

**Practice-oriented controversies and borrowed epistemic support in  
current evolutionary biology: phylogeography as a case study.**

**Alfonso Arroyo-Santos**

**Posgrado en Filosofía de la Ciencia, IIF-UNAM**

**agripas@gmail.com**

**Mark E. Olson**

**Instituto de Biología, UNAM**

**molson@ibiologia.unam.mx**

**Francisco Vergara-Silva**

**Instituto de Biología, UNAM**

**fvs@ibiologia.unam**

Although there is increasing recognition that theory and practice in science are often inseparably intertwined, discussions of scientific controversies often continue to focus on theory, and not practice or methodologies. As a contribution to constructing a framework towards understanding controversies linked to scientific practices, we introduce the notion of *borrowed epistemic credibility* (BEC), to describe the situation in which scientists exploit fallacious similarities between accepted tenets in other fields to garner support for a given position in their own field. Our proposal is based on the analysis of a recent controversy in phylogeography, a biological subdiscipline concerned with the study of the historical causes of variation in genetic diversity within species in concrete biogeographical locations. Through a review of the arguments that support the two conflicting phylogeographic schools, we show that 'theory' plays essentially no role as a foundation of the controversy, whereas both sides borrow epistemic credibility from sources such as formal logic, similarity of results to those in other scientific areas, the authority of prominent scientists, or the presumed superiority of quantitative vs. verbal reasoning. Our case study underscores the indivisibility of theory and practice and provide a means to re-examine important philosophical issues such as the meaning of inference, rationality, justification, and objectivity in scientific practice.

Keywords: scientific practice, scientific controversy, epistemic credibility, phylogeography, statistical inference, Bayesianism, Popperianism

## 1. Introduction

As scientists turn increasingly to computer simulations and statistical inferences to support their research agendas, many recent scientific controversies are centered on practical rather than theoretical issues. From a philosophy of science perspective, this feature of contemporary scientific research has been noticed by Winsberg, who claims that in computer simulation studies any underlying theory is just one of the many factors involved in the outcome of the simulations and, as a consequence, theoretical knowledge may bear no direct substantive relationship to the knowledge generated by the simulation (see Winsberg 1999, 2006, 2010). Removal from theory is likewise far in the case of statistical inferences, because the generation of results is left in the hands of tests that were not developed for dealing with the specific situation being studied. With the ever more frequent use of simulations and statistical inferences, practical controversies are certain to be ever more frequent.

Nevertheless, many philosophers of science still treat scientific controversies as though they mostly occurred at the theoretical level. For example, in a frequently cited study, McMullin (1987) called theoretical disputes "the commonest source of controversy in science" (p. 66). With a similar emphasis on theory, Dietrich and Skipper (2007) treated ways in which scientists exploit underdetermination to favor particular views by "playing off the multidimensional framework of theory evaluation" (p. 295). Although studies of disputes at the theory level are important, placing so much emphasis on theories may obscure that most controversies are likely linked to the interpretation of results obtained by means of methodologies, modeling tools, and other similar resources –in short, with scientific practice.

In this paper we present the case of a scientific controversy taking place at the level of methodologies where two competing groups in the field of phylogeography apparently debate on what is the better inference tool for their field. We say apparently because philosophical analysis shows that this practical controversy centers not so much on the methods themselves but on the issues they embody. Here we will give an example of how scientists address a methodological

controversy not by appealing to theory but instead to an array of extratheoretical considerations.

In contrast to other metascientific analyses that have often highlighted how scientists refer to *epistemic virtues* such as simplicity, predictive power, or internal coherence to choose between competing theoretical alternatives (for a review see Pournari 2008, Nola and Sankey 2007, Quine and Ullian 1978 ), in our case study, epistemic virtues alone have not been appealed to. Instead, competing authors are predominantly searching for what we call *epistemic credibility*. In the context we analyze here, epistemic credibility is a means to justify a practice in one field by appealing to a set of values, norms, virtues or practices that are considered as the signature of good science in similar fields. Because scientists seek to transfer merits of acceptance effective in other fields to their own, we refer to the *borrowing* of epistemic credibility. We claim that the idea of *borrowed epistemic credibility* (BEC) is useful to understand how a scientific controversy that could seem to be a simple disagreement over fine points of method in fact embodies deep disagreements over what constitutes acceptable scientific practice.

We illustrate the role of BEC in a controversy currently taking place within a sub discipline of evolutionary biology called phylogeography. Phylogeography has been described by one of its main proponents as a geography-based bridge between phylogenetics and population genetics (Avice 2009; see also Hickerson et al. 2010), provides an outstanding example of an area heavily driven by methodologies, in many cases computer programs with ever-increasing degrees of sophistication (e.g. Posada and Crandall 2001). As a case study, phylogeography is ideal for our purposes, because theory has only been marginally involved in the specifics of the conflict between the two main stances currently dominating the field, with scientists instead largely borrowing epistemic credibility from resources that do not belong to phylogeography *per se*. The current controversy within phylogeography is also ideal for illustrating the distinct way in which BEC is used, *vis-à-vis* the more widely acknowledged role of epistemic virtues.

