, or in environmental circumstances. Thus local races may be very different as populations, although individuals may overlap in all characters . . .

Fisher’s essential point was that the ability to investigate meaningful population differences from data such as human blood-group frequencies depends on the accumulation of information from a variety of blood-group systems, no one of which will reveal the phylogenetic structure by itself. The same is true today when the wealth of genetic material available for analysis in all species is effectively boundless.

At the Galton Laboratory, University College London, Fisher established a Serological Unit in 1935, which was funded by the Rockefeller Foundation to study human blood groups and other genetic markers with a view to initiating a linkage map of humans (EDWARDS 2004a). One of those who joined the unit was R. R. Race, who, after the Second World War, published with Ruth Sanger the pioneering book *Blood Groups in Man* (RACE and SANGER 1950) to which Fisher contributed the foreword. Race and Sanger became among Fisher’s closest friends and kept him abreast of the latest developments in serology. At the same time (1951), the Nuffield Foundation made a grant to the Royal Anthropological Institute, London, “to aid the study of blood groups as a clue to the understanding of human diversity,” and this enabled A. E. Mourant to write *The Distribution of the Human Blood Groups* (MOURANT 1954; foreword by H. J. Fleure, from which the above quotation about the purpose of the grant is taken).

¹I gave an earlier and more detailed account of the Pavia developments in a talk entitled, “History and philosophy of phylogeny methods” at the EC Summer School “Methods for Molecular Phylogenies,” Newton Institute, Cambridge, United Kingdom, on August 10, 1998, on which parts of the present article are based. The typescript of the talk was distributed to participants and is available from the author.

Fisher’s Galton Unit was evacuated to Cambridge during the war, and Fisher was reunited with it on becoming Professor of Genetics in 1943. When, in 1945, Race was promoted to be the Director of the Galton Unit following the death of the first director, G. L. Taylor, his position was filled by Mourant (MOURANT 1995). Therefore, for a brief period, Fisher, Race, Sanger, and Mourant were in regular contact in Cambridge. Fisher made great efforts to persuade the University of Cambridge to offer both Race and Mourant positions in his department after the war ended in 1945, but without success. In 1946, both took up posts at the Lister Institute in London as Directors of Medical Research Council Units (Box 1978) but continued to keep in close touch with Fisher.

In 1946 on a visit to Milan, Race had been introduced to Cavalli (*b.* 1922; later Cavalli-Sforza), then a young bacteriologist with a medical degree and an interest in genetics who “mentioned that he was then grappling with the study of Fisher’s statistical books” (Box 1978). Race wrote to Fisher about how impressed he had been with him. Cavalli was hoping to spend some time in England, and in March 1948 he arrived at the John Innes Horticultural Institution on a scholarship to spend 6 months working in the laboratory of K. Mather, a pre-war colleague of Fisher’s at University College. In July he traveled to Stockholm for the 8th International Congress of Genetics and met Fisher for the first time (he had already seen and heard him at meetings of the Genetical Society) “and after five minutes of conversation, he offered me a job at Cambridge in his department for setting up a laboratory to work on crossing over in bacteria” (CAVALLI-SFORZA 1990; CAVALLI-SFORZA and CAVALLI-SFORZA 2005). Cavalli accepted and started in Cambridge in October. Both Race and Mather also attended the Stockholm Congress and perhaps acted as midwives to Fisher’s offer. Cavalli spent 2 years in Cambridge and has written (CAVALLI-SFORZA 1991) about the genesis during his time at Cambridge of his idea of reconstructing evolutionary trees:

More than 40 years ago, when I was studying bacterial genetics in the laboratory of Sir Ronald A. Fisher of the University of Cambridge, the place was saturated with mathematical theorizing. Thus, it is not surprising that I started thinking about a project so ambitious it seemed almost crazy: the reconstruction of where human populations originated and the paths by which they spread throughout the world. I reasoned that the task could be accomplished by measuring how closely living populations are related to one another and by deducing from this information a comprehensive family tree.

Cavalli returned to Milan in October 1950 as a bacteriologist, taking with him the lasting friendship, personal and scientific, of Race, Sanger, Mourant, and Fisher. These were the world’s leaders in blood-group genetics, in the application of genetics to anthropology, and in statistics, especially as applied to problems of genetical



FIGURE 1.—Italy, 1963: standing—Luca and Alba Cavalli-Sforza and three of their children; seated—Hiroko Kimura, Anthony Edwards, Akio Kimura, and Thomas, Catharina, and Ann Edwards. Photograph by Motoo Kimura; reproduced by courtesy of Hiroko Kimura.

estimation. All that was missing from the realization of Cavalli's project was a suitable academic environment and adequate computing facilities.

The decade 1950–1960 saw astonishing advances in electronic computers for scientific use. In Cambridge, one of the leading centers, the pioneer machine EDSAC 1 performed its first calculation on May 6, 1949, and its last on July 11, 1958, by which time its successor EDSAC 2 had taken over (WILKES 1985). Appropriately, the first use of an electronic computer in biology was to tabulate the solution to a differential equation for Fisher (FISHER 1950).

By 1961, Cavalli-Sforza (as he became in 1950) had left bacteriological work and moved into the field of human population genetics with a professorship at the University of Pavia. The university had a modern computer on order, an Olivetti Elea 6001. Research funds were available to use it for human genetics and in particular to make associated appointments. Early in the year Cavalli-Sforza came back to his old Cambridge department and sought me out. As an undergraduate I had initially read physics and allied subjects but in 1956 applied to do genetics under Fisher in his last year as professor (see EDWARDS 1990). By 1961, I had completed a Ph.D. on the statistics of the human sex ratio, out of which had already come a number of publications. As part of my course I had been sent to Race and Sanger's laboratory to learn elementary blood grouping and had met Mourant as well. I had also been sent to Mather, by then in Birmingham, to study chromosomes. I had spent a month in Uppsala working with M. Fraccaro, an Italian colleague of Cavalli-Sforza's. Any of these researchers might have mentioned my name, as Fisher might have, for I had been the last undergraduate to be admitted to his department before he retired in 1957. Cavalli-Sforza offered me an appointment in Pavia and, after a short visit there, I accepted.

Like Cavalli-Sforza, I was not trained as a statistician and had similarly learned most of my statistics from Fisher's books (EDWARDS 2005). As a research student I had taught myself to use the Cambridge computer EDSAC 2 to tackle problems in genetics (EDWARDS 1961). It would be hard to imagine a better training for the problem of constructing phylogenetic trees from genetical data using statistical methods and electronic computers.

During a visit to Pavia in 1963 by Motoo Kimura and his wife Hiroko and son Akio, the Cavalli-Sforzas invited the Kimura and Edwards families to accompany them on an outing to the Alps, during which Motoo Kimura took a group photograph (Figure 1). This appears to be the only photograph showing Cavalli-Sforza and me together at the time of our collaboration.

PHYLOGENETIC ANALYSIS BEFORE THE COMPUTER

Acceptance of Darwin's view that *Homo sapiens* formed a single species whose races descended from a common stock led rapidly to the construction of phylogenies, most memorably that by HAECKEL (1876; reproduced in EDWARDS 2001), which placed the origin in Lemuria, a mythical continent in the Indian Ocean that delivered branches of humankind to Africa, India, the Far East, and Australasia before conveniently disappearing beneath the waves. Haeckel recognized 12 races, some of which he subdivided. All such phylogenies were, of course, based on classification by observable or measurable anthropological characters. Population differences in the *ABO* blood-group gene frequencies were first noticed during World War I and, as more blood groups were discovered and samples were collected from populations worldwide, "contour" maps depicting the frequencies of particular genes were derived (see

especially MOURANT 1954, *The Distribution of the Human Blood Groups*). BOYD (1952), in *Genetics and the Races of Man*, reviewed attempts to use blood-group information to define races, coming down in favor of six genetically defined races, although with no provision for a Pacific race because the blood-group evidence there was too discordant. But he stopped short of suggesting a phylogenetic tree, inclining in any case toward a polyphyletic origin. The use of the blood groups to construct a phylogenetic tree of human populations is probably the first example of the use of genetic data in phylogeny because it is unlikely that a comparable amount of data would have accrued for any other species.

At first sight it might seem natural to describe the development of the new methods of phylogenetic analysis against the background of taxonomic practice in the middle of the 20th century, but in fact it took place almost independently of the field of taxonomy. Rather, it was an attempt by geneticists and statisticians to build from scratch methods designed primarily to handle human blood-group gene-frequency data of the kind already mentioned. Unlike contemporary practice in taxonomy, the new methods of phylogenetic analysis belonged to the quite separate tradition of genetical model building and statistical estimation theory used, for example, in genetic linkage (MATHER 1951; BAILEY 1961).