### **Borrowed epistemic credibility (BEC)**

The basic idea of borrowed epistemic credibility (BEC) is that for a number of reasons, there are cases in which scientists cannot appeal to empirical arguments to justify their theories, methodologies, or even their results and consequently they must appeal to other non-empirical elements. In our case study, scientists appeal to a large set of elements that may include epistemic virtues but also norms, values, practices, and other factors considered important in their field or in similar areas of research. The elements that scientists select for these sets are important because they represent what the members of the field mean by objectivity, rationality, or other considerations that a given community takes as the standards of 'good scientific practice'. Usually, scientists seek to justify their theories, methodologies, or results by connecting them in some way with these sets of considerations. In the case of phylogeography, where the debate has centered on methodologies, scientists defending rival methodologies can appeal to epistemic credibility coming from ecology, population genetics, biogeography or molecular biology, to name some examples. We will call phylogeography the *target field*, whereas ecology, population genetics, and molecular biology will be called the *source fields*. Methodologies coming from the target field need not exhibit the elements borrowed from the source fields. The key point of our analytical proposal is that, when borrowing credibility, scientists construct rhetorical devices to claim that if the defended methodology resembles in any sense of the term methodologies, theories, models or results coming from the source fields, then these *should somehow also exhibit the virtues of the target fields*. According to our analysis, when the defended methodology has successfully

borrowed epistemic credibility from the source fields, it has also been effectively –though artificially– endowed with virtues that in turn justify the results obtained with those methodologies.

An issue concerning *virtues* is worth clarifying at this point. In the context of this paper, we will speak of virtues when making reference to the members of a set formed by epistemic virtues, norms, practices, tacit knowledge and other elements associated to what a particular community regards as objective, scientific, rational, and in general, what it takes to be the standards of ‘good science’. Second, epistemic credibility is not borrowed from an abstract set of elements but from prestigious work or from prestigious individuals whose work could be seen as exemplifying some important virtue. For example, if A is a prestigious scientific research coming from a source field and B is a work in search of credibility, and if it can be said that somehow there is a similarity between A and B, then borrowing epistemic credibility amounts to saying that if A and B are comparable, then B also exhibits A’s virtues, which in turn justifies that members of the community accept B. Note that the burden of proof for B lies entirely on the argument constructed around the borrowing of epistemic credibility. As a result, we claim that borrowing epistemic credibility should be viewed as a fallacy because if B truly had the imputed virtues, then the comparison with A would serve to highlight B’s particular virtues, and therefore B would not need to borrow credibility–epistemic or otherwise– from elsewhere.

The similarities invoked between scientific research A and B can be of different types. For example, robustness can be considered one type. B borrows epistemic credibility in virtue of its results being the same or analogous to those obtained in A. Another category is bootstrapping, in which the results in B are not the same but resemble the kind of results generated by A. Appeals to logical inference may also be a kind of borrowed credibility, when B’s results are justified because they conform to a notion of logical inference represented by A. In a similar way, quantification might be regarded as another kind of borrowed credibility, when quantitative results in A are presented as superior to qualitative ones obtained elsewhere. In evolutionary biology, this is a very important source of epistemic credibility given that much research has traditionally been qualitative. In the case of phylogeography, one controversial method earned its popularity in being the first available quantitative method. Finally, epistemic credibility may also be obtained from sources such as tradition or appeals to the authority of prominent scientists whereby the results obtained using a certain methodology are justified because an expert or a group of individuals regarded as experts in their field back a given methodology. This is only a short list of possible sources of epistemic credibility, but it could be argued that the same strategy can be used to justify theories, models, or hypotheses. In what follows, we will show how phylogeography illustrates biologists’ deployment of BEC to advance their positions.

Before developing our philosophical analysis, we briefly introduce the basics of phylogeography through a review of the centerpiece of phylogeographic theory –the so-called *coalescent*– and illustrate how this otherwise important theoretical component has no say in the current debate. Then, we discuss more specifically the current controversy surrounding a specific phylogeographic method called *nested clade phylogeographic analysis* (NCPA) and illustrate how a controversy that is apparently restricted to details of methodologies in fact reflects different points of view regarding how science is to be carried out –that is, regarding scientific practice. We then argue that phylogeographic theory is a bystander in the phylogeographic controversy, and that the arguments in this controversy largely rest on different sources of BEC.

## **2. Genetics and geography: the basic goals of phylogeography**

Phylogeography was born in the late 1980s as an attempt to unify the fields of phylogenetics and

population biology in an explicitly geographical context. The discipline was christened in a 1987 paper (Avise et al. 1987; see also Avise 2009; Hickerson et al., 2010) in which the authors stated that "... (a) phylogenetic interrelationships among [segments of DNA] and (b) geographic distributions of the phylogenetic groupings... constitute concerns of a discipline that might be termed intraspecific phylogeography." (pp. 516-517). Defined this way, phylogeographic studies take ways of thinking about the evolutionary relationships between species, known as phylogeny, and apply them within species; the 'geography' suffix alludes to the role of the analysis of the spatial distribution of genetic variation. Empirical phylogeographic studies start with the gathering of DNA sequence data from as many individuals and from as many populations as possible from the species of interest. These data are then analyzed with a variety of algorithms to sort the samples into groups via similarity criteria (Avise 2009, Posada and Crandall 2001; Woolley et al. 2008). Once a hypothesis of the relationships between the populations of a species has been reconstructed, it is then possible to assess the distribution of genetic variation with respect to geography. For example, Avise (1992) detailed a case in which a suite of coastal animals including horseshoe crabs, oysters, sparrows, sea bass, and a species of coastal turtle, all with ranges spanning both the Atlantic coast as well as the US portion of the Gulf of Mexico, showed strong divergence in genetic variants between the Gulf and the Atlantic. The authors found that, in all of these species, the populations in the Gulf shared more similarities with the other Gulf populations than with those in the Atlantic and vice versa. The associations between geography and genetics are used to make causal inferences regarding the relative role of "recurrent processes," such as continual genetic interchange, as opposed to unique events such as the rising of a mountain range (Fig. 1).

Although the goals of phylogeographic studies are similar, there are marked differences in the methods used. Before treating in more detail the current debate in phylogeography, we show how the methods-driven nature of the field leads to controversies that take place at a remove of theory. To that end, we briefly review the key theoretical element of phylogeography generally referred to as the coalescent.