The closest connection to taxonomic methods was with the “numerical taxonomy” promoted by SNEATH and SOKAL in a long article in *Nature* in March 1962 and subsequently in their book, *Principles of Numerical Taxonomy* (SOKAL and SNEATH 1963). In Pavia, we were familiar with this development through personal contact with Sneath, but it hardly overlapped with our thoughts because of the absence from it of any approaches using statistical estimation procedures. Of course, it was encouraging to find that others were also grappling with the problem of coaxing the early electronic computers into producing useful taxonomic output, and the discussion of clustering methods by Sokal and Sneath will have been helpful. But the main point to note is that, insofar as Sokal and Sneath were building on contemporary taxonomy, their book is a convenient reflection of that field as it stood in 1963. SNEATH (1995) reviewed the development in “Thirty years of numerical taxonomy,” although to describe numerical phylogenetics then as one of the “four areas of numerical taxonomy that were only faintly foreshadowed in 1963” is perhaps more monophyletic than the evidence can bear. FELSENSTEIN (2004) noted that MICHENER and SOKAL (1957), in “A quantitative approach to a problem in classification,” one of the founding articles of numerical taxonomy, derived a tree whose “interpretation as a phylogeny was made by Michener; Sokal saw it as a classification that did not necessarily have any validity as a phylogeny.”

Another useful guide to taxonomic thinking at the time is the book, *Phenetic and Phylogenetic Classification*,

from the Systematics Association meeting held in Liverpool (United Kingdom) in April 1964 (HEYWOOD and McNEILL 1964b). The meeting had also featured prominently in *Nature*: “The main aim of the symposium was to consider the principles which should be followed in the construction and arrangement of taxonomic groups and how these principles differ if the purpose of the classification is evolutionary or not” (HEYWOOD and McNEILL 1964a). Apart from our own article at the meeting (EDWARDS and CAVALLI-SFORZA 1964b), to which I return below, again there is no evidence of any overlap with our statistical genetics approach.

To maintain that contemporary taxonomic thinking made little contribution to the development of our methods of phylogenetic analysis is not, of course, to deny that with the benefit of hindsight certain similarities could be recognized. This is particularly true of the “method of minimum evolution,” called elsewhere the “principle of parsimony.” I return to this point after describing our developments.

THE DEVELOPMENT OF THE METHODS

Least-squares estimation on an additive tree: When I took up my post in Pavia in October 1961, the promised computer, an Olivetti Elea 6001, was late in delivery, so that it was not possible to start work on the computer analysis of the human population data as planned. Instead, in many discussions over the next few months, Cavalli-Sforza introduced me to his idea of developing methods for a computer-based construction of evolutionary trees from contemporary data, his particular interest being, as Fisher’s had been, the analysis of human populations using information from the blood groups. Initially, I was skeptical. I had studied linkage estimation theory under Fisher himself and I knew how difficult it was. I had visions of evolutionary tree estimation being much the same but with the addition of the need to estimate the form of the tree itself, surely a fatal complexity: my intuition was that there would be insufficient data for the task.

Cavalli-Sforza was very persuasive, however. Let us choose, he said, a number of distinct human populations, say 15, and from their gene frequencies at various blood-group loci let us compute the pairwise genetic distances between them. Now imagine a phylogenetic tree uniting the 15 populations, with a particular topology but unknown branch lengths. To each pairwise distance there corresponds a distance along the branches of the tree, being the sum of the relevant branch segments. These can now be estimated by the method of least squares. Different topologies can be tried and the one with the least residual sum of squares adopted. Because the data for the method are naturally presented as a square matrix of pairwise distances between the populations, this has often been called

the “distance matrix” method. It was wholly Cavalli-Sforza’s idea, from its inception right up to the specification of the estimation procedure. As a matter of record, the method was first mentioned in an article on cluster analysis presented at the International Biometrics Congress in September 1963 (EDWARDS and CAVALLI-SFORZA 1963b):

Alternatively, an additive model may be appropriate, in which the lengths of the segments are so estimated that the observed distance between two points is closely approximated by the expected distance along the tree, formed by the sum of the connecting segments. Least squares provides an elegant estimation procedure...

My first task was to write the computer program to implement Cavalli-Sforza’s proposal. The Olivetti computer was operational by May 1962, but as yet without any high-level language such as FORTRAN so that progress was quite slow. A FORTRAN II compiler became available early in 1963 and probably by June the additive tree least-squares program was working for 15 populations. I had made a special visit to Mourant in London and with his help had selected 15 human populations from around the world for which there was adequate information about gene frequencies for the commoner blood groups. The method and the results were reported by Cavalli-Sforza in an invited paper at the 11th International Genetics Congress at The Hague in September 1963 (Figure 2; CAVALLI-SFORZA and EDWARDS 1965).

The method of minimum evolution: The article also referred to an alternative method in which the criterion to be minimized was not the residual sum of squares but simply the total branch length of the tree in the multi-dimensional character space. This came about through my realizing that, although the points representing the 15 populations were displayed in a Euclidean space, the tree generated by Cavalli-Sforza’s least-squares procedure was not. Of course, one could simply add the condition that the nodes of the tree must be points in the same space, but, thinking about this early in the spring of 1963, it suddenly occurred to me that of all the possible trees in the space, one must be the shortest overall. Given that the metric of the character space was an “amount” of evolution, this shortest tree was the one that joined all the data points invoking the minimum amount of evolution.

This intuitive algorithmic solution was obviously quite sensible, for points close to each other in the character space would tend to be on the same branch of the tree. But it was also very easy to write the associated computer program EVOMIN to search for this minimum Steiner tree in a multi-dimensional space. A copy of EVOMIN still exists. I presented the idea at the summer meeting of the Genetical Society of Great Britain in London in July 1963 (Figure 3; EDWARDS and CAVALLI-SFORZA 1963a) and its use on the blood-group data was then reported at the Genetics Congress in September.

L. L. Cavalli-Sforza and A. W. F. Edwards

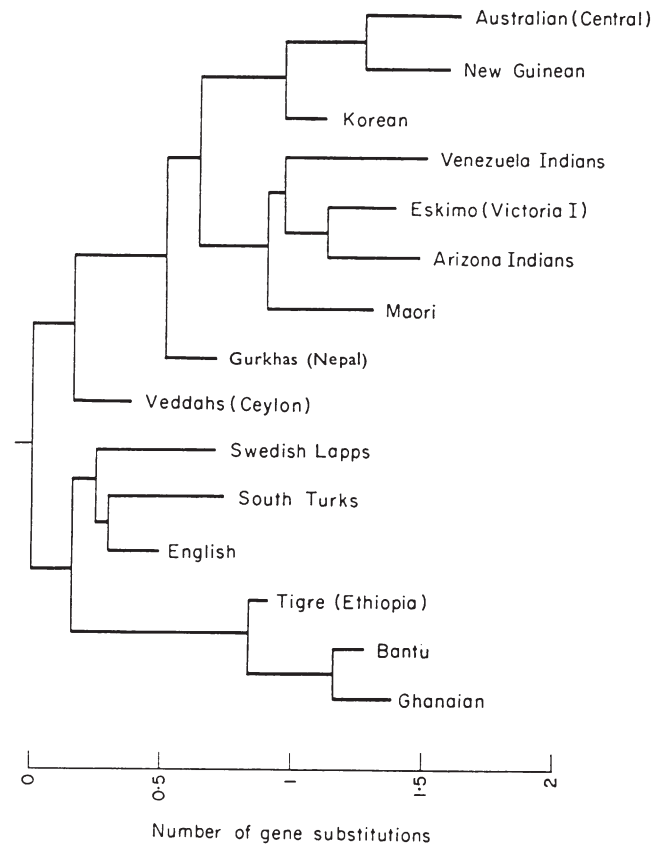


FIGURE 2.—The first numerical phylogeny produced by a combination of least-squares and parsimony methods (11th International Congress of Genetics, 1963, CAVALLI-SFORZA and EDWARDS 1965). The blood-group gene frequencies of the 15 populations used to produce Figure 2 have never been published but are available from the author.

The method of maximum likelihood: Just as I had not been entirely satisfied with Cavalli-Sforza’s additive-tree method, so he was not satisfied with the logic of my method of minimum evolution, and in truth neither was I. It just seemed a very neat computable idea that would give a sensible result. We both strove for something better, something more statistical, and during many discussions came to see that the basis of any statistical method comparable to, say, the estimation of recombination fractions in genetic linkage, would be a stochastic model for the evolutionary process whose parameters (including the form of the tree) we would then estimate. With this clarification, we chose a branching diffusion process as our model for evolution. We added a time dimension to the character space, and the challenge became one of estimating not only the form (topology) of the tree but the position in this space-time of each of its nodes. Naturally, the chosen estimation procedure was Fisher’s method of maximum likelihood. The development was first reported at the Systematics Association meeting in April 1964 (EDWARDS and CAVALLI-SFORZA 1964b).