### **3. The coalescent**

The theory of coalescence was developed independently by Kingman (1982), Hudson (1983), and Tajima (1983) with the aim of tracing present day genetic lineages back in time to their most recent common ancestor. Coalescence is the reverse of divergence: as we move forward in time, we can think of an individual DNA molecule replicating and siring two new lineages, an event called a divergence. Looking backward in time, whenever two lineages merge into the same ancestor, we say that they "coalesce." Coalescent theory translates this intuition of coalescence into a series of idealized mathematical models whose main goal is to calculate the time elapsed between the most recent common ancestor and the genetic variants found in present day populations. Given the highly idealized nature of the models, scientists modify some of the premises of the coalescent model to approximate results from real-life populations. For example, coalescent times are expected to be much longer for genetic loci that recombine than for those that do not (McVean and Cardin 2005). Basic versions of the coalescent assume that all members of a population are equally fit, a condition never observed in nature. In response, variation in factors such as survivorship or fecundity is included in the model (O'Fallon et.al. 2010). By the same token, levels of genetic diversity can be taken into account in calculating population size, wherein a small population with high genetic diversity could be considered as representing more individuals in a genetic sense than a very large but genetically uniform population (Nei and Takahata 1993). Similarly, different mutation rates will also be associated with different coalescence times *ceteris paribus*. All of these modifications are introduced to make the models as close to the actual world

as possible, but as models become more realistic, the resulting algorithms become less tractable as the number of possible evolutionary scenarios that can be depicted increases enormously with each new variable (e.g. McVean and Cardin 2005). It is in this context that methods that apply the coalescent to phylogeography proliferate, as each emphasizes different combinations of realism, generality, and inferential power. In the next section we will review the controversy that has arisen between the two most popular of such methods.

#### **4. Coalescent theory in nested clade phylogeographic analysis versus the approximate Bayesian computation approach**

Nested clade phylogeographic analysis (NCPA) is the brainchild of American geneticist Alan Templeton (Templeton 1987, 1998, 2001, 2002, 2006, 2008, 2010a; Templeton et al. 1995) and is widely considered to be the most popular phylogeographic method (Petit 2008a, 2008b). NCPA starts with the sequencing of DNA regions from samples from throughout the range of a given species and estimating their genealogical relationships. The resulting branching diagrams are known as 'haplotype networks' (Fig. 2)<sup>1</sup>. The "nested clade" aspect of NCPA refers to a step in which the haplotypes are arranged into a nested hierarchy of groups that are separated by one or more mutations<sup>2</sup>. The procedure then calculates by null-hypothesis testing a pair of indices that reflect expectations regarding how widely haplotypes are distributed, how abundant they are, and how far they have moved historically. Significantly large or small indices are taken as indicating processes such as geographical range expansion or shifts in the center of distribution. To interpret the indices, Templeton et al. (1995) provide an inference key that forms the basis for all NCPA studies. The inferences drawn with the key are supposed to be deductive consequences of what should obtain given the general expectations of coalescent theory. The idea is that different scenarios, e.g. panmixia, restricted gene flow with isolation by distance, sudden range expansion, long-distance dispersal, or events that divide one large population in two, should produce differing patterns of significance of the main NCPA indices (see Templeton 2009a, 2009b).-

Detractors of NCPA charge that phylogeographic inferences are best evaluated by a family of Bayesian-based statistical models, the most popular of which is called approximate Bayesian computation (ABC). To test phylogeographic hypotheses with ABC, a scientist proposes a small set of hypotheses that might account for the geographical distribution of genetic variation observed today. Given the data set, which consists of DNA sequences, the evolution of the set of samples is modeled back in time. Different values for variables such as population size and subdivision, number of generations, mutation and migration rates, and other factors of interest, are built into the model. These factors influence the rate of sorting processes, and thus how likely two haplotypes are to coalesce into one. Because sorting processes are stochastic, the exact topologies of the genealogies in each round of modeling will often differ. By repeating the model many times, it is possible to see how often results support or contradict a given hypothesis (for a more detailed discussion on Bayesian methods see Beaumont et.al. 2010, Nielsen and Beaumont 2009, or Huelsenbeck 2001).

##### **4.1 The coalescent as a bystander in the controversy**

Coalescent theory is reflected in very different ways in NCPA and the ABC approaches. However, neither side in the controversy has questioned the very different ways in which the coalescent is shaped into statistical expectations, e.g. that the NCPA key's assumptions are flawed, or that genetic lineage divergence/coalescence should not be modeled. Templeton, the author of NCPA, goes so far as to advocate the use of both approaches (Templeton 2009a). In general, criticism of NCPA is directed at its statistical aspects. For example, Knowles and Maddison's (2002) major

concern about this topic is that, in their view, single-locus NCPA ignored the randomness associated with the coalescent process and the accumulation of mutations. In other words, their concern has to do not with coalescent theory, but with the statistical error in the inference. Knowles and Maddison (2002) ask “Does NC[P]A then fill the role desired of a flexible, statistical inference procedure?” (p. 2631) and answer that NCPA “does not attempt to distinguish statistically among alternative interpretations, nor does it provide an estimate of the uncertainty in its conclusions.” They contend that, “for any interpretation derived from Templeton’s inference key, we cannot ascertain the confidence limits on the reconstructed history, whether they are so broad as to include many unconsidered alternatives, or if an alternative hypothesis would be almost equally well supported by the data.” (p. 2624).