THE RECONSTRUCTION OF EVOLUTION

A. W. F. EDWARDS and L. L. CAVALLI-SFORZA

*International Laboratory of Genetics and Biophysics, Naples,
Pavia Section: Istituto di Genetica, Universita di Pavia*

Darwin's theory of evolution invites us to believe that closely similar species are closely related. In cases in which there is no fossil record, or other means of following evolutionary history, this concept provides the only means of drawing inferences about the course of evolution, and it is important to see whether it can be defined sufficiently precisely to be used for estimating the most likely form of an evolutionary tree solely from data on the similarities and differences between species. A detailed consideration of this problem leads to the conclusion that the following principle may be employed :

THE PRINCIPLE OF MINIMUM EVOLUTION

The most plausible estimate of the evolutionary tree is that which invokes the minimum nett amount of evolution.

This principle will be discussed, the system of estimation to which it leads expounded, and some examples of its application given.

Thus, toward the end of 1963 was born the "likelihood approach" to phylogenetic analysis, although it was to be a long time before all the logical and computational difficulties were ironed out. I spent much of my 1964–1965 year at Stanford University struggling to extend the four-population prototype program to 15 populations, reporting on it at the November 1965 meeting of the Genetical Society in London. Eventually, the program EVOTREE was distributed, but it was soon overtaken by packages such as Felsenstein's PHYLIP. A copy of EVOTREE exists, having recently been found by Felsenstein. The subsequent development of the approach is considered briefly below.

In retrospect, the real contribution from Pavia in 1962–1964 was the embedding of modern phylogenetic analysis in the Fisherian tradition of stochastic model building and efficient statistical estimation by maximum likelihood and the use of genetical data. The detailed articles on this development are CAVALLI-SFORZA and EDWARDS (1966, 1967) and EDWARDS (1970); a number of other relevant contemporary publications are included in the list of references. Perversely, however, it was the earlier distance-matrix and minimum-evolution methods that caught the popular scientific imagination through being promoted by others after Cavalli-Sforza and I had introduced the

maximum-likelihood approach. The next section summarizes this.

CONTEMPORARY INTRODUCTIONS OF
DISTANCE-MATRIX AND
MINIMUM-EVOLUTION METHODS

As indicated in the quotation from K. O. May given in the Introduction, there is little point in priority-hunting when it comes to ideas. One can report only what happened and leave it at that. In his book *Inferring Phylogenies*, FELSENSTEIN (2004) devotes a chapter to history and philosophy that leaves no stone unturned. Distance-matrix methods using least squares became popular through the article by FITCH and MARGOLIASH (1967), who used a weighted variant of Cavalli-Sforza's method. I had discussed this method with Fitch during a visit to Madison, Wisconsin, in the summer of 1965.

The case of the method of minimum evolution is more interesting because it was clearly an idea whose time had come. It became popular under the name "parsimony" through the article by CAMIN and SOKAL (1965). Sokal in fact had heard Cavalli-Sforza deliver our paper in September 1963 at the 11th International Genetics Congress at which the method of minimum evolution was mentioned, and I discussed it with him

FIGURE 3.—The first announcement of the method of minimum evolution or parsimony (EDWARDS and CAVALLI-SFORZA 1963a).

afterward. Their use of parsimony, however, was independent, having originated with Camin (SOKAL 1966). Others who are known to have had the idea include ECK and DAYHOFF (1966); for further information, see EDWARDS (1996, 2004b).

It has occasionally been represented that the taxonomic methods of W. Hennig (1913–1976) involved an implicit use of the idea of minimum evolution on a tree. My own reading of his work failed to find this (EDWARDS 1996); for further views, see SOBER (1988), SNEATH (1995), and FELSENSTEIN (2004).

Two things should be said about the 1963 announcement of the method of minimum evolution (originally called a “principle”). First, it was so explicitly stated that one could—and did—immediately start thinking about computer programs to implement it, for both discrete and continuous data: “The most plausible estimate of the evolutionary tree is that which invokes the minimum net amount of evolution.” Second, in addition to its being presented at a meeting of Britain’s Genetical Society, the abstract containing the principle in full appeared in 1963 in two of the leading genetics journals, *Heredity* (Figure 3) and the *Annals of Human Genetics* (EDWARDS and CAVALLI-SFORZA 1963a). No earlier publication of such a principle has been noticed.