Similarly, in the most recent critique of NCPA, Panchal and Beaumont (2010) identify numerous problems having to do with the statistical analysis of NCPA but not with any underlying phylogeographic theory. Rather than finding errors in the way that NCPA flows from coalescent theory, Panchal and Beaumont follow authors such as Knowles and Maddison (2002) in asserting that the method suffers from high false-positive rates. Other statistical criticisms include that the method’s testing approach cannot distinguish between scenarios such as restricted gene flow caused by isolation by distance or caused by long-distance dispersal (see pp. 15-17). Regarding the coalescent, they only say that the theory for dating clades in NCPA is based on the standard coalescent but is then applied to scenarios that explicitly do not correspond to it (p.15). Therefore, the only reference to coalescent theory has nothing to do with disagreements regarding the theory itself but with its statistical implementation. Similarly, Templeton’s worries also have everything to do with methodology. In a recent letter, he said that his “main objection to ABC was that it can produce posterior ‘probabilities’ that are not true probabilities” (Templeton 2010b: 488) and that “the potential of ABC is currently not realized because of serious statistical and mathematical flaws” (ibid: 489).

Remarkably, mention of theory in criticisms of NCPA only appears in Beaumont et al. (2010), a good eight years after the critique of Knowles and Maddison (2002), which began the controversy. However, the criticism of Beaumont et al. (2010) has to do not with theory but with the specific assumptions regarding population demographics that a given coalescent equation can apply to, arguing that the equation used in NCPA is too simple for the situations in which it is applied. Here we find no real disagreement that can be considered one of theory; the authors are not pitting different derivations of the coalescent against one another. Instead, they are arguing over what level of biological realism needs to be reflected in the jump from the idealized models of coalescent theory to dealing with real genetic data.

To us, Panchal and Beaumont and Templeton’s conclusions show that the controversy in phylogeography has nothing to do with theory but with methodological issues of both NCPA and ABC. If the problem were the coalescent, discussion would be centered not on the kinds of things NCPA or ABC cannot do but in the things that each solves to show how one model does not follow from theory as the other one does. Instead, all the discussion is centered on the statistical limitations of both modeling approaches, which is why, in the end, we can say that the coalescent has no saying in this debate.

## **5. The nested controversies: the sources of epistemic credibility.**

Because phylogeographers spend a great deal of time discussing the pros and cons of statistical

modeling approaches, the phylogeography debate might seem to be simply a repeat of the long controversy between frequentist and Bayesian interpretations of probability (i.e. Aldrich 2008, Aldrich 2002, Howie 2002, Efron 1986) in a phylogeographic context. For example, in a recent review of the phylogeographic controversy, Bloomquist and collaborators say that the debate “has focused around two general points: (i) does single-locus and multi-locus NCPA have an inherently high false-positive rate and does this preclude its use? and (ii) do model-based methods or NCPA provide a more appropriate way to analyze phylogeographic data?” (Bloomquist 2010, 627). If the NCPA controversy were simply a methodological dispute then it would be resolved when both methods were rigorously tested, showing that one is more reliable than the other. However, this is not likely to happen any time soon because the very notions of what would constitute proof and the processes to generate this proof differ between the two communities. NCPA proponents lean toward a sort of deductivism and supporters of ABC toward a more abductive approach. Templeton’s method carries out a set of comparisons between the observed geographic and genealogical distributions of genetic variation. The statistical outcome of NCPA is a set of yes-no decisions, which is given biological meaning in the context of the inference key. ABC instead compares the computer modeling of different evolutionary scenarios and among these uses statistical evaluations to decide which seems the likeliest among the alternatives modelled. While none of these approaches can be unequivocally categorized in any of the typical categories of inference, they clearly show different inclinations regarding notions of satisfactory scientific inference.

Our analysis shows that the debate actually has nothing to do with quarreling interpretations of probability because competing groups have never said that either null-hypothesis testing or Bayesian modeling are wrong in any statistical sense (remember that Templeton actually advocates using both approaches). Given that there are also no real theoretical differences between the two competing groups (meaning that, as we have shown, the coalescent plays no relevant role), and that the controversy cannot be settled by the results generated by the very methodologies in dispute, we claim the controversy actually takes place in the sources of epistemic credibility that have been used to defend either the NCPA or ABC approach.

In what follows we present some examples to show how both ABC and NCPA supporters defend their positions not on the intrinsic value of their methodologies but on a set of extrinsic virtues. In doing so, phylogeographers argue that method x should be preferred because it represents what the members of the community (of phylogeographers, evolutionary biologists, or biostatisticians) take for good science.

One of the most popular sources of epistemic credibility is what we have called logical inference, meaning an appeal to what each phylogeographer understands as a valid logical inference. For example, both sides seem to agree on Popperian falsificationism as an ideal of scientific inference. NCPA has been defended as a Popperian approach because it successively subjects null hypotheses to rejection (Templeton 2009). Templeton charges that, if the pool of all plausible hypotheses is considered, then successively rejecting competing hypotheses will lead to a “strong” inference (in the sense of Platt 1964, see also Chamberlin 1890, Beard and Kushmerick 2009). In contrast to NCPA, because ABC can only compare the relative fits of a small number of modeled scenarios, it cannot be considered strong scientific inference. The conclusion drawn is that NCPA should be preferred because it conforms to the rules of (some sort of) formal logic, e.g. when Templeton mentions that “the statistics or probabilities used to measure the goodness of fit of the models obey the constraints imposed by formal logic” (2010b: 6376). Supporters of ABC do not refute the view of Popperian falsificationism as a valid approach, and indeed seem to share it with supporters of NCPA. Instead, they accuse NCPA of being inductive, traditionally presented in evolutionary



biology as the antithesis to Popper and indeed to science (Mayr 1982), such as when Beaumont and Panchal (2008, p. 2564) say that "Templeton (2008) cites Popper (1959) in support of the NCPA approach against model-based statistical analysis. However, we would suggest that although NCPA consists of a large number of hypothesis tests based on permutation methods, in the end it follows an inductivist paradigm of trying to derive a general explanation directly from the data... By contrast to NCPA, in model-based analysis [such as the ABC approach], one model is pitted against another in the face of the data, and this, surely, is a more valid scientific approach" (Beaumont and Panchal 2008: 2564).

Another variant is disagreement over the value of verbal versus quantitative reasoning. For example, Beaumont and Panchal (2008) charged that "A verbal, reasoned, argument is presented in Templeton et al. (1995) to justify the method, and the inferences it makes, not dissimilar in style and authority to the *Corpus Aristotelicum*. The authors of 265 papers that have used NCPA are, in a sense, appealing to this authority. One needs to ask: is this science?" (p. 2564). Appealing to formal logic or to one type of reasoning as superior to another is not a derivation from theory but instead represents a borrowing of epistemic credibility from what phylogeographers understand as a valid logical inference.

Other examples of epistemic credibility include notions of robustness, which refers to approaches that repeat earlier conclusions, as well as familiarity, which refers to those that produce results congruent with prevailing views. Examples can be found in Templeton's critique of Fagundes's use of ABC to contrast three different models of human evolution (Templeton 2008, 2009, 2010). Templeton charges that one of Fagundes's models is contrary to the prevailing notion of isolation by distance between humans living in Eurasia and Africa, noting that "it is patent that the parameter values chosen by Fagundes et al. (2007) are strongly discrepant with the empirical data on autosomal coalescent times" (Templeton 2009, p. 323). Templeton invokes robustness and familiarity of the results, alluding to other controversies in evolutionary biology to show that ABC does not reach the conclusions generated in similar fields. Again, robustness and familiarity are not theoretical issues but instead discredit rival positions by showing that they are inconsistent with accepted results.

As a final instance, epistemic credibility is frequently borrowed from different representations of authority. A conspicuous example is Beaumont et al. (2010), in which 22 authors unite in a single paper to express their reserves regarding NCPA and their support of ABC. The message of this surfeit of authors would seem to be that the endorsements of many scientists against one approach implies that it is incorrect. A similar implication that solitariness is associated with the incorrect position is when Beaumont and Panchal (2008, p. 2563) note that "there is a disagreement between Templeton (2004, 2008), who suggests the method works well, and three independent groups (Knowles & Maddison 2002; Petit & Grivet 2002; Panchal & Beaumont 2007), who believe that they have demonstrated that it does not. As far as we are aware, there are currently no publications other than those of Templeton and co-workers to support the accuracy or efficacy of NCPA." The preceding examples are appeals to the authority implied by consensus among many scientists. Another approach is to appeal to the authority of a single prominent figure, such as the statement of Knowles (2008 p. 2712), when she refers to authors who voice "other concerns over the validity of NCPA's inferences." Among these authors she cites Felsenstein (2004), one of the leading developers of the methods used for reconstructing the evolutionary relationships of organisms. Given his prominence in systematic biology, his verdict against NCPA would be weighty. However, in Felsenstein (2004), the only reference to NCPA we could locate was on p. 484: "A more statistical approach was taken by Templeton (1998), using the nested clade analysis tree reconstruction methods introduced earlier by Templeton et al. (1988). Although well-

defined enough to be implemented by computer programs (Clement, Posada, Crandall, 2000; Posada, Crandall, and Templeton, 2000), these methods do not attempt to take into account the uncertainty of the estimate of the tree, and there has been little study of their statistical properties. A notable exception is the paper by Knowles and Maddison (2002). Although the need to use manual steps in the analysis limited the number of replications they could make, they found that the single-tree approach was problematic." In contrast to the assertion of Knowles (2008), Felsenstein does not express anything that can be construed as "other concerns" beyond what had already been discussed in the literature. As a result, we can only interpret Knowles's cite of Felsenstein as an attempt to borrow credibility from his authority to shore up her position. Finally, supporter of ABC Petit (2008a, p. 1404) appeals to the authority of institutions when he concludes that "...the results of Panchal & Beaumont (2007) convince me that reputable journals should (i) discourage the use of the NCPA method for single locus data sets..., and (ii) still be suspicious of NCPA analyses based on multiple loci," the subtle aspersion being that any journal that henceforth publishes an NCPA study should be held in suspicion.

We trust that with these examples of appeals to logical inference, robustness, familiarity, and to authority, we have illustrated ways that scientific controversy can take place at a remove from theory, and that a major element of the phylogeography controversy is the borrowing of epistemic credibility. Moreover, our survey suggests that authors defend not only a particular methodology but perhaps more importantly, a vision of what phylogeography and indeed science ought to be.

## **6. Epistemic credibility and the theory-practice divide**

Thinking of theory as standing separately from practice has a long tradition in the philosophy of science. As a result, philosophers of science have an extensive battery of conceptual tools at their disposal to think about the theory-practice relation. The vocabulary of laws and axioms, of the distinction between models and theories, syntactic and semantic relations to the world, as well as notions of theories as explanatory devices all stem from the conception of theory as being in some way distinct from practice. The view of theory as interacting inextricably with practice has less of a tradition, and because of the complexity of this view, much conceptual machinery remains to be built (see for example Burian 2005, Keller 2002, Rheinberger 1997, Pickering 1995). Here we offer the notion of *borrowed epistemic credibility* (BEC), which we believe can help better understand how theory and practice interact to construct scientific knowledge. We hope we have convinced the reader that scientists do borrow epistemic credibility from many sources, and that these sources reflect the relevance of extra-theoretical elements in the construction of specialized scientific knowledge.