EPILOGUE: THE SUBSEQUENT DEVELOPMENT OF LIKELIHOOD METHODS

Distance-matrix and minimum-evolution methods enjoyed considerable popularity partly because they were essentially algorithmic and thus relatively easy to program for computers and perhaps also because they bore some resemblance to traditional taxonomic practice and were therefore more accessible to taxonomists. Even so, they were attempts to “find” a phylogenetic tree and were soon seen to be no substitute for the application of the theory of statistical estimation to the supposed stochastic model that had generated the tree in the first place.

The concept of likelihood was one of R. A. Fisher’s great contributions to statistics [for further information, see my book *Likelihood* (EDWARDS 1992)]. The method of maximum likelihood rapidly became, and has remained, the standard method for the estimation of statistical parameters, so it was natural for EDWARDS and CAVALLI-SFORZA (1964b) to propose its adoption in the case of phylogenetic trees. The initial difficulties of implementation (see especially CAVALLI-SFORZA and EDWARDS 1967) were surmounted only once the mistake of treating all the unknowns in the tree as parameters in the true sense was recognized (FELSENSTEIN 1968; EDWARDS 1970).

Progress, largely by Felsenstein himself starting with his Ph.D. thesis (FELSENSTEIN 1968), is best traced through his book *Inferring Phylogenies* (FELSENSTEIN 2004). Computationally feasible methods were developed for both continuous data such as gene frequencies

and discrete data such as DNA sequences [these of course were not available in 1964, but EDWARDS and CAVALLI-SFORZA (1964b) had already observed that in a discrete case the character space would be a lattice of points]. Computer packages incorporating likelihood methods were written and made widely available.

Some indication of the dominance that the statistical approach via maximum-likelihood and likelihood-based Bayesian methods has now achieved is in the book *Computational Molecular Evolution* by YANG (2006), where the author states in his preface:

The time appears ripe to summarize the methodological advancements in the field, and this book is such an attempt. I make no effort to be comprehensive in the coverage. There is hardly such a need now, thanks to recent publication of Joseph Felsenstein’s (2004) treatise, which has discussed almost everything relevant to phylogenies. Instead I take the view that molecular evolutionary analysis, including reconstruction of phylogenies and inference of the evolutionary process, is a problem of statistical inference (CAVALLI-SFORZA and EDWARDS 1967). Thus well-established methods such as likelihood and Bayesian are described as standard.

(For the avoidance of misunderstanding, this should not be taken to mean that Felsenstein’s view is any different: it was, after all, the main argument of his Ph.D. thesis.)

On the occasion of the presentation of 12 Darwin–Wallace Medals in 2008 by the Linnean Society (Professor Felsenstein being one of the recipients), Professor Mohamed Noor responded on behalf of all the medalists. He described his choice of “three major innovations [which] stand out as having played a major role in driving progress in understanding evolution these past 50 years”: (1) “the rise of the field of phylogenetic systematics”; (2) “the explosion of computational power and availability;” and (3) “the amazing growth of the field of genetics.” Of the first, he said:

We now know that physical or DNA-sequence similarity does not necessarily demonstrate closer relationship, and we apply quantitative formulae to sort relationships using shared-derived characters in particular. This field also heralded the application of model-based statistical methods for inferring species relationships, including both Bayesian and maximum likelihood based approaches. Overall, phylogenetic systematics is thriving and providing an irreplaceable framework that essentially didn’t exist for understanding how and when evolutionary processes operated in the past. (NOOR 2009)

I am indebted to Professor J. Felsenstein for his help, to Professor Cavalli-Sforza for answering several queries about times long ago, and, as ever, for a collaboration of consuming interest.

LITERATURE CITED

- For the sake of historical completeness, this list includes some early references not mentioned in the text.*
 BAILEY, N. T. J., 1961 *Introduction to the Mathematical Theory of Genetic Linkage*. Clarendon Press, Oxford.
 BOX, J. F., 1978 *R. A. Fisher, The Life of a Scientist*. John Wiley & Sons, New York.

- BOYD, W. C., 1952 *Genetics and the Races of Man*. Little, Brown, Boston.
- CAMIN, J. H., and R. R. SOKAL, 1965 A method for deducing branching sequences in phylogeny. *Evolution* **19**: 311–326.
- CAVALLI-SFORZA, L. L., 1966 Population structure and human evolution. *Proc. R. Soc. Lond. B Biol. Sci.* **164**: 362–379.
- CAVALLI-SFORZA, L. L., 1990 Recollections of Whittingehame Lodge. *Theor. Popul. Biol.* **38**: 301–305.
- CAVALLI-SFORZA, L. L., 1991 Genes, peoples and languages. *Sci. Am.* **265**(5): 72–78.
- CAVALLI-SFORZA, L. L., and F. CAVALLI-SFORZA, 2005 *Perché la scienza*. Mondadori, Milan.
- CAVALLI-SFORZA, L. L., and A. W. F. EDWARDS, 1965 Analysis of human evolution. Proceedings of the 11th International Congress of Genetics, The Hague 1963, *Genetics Today* **3**: 923–933. (N.B. *The year of this paper is often given as 1964, which is indeed the date on the offprints.*)
- CAVALLI-SFORZA, L. L., and A. W. F. EDWARDS, 1966 Estimation procedures for evolutionary branching processes. *Bulletin of the International Statistical Institute (Proceedings of the 35th session)* **41**: 803–808.
- CAVALLI-SFORZA, L. L., and A. W. F. EDWARDS, 1967 Phylogenetic analysis: models and estimation procedures. *Evolution* **21**: 550–570 and *Am. J. Hum. Genet.* **19**: 233–257.
- CAVALLI-SFORZA, L. L., I. BARRAI and A. W. F. EDWARDS, 1964 Analysis of human evolution under random genetic drift. *Cold Spring Harb. Symp. Quant. Biol.* **29**: 9–20.
- ECK, R. V., and M. O. DAYHOFF, 1966 *Atlas of Protein Sequence and Structure 1966*. National Biomedical Research Foundation, Silver Spring, MD.
- EDWARDS, A. W. F., 1961 The population genetics of sex-ratio in *Drosophila pseudoobscura*. *Heredity* **16**: 291–304.
- EDWARDS, A. W. F., 1965 The reconstruction of evolution: estimation by maximum likelihood. (Abstract). *Genetical Society programme*, November. 1965.
- EDWARDS, A. W. F., 1966 Studying human evolution by computer. *New Sci.* **30**: 438–440. Reprinted in *Evolution*, edited by M. RIDLEY, 1997. Oxford University Press, Oxford.
- EDWARDS, A. W. F., 1969a Genetic taxonomy, pp. 140–142 in *Computer Applications in Genetics*, edited by N. E. MORTON. University of Hawaii Press, Honolulu.
- EDWARDS, A. W. F., 1969b Genetic taxonomy. (Abstr.) *Biometrics* **25**: 610.
- EDWARDS, A. W. F., 1970 Estimation of the branch points of a branching diffusion process. *J. R. Stat. Soc. Series B* **32**: 155–174.
- EDWARDS, A. W. F., 1990 Fisher, W, and the fundamental theorem. *Theor. Popul. Biol.* **38**: 276–284.
- EDWARDS, A. W. F., 1992 *Likelihood*, Expanded Edition. Johns Hopkins University Press, Baltimore.
- EDWARDS, A. W. F., 1996 The origin and early development of the method of minimum evolution for the reconstruction of evolutionary trees. *Syst. Biol.* **45**: 79–91.
- EDWARDS, A. W. F., 2001 La ricostruzione degli alberi evolutivi, pp. 23–30 in *Le radici prime dell'Europa*, edited by G. BOCCHI and M. CERUTI. Mondadori, Milan.
- EDWARDS, A. W. F., 2003 Human genetic diversity: Lewontin's fallacy. *BioEssays* **25**: 798–801.
- EDWARDS, A. W. F., 2004a Mendelism and man 1918–1939, pp. 33–46 in *A Century of Mendelism in Human Genetics*, edited by M. KEYNES, A. W. F. EDWARDS and R. PEEL. CRC Press, Boca Raton, FL.
- EDWARDS, A. W. F., 2004b Parsimony and computers, pp. 181–190 in *Milestones in Systematics*, Special Volume Series 67, edited by D. M. WILLIAMS and P. L. FOREY. Systematics Association, London.
- EDWARDS, A. W. F., 2005 R. A. Fisher, Statistical Methods for Research Workers, pp. 856–870 in *Landmark Writings in Western Mathematics, 1640–1940*, Ed. 1, edited by I. GRATTAN-GUINNESS. Elsevier Science Publishers, Amsterdam.