We do not claim that scientific controversies closely linked to practice and/or methodology exclusively rest on epistemic credibility. Instead, we argue that epistemic credibility is likely to play an important role in controversies where empirical results alone cannot help settle differences. For example, we detect indirect appeals to epistemic credibility in Dietrich and Skipper's 2007 account of the molecular clock controversy, in which both theoretical and atheoretical elements played an important role. We find a similar analytical context in Winsberg (2006, p. 2), who states that the credibility of a simulation model comes not only from its governing theory, but also "from the antecedently established credentials of the model building techniques employed by the simulationists". In these examples, and similar to what we have seen in phylogeography, scientists have had to look for arguments beyond the theory and its empirical consequences to settle a theoretical dispute or to justify particular modeling practices. However, the elaborations of Dietrich & Skipper and Winsberg make reference to extra-theoretical elements in the context of epistemic virtues and/or the social aspects of science as surrounding theoretical disputes; in

contrast, our account and framework place BEC at the core of theoretical disputes, not at the periphery.

Our framework also gives evidence of two great traditions within evolutionary biology, showing that this field is by no means a single unified body of thinking. For example, theory in the form of the coalescent may play no role in the controversy but Popperian falsificationism is often cited. This is not a particular love affair phylogeographers have with Popperianism, but is instead included in a long-lasting relation that prominent research groups within systematics and evolutionary biology have had with Popper's vision of science since the 1960s. As discussed by Rieppel (2008; see also Stamos 1996; Helfenbein and DeSalle 2005; Faith 2005), falsification of hypotheses has been perceived as a crucial component of good scientific practice for decades by a large number of systematists and evolutionary biologists. Templeton's borrowing of epistemic credibility from Popper follows this tradition, as do ABC supporters, who do not attempt to challenge the Popperian ideal but instead try to show how NCPA is not really an example of falsificationism. In a similar way, one can trace back sources of epistemic credibility in the debate to other important values, norms, traditions, or practices particular to different schools within evolutionary biology. To name some examples, different schools defend their own interpretations of statistical inference (see Sarkar 1992), the value of graphical versus mathematical representation (see Petit 2007), or the value of graphical versus mathematical reasoning (Beaumont et.al. 2010). Perhaps this example of an intellectual tradition within a guild of scientists illustrates the social dimension of knowledge (see for example Wusch 2002). From this point of view, normativity is a rational standard provided by a community (see Wusch 2002, chapters 8 and 9). Applied to our case, phylogeographers in disagreement represent two communities within evolutionary biology that have quarreled for decades over the correct way to conceptualize their field, evolutionary history, and science itself. The norms and values defended by each community are made explicit in the sources from which epistemic credibility is borrowed by NCPA and ABC supporters. If this is correct, then it should be possible to reconstruct the intellectual history of evolutionary biology to show how and why (at least) two parties emerged and what their differences are. Perhaps such analysis could help find ways to work out the differences.

Our study also highlights other pertinent questions of the theory-as-practice view. For example, science evolves via a continual anastomosis of practices and concepts. Phylogeography is an excellent example, because it is a synthetic discipline forged by the union of phylogenetics, biogeography, and population genetics. Other synthetic disciplines include evolutionary developmental biology and metabolomics. If phylogeography is any guide, then it would seem that these fields do not emerge out of the construction of novel theory but as the pulling together in the laboratory of disparate research traditions. Theory, along with practice generally construed, may be inherited from the different fields being brought together, but it is by no means clear why certain elements make the step to the new field, or how theory originated within the new field is constructed. However, the conceptual tools that we propose here can illuminate some aspects of the forging of synthetic disciplines, by exposing the web of distinct elements imported from other fields and how they are used to construct a new discipline. Again, the notion of epistemic credibility is vital as the new discipline imports not only theory and methodologies but— as the phylogeographic debate illustrates— what the members of the new community should take for rational, objective, or scientific.

Borrowing epistemic credibility is not restricted to phylogeography. It is clear that in novel synthetic fields, the progenitor disciplines must play an important role as the source fields for epistemic credibility. The clearest examples come from disciplines in which the merits of theories or practices in general cannot be evaluated by experimentation, that is, areas in which the

experiments needed to resolve a controversy conclusively would be so difficult to carry out and interpret that scientists must appeal to other sources to defend the empirical data. Take, for example, the case of the danger theory in immunology (Matzinger 1994). Danger theory was introduced to explain some of the shortcomings of the clonal selection theory, the field's champion theoretical corpus. A key distinction of danger theory has to do with the signals that turn the immune system on and off, a difference that should translate into important medical consequences most notably in cancer treatment or transplants. However clear the consequences of the theory, experiments are very difficult to come by as it is not clear how to manipulate something as complex as the immune system to generate clear experimental results. Nonetheless, danger theory has slowly gained popularity, as its supporters have successfully defended it appealing to numerous sources of epistemic credibility. For example, Kuhn's paradigms (i.e. Matzinger 2007) to defend the relevance and importance of danger theory, informatics where there are successful intrusion detecting systems based on the intuitions provided by the immunological theory that indirectly support the notion that *danger* is an important element for any surveillance and protection mechanism (i.e. Ou 2010, Vella 2010), and to a theory of meaning and reference to contextualize the problems raised by the clonal selection theory and the advantages of a danger theory (Matzinger 1994).

## 7. Conclusion

The polemics surrounding NCPA illustrate how a seemingly methodological controversy is in effect a larger debate confronting two visions of what one scientific field should be. This larger debate is evident in the numerous sources of epistemic credibility borrowed by the participants and the scant presence of coalescent theory. From a philosophical perspective, we illustrate a case of scientific controversy that takes place largely in the realm of practices. In our analysis, the interplay of theory and practice-based controversies, theory evaluation using epistemic virtues, and the web of conceptual relationships traced by the borrowing of epistemic credibility, contrasts with the traditional conception of theory as separate from practice. Instead, our analysis adds support to the increasing recognition of the inseparable reciprocity between concepts and practice in science.

Our analysis also highlights an important social component to the practice of science and the generation of knowledge. The sources of epistemic credibility invoked in this particular controversy are not simply rhetorical devices used to defend a particular methodology but trace the historical evolution of the field of genealogical studies by exposing the epistemic, methodological, and theoretical commitments shared by its different communities. In our reconstruction it can be seen how the controversy transcends the particulars of ABC versus NCPA to deal with numerous fundamental disagreements in terms of the importance of gene trees, the need to automate the inference process, the importance of formal inferences, and in general the sources of epistemic credibility that justify the results of a given method. Perhaps then, this controversy is another chapter in a long debate evolutionary biologists have had to understand better their field, evolution, and science, and shows how complex practices, in which the distinction between theory and practice is difficult to demarcate, begs revision of notions of the nature of explanation and how scientists define the appropriate structure of inferences.

<sup>1</sup>'Haplotype' denotes any particular DNA variant. The DNA segments used in these studies have numerous positions with different states across the different individuals sampled. These states can be manifested as substitutions of different bases (G, A, T, or C) or by the presence or absence of a base at a given site, known as an insertion-deletion event or indel. Different individuals have different combinations of substitutions and indels. Each unique combination of states in the same molecule of DNA is referred to as a haplotype.

<sup>2</sup>'Clade' is the term used to denote a phylogenetic grouping that includes an ancestor and all of its descendants.

## References

Aldrich J (2002) How Likelihood and Identification Went Bayesian. *International Statistical Review / Revue Internationale de Statistique* 70: 79-98.

Aldrich J (2008) R. A. Fisher on Bayes and Bayes' Theorem. *Bayesian Analysis* 3: 161–170 .

Avise JC (1992) Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos* 63:62-76.

Avise JC (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography* 36: 3–15.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* 18:489-522

Beard DA, Kushmerick MJ (2009) Strong inference for systems biology. *PLoS computational biology*, 5 (8): e1000459.

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics* 5: 251-261

Beaumont M, Panchal M (2008) On the validity of nested clade phylogeographical analysis. *Molecular Ecology* 17: 2563–2565

Beaumont MA, Nielsen R, Robert C, Hey J, Gaggiotti O, Knowles L, Estoup A, Panchal M, Corander J, Hickerson M, Sisson SA, Fagundes N, Chikhi L, Beerli P, Vitalis R, Cornuet JM, Huelsenbeck J, Foll M, Yang ZH, Rousset F, Balding D Excoffier L (2010) In defence of model-based inference in phylogeography. *Mol. Ecol.* 19: 436 – 446.

Bloomquist EW, Lemey P, Suchard MA (2010) Three roads diverged? Routes to phylogeographic inference. *Trends Ecol Evol.* 25: 626-32.

Burian R (2005) *The Epistemology of Development, Evolution, and Genetics*. Cambridge University Press. NY.

Castelloe J, Templeton RA (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. *Molecular Phylogenetics and Evolution* 3: 102-113.

- Ou C-M, Ou C-R (2010) Agent-Based Immunity for Computer Virus: Abstraction from Dendritic Cell Algorithm with Danger Theory. *GPC 2010*: 670-678
- Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9: 1657–1659.
- Diniz Filho JAF, Telles MPDC, S. L. Bonatto et al. 2008. Mapping the evolutionary twilight zone: molecular markers, populations and geography. *Journal of Biogeography* 35: 753-763.
- Efron B (1986) Why isn't everyone a Bayesian? *Am. Stat.* 40: 1-5
- Fagundes NJR, Ray N, Beaumont M et al. (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences*, 104, 17614–17619.
- Felsenstein, J (2004) *Inferring Phylogenies*. Sinauer Associates, Sunderland, Mass.
- Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD (2010) Phylogeography's past, present, and future: 10 years after *Awise*. *Molecular Phylogenetics and Evolution* 54: 291–301
- Howie D (2002) *Interpreting Probability: Controversies and Developments in the Early Twentieth Century*. Cambridge University Press, New York.
- Hudson RR (1983) Properties of a neutral allele model with intragenic recombination. *Theor. Pop. Biol.* 23: 183–201.
- Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294: 2310-2314.
- Platt JR (1964) Strong inference. *Science* 146: 347-353
- Joly S, Stevens MI, van Vuuren BJ (2007) Haplotype networks can be misleading in the presence of missing data. *Systematic Biology* 56: 857-862.
- Keller EF (2002) *Making Sense of Life : Explaining Biological Development with Models, Metaphors, and Machines*. Harvard University Press, Cambridge
- Kingman JFC (1982) The coalescent. *Stochast. Proc. Appl.* 13: 235–248.
- Knowles LL (2004) The burgeoning field of statistical phylogeography. *J. Evol. Biol.* 17: 1-10.
- Knowles LL (2008) Why does a method that fails continue to be used? *Evolution* 62: 2713-2717.
- Knowles LL, Maddison WP (2004) Statistical phylogeography. *Mol. Ecol.* 11: 2623-2635.
- Helfenbein KG, DeSalle R (2005) Falsifications and corroborations: Karl Popper's influence on systematics. *Molecular Phylogenetics and Evolution* 35: 271–280.
- Pournari M (2008) The Distinction Between Epistemic and Non-Epistemic Values in the Natural Sciences. *Sci & Educ.* 17:669–676
- Vella M, Roper M, Terzis S (2010) Danger Theory and Intrusion Detection: Possibilities and Limitations of the Analogy. *Lecture Notes in Computer Science*, 6209: 276-289.
- Kusch M (2002) *Knowledge by Agreement: The Programme of Communitarian Epistemology*. The Clarendon Press, Oxford.