- EDWARDS, A. W. F., and L. L. CAVALLI-SFORZA, 1963a The reconstruction of evolution. (Abstr.) *Heredity* **18**: 553, and *Annals of Human Genetics* **27**, 104–105.
- EDWARDS, A. W. F., and L. L. CAVALLI-SFORZA, 1963b A method for cluster analysis. Preprints of the 5th International Biometrics Conference, Cambridge, UK, September, 1963.
- EDWARDS, A. W. F., and L. L. CAVALLI-SFORZA, 1964a A method for cluster analysis. (Abstr.) *Biometrics* **20**: 383.
- EDWARDS, A. W. F., and L. L. CAVALLI-SFORZA, 1964b Reconstruction of evolutionary trees, pp. 67–76 in *Phenetic and Phylogenetic Classification*, Systematics Association Publication No. 6, London. Reprinted in *Cladistic Theory and Methodology*, edited by T. DUNCAN and T. F. STUESSY, 1985. Van Nostrand Reinhold, New York.
- EDWARDS, A. W. F., and L. L. CAVALLI-SFORZA, 1965 A method for cluster analysis. *Biometrics* **21**: 362–375.
- FELSENSTEIN, J., 1968 Statistical inference and the estimation of phylogenies. Ph.D. Thesis, University of Chicago, Chicago.
- FELSENSTEIN, J., 2004 *Inferring Phylogenies*. Sinauer, Sunderland, MA.
- FISHER, R. A., 1925 *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh.
- FISHER, R. A., 1950 Gene frequencies in a cline determined by selection and diffusion. *Biometrics* **6**: 353–361.
- FISHER, R. A., 1956 Blood-groups and population genetics. Proceedings 1st International Congress of Human Genetics. *Acta Genet. Stat. Med.* **6**: 507–509.
- FISHER, R. A., 1959 Natural selection from the genetical standpoint. *Aust. J. Sci.* **22**: 16–17.
- FITCH, W. M., and E. MARGOLISH, 1967 Construction of phylogenetic trees. *Science* **155**: 279–284.
- GOMBERG, D., 1968 “Bayesian” postdiction in an evolution process. Istituto di Genetica, Pavia. (*Unpublished.*)
- HAECKEL, E., 1876 Frontispiece, *History of Creation*, Vol. II. King, London.
- HEYWOOD, V. H., and J. MCNEILL, 1964a Phenetic and phylogenetic classification. *Nature* **203**: 1220–1224.
- HEYWOOD, V. H., and J. MCNEILL (Editors), 1964b *Phenetic and Phylogenetic Classification*, Systematics Association Publication No. 6, London.
- MATHER, K., 1951 *The Measurement of Linkage in Heredity*, Ed. 2. Methuen, London.
- MAY, K. O., 1975 Historiographic vices. II. Priority chasing. *Historia Mathematica* **2**: 315–317.
- MICHENER, C. D., and R. R. SOKAL, 1957 A quantitative approach to a problem in classification. *Evolution* **11**: 130–162.
- MOURANT, A. E., 1954 *The Distribution of the Human Blood Groups*. Blackwell, Oxford.
- MOURANT, A. E., 1995 *Blood and Stones: An Autobiography*. La Haule, Jersey, Channel Islands.
- NOOR, M., 2009 (In) Presentation of the Darwin–Wallace Medals 2008. *Linnean* **25**(2): 3–9.
- RACE, R. R., and R. SANGER, 1950 *Blood Groups in Man*. Blackwell, Oxford.
- SNEATH, P. H. A., 1995 Thirty years of numerical taxonomy. *Syst. Biol.* **44**: 281–298.
- SNEATH, P. H. A., and R. R. SOKAL, 1962 Numerical taxonomy. *Nature* **193**: 855–860.
- SOBER, E., 1988 *Reconstructing the Past*. MIT Press, Cambridge, MA.
- SOKAL, R. R., 1966 Numerical taxonomy. *Sci. Am.* **215**(6): 106–116.
- SOKAL, R. R., and P. H. A. SNEATH, 1963 *Principles of Numerical Taxonomy*. Freeman, San Francisco.
- VENN, J. (Compiler), 1898 *Biographical History of Conville and Caius College*, Volume II, Cambridge University Press, Cambridge, UK.
- WILKES, M. V., 1985 *Memoirs of a Computer Pioneer*. MIT Press, Cambridge, MA.
- YANG, Z., 2006 *Computational Molecular Evolution*. Oxford University Press, Oxford.