- Matzinger P (1994) Tolerance, Danger, and the Extended Family. *Ann. Reviews of Immunology*, 12: 991-1045.
- Matzinger P (2007). Friendly and dangerous signals: is the tissue in control? *Nat. Immunol.* 8: 11-13.
- Mayr E (1982) *The Growth of Biological Thought: Diversity, Evolution and Inheritance*. Harvard University Press, Cambridge.
- McMullin E (1987) Scientific controversy and its termination. In: Engelhardt HT, Caplan AL (eds) *Scientific controversies: Case studies in the resolution and closure of disputes in science and technology*. Cambridge University Press, Cambridge, pp. 49–91.
- McVean GA, Cardin NJ (2005) Approximating the coalescent with recombination. *Philos Trans R Soc Lond B Biol Sci.* 360:1387-93.
- Dietrich MR, Skipper RA (2007). Manipulating Underdetermination in Scientific Controversy: The Case of the Molecular Clock. *Perspectives on Science* 15: 295-326.
- Nei M, Takahata N (1993) Effective population size, genetic diversity, and coalescence time in subdivided populations. *J Mol Evol.* 37:240-244.
- Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Mol. Ecol.* 18: 1034-1047.
- O'Fallon BD, Seger J, Adler FR (2010) A continuous-state coalescent and the impact of weak selection on the structure of gene genealogies. *Mol Biol Evol.* 27:1162-1172.
- Olsen KM, Schaal BA (1999) Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. *PNAS* 96: 5586-5591.
- Panchal M, Beaumont MA (2010) Evaluating nested clade phylogeographic analysis under models of restricted gene flow. *Syst Biol.* 59:415-432.
- Panchal M (2007) The automation of nested clade phylogeographical analysis. *Bioinformatics* 23: 509-510.
- Petit RJ (2008<sup>a</sup>) The coup de grâce for nested clade phylogeographic analysis? *Mol. Ecol.* 17: 516–518.
- Petit RJ (2008<sup>b</sup>) On the falsifiability of the nested clade phylogeographic analysis method. *Mol. Ecol.* 17: 1404.
- Petit RJ, Grivet D (2002) Optimal randomization strategies when testing the existence of a phylogeographic structure. *Genetics* 161: 469-471.
- Pickering A (1995) *The mangle of practice: time, agency, and science*. University of Chicago Press, Chicago.
- Popper KR (1959) *The Logic of Scientific Discovery*. Hutchinson, London.
- Posada D, Crandall KA (2001) Intraspecific phylogenetics: Trees grafting into networks. *Trends in Ecology and Evolution* 16: 37-45.
- Posada D, Crandall KA, Templeton AR (2000) GeoDis: A program for the Cladistic Nested Analysis of the Geographical Distribution of Genetic Haplotypes. *Mol. Ecol.* 9:487-488.

- Quine WV, Ullian, JS (1978) *The Web of Belief* Random House, NY.
- Rheinberger HJ (1997) *Toward a history of epistemic things: synthesizing proteins in the test tube.* Palo Alto. Stanford University Press.
- Rieppel O (2008) 'Total evidence' in phylogenetic systematics. *Biol Philos.* 24: 607-622.
- Nola R, Sankey H (2007). *Theories of Scientific Method: An Introduction.* Acumen, London.
- Sarkar S (ed) (1992) *The Founders of evolutionary genetics (a centenary reappraisal).* Kluwer Academic Publishers, Dordrecht.
- Stamos DN, (1996) Popper, falsifiability, and evolutionary biology. *Biol Philos.* 11: 161-191.
- Chamberlin TC (1897) The method of multiple working hypotheses. *J- Geol.* 5: 837.
- Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105, 437-460.
- Templeton AR. 2010a Coalescent-based, maximum likelihood inference in phylogeography. *Mol Ecol.* 2010 Feb;19(3):431-5; discussion 436-46.
- Templeton AR. 2010b. Correcting approximate Bayesian computation. *Trends Ecol Evol.* 2010 Sep;25(9):488-9; author reply 490-1. Epub 2010 Jul 7. No abstract available.
- Templeton AR. 2010c Coherent and incoherent inference in phylogeography and human evolution. *Proc Natl Acad Sci U S A.* 2010 Apr 6;107(14):6376-81. Epub 2010 Mar 22.
- Templeton, A. R. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* 117: 343-351.
- Templeton, A. R. 2001. Using phylogeographic analyses of gene trees to test species status and processes. *Mol. Ecol.* 10: 779-791.
- Templeton, A. R. 2002. "Optimal" randomization strategies when testing the existence of a phylogeographic structure: A reply to Petit and Grivet. *Genetics* 161: 473-475.
- Templeton, A. R. 2006. *Population genetics and microevolutionary theory.* Wiley-Liss, NJ.
- Templeton, A. R. 2008. Nested clade analysis: an extensively validated method for strong phylogeographic inference. *Mol Ecol.* 17: 516-518.
- Templeton, A. R. 2009a. Statistical hypothesis testing in intraspecific phylogeography: nested clade phylogeographical analysis vs. approximate Bayesian computation. *Molecular Ecology* 18: 319-331.
- Templeton, A. R. 2009b. Why does a method that fails continue to be used: the answer. *Evolution* 63: 807-812.
- Templeton, A. R., E. Routman, and C. A. Phillips. 1995. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* 140: 767-782.
- Winsberg Eric (2006). *Models of Success Versus the Success of Models: Reliability Without Truth.* *Synthese* 152 (1).



Winsberg Eric (2009). Computer Simulation and the Philosophy of Science. *Philosophy Compass* 4 (5):835-845.

Winsberg Eric (2010). *Science in the Age of Computer Simulation*. The University of Chicago Press.

Woolley SM, Posada D, Crandall KA, 2008 A Comparison of Phylogenetic Network Methods Using Computer Simulation. *PLoS ONE* 3(4): e1913. doi:10.1371/journal.pone.0001913