

The phylogeography debate and the epistemology of model-based evolutionary biology

Alfonso Arroyo-Santos · Mark E. Olson ·
Francisco Vergara-Silva

Received: 20 March 2012 / Accepted: 17 June 2013
© Springer Science+Business Media Dordrecht 2013

Abstract Phylogeography, a relatively new subdiscipline of evolutionary biology that attempts to unify the fields of phylogenetics and population biology in an explicit geographical context, has hosted in recent years a highly polarized debate related to the purported benefits and limitations that qualitative versus quantitative methods might contribute or impose on inferential processes in evolutionary biology. Here we present a friendly, non-technical introduction to the conflicting methods underlying the controversy, and exemplify it with a balanced selection of quotes from the primary biological literature, to invite the philosophy of biology community to pay attention to the elements that have played a primary role in its presumed resolution. We also present the basic features of our own metascientific take on the debate, and point out—as a preliminary step in preparation for upcoming, more detailed treatments—the importance that appeals to authority in fields external to phylogeography per se have played in the current status of this highly visible evolutionary biology dispute.

Keywords Phylogeography · Scientific controversies · Evolutionary biology · Popperianism · Statistical approaches · Borrowed epistemic credibility

A. Arroyo-Santos (✉)
Facultad de Filosofía y Letras, UNAM, Circuito Interior Ciudad Universitaria S/N, 04510 Mexico, DF, Mexico
e-mail: agripas@gmail.com

A. Arroyo-Santos
Centro de Información Geoprospectiva, Berlín 29A, Del Carmen Coyoacán, 04100 Mexico, DF, Mexico

M. E. Olson · F. Vergara-Silva
Instituto de Biología, UNAM, Circuito Exterior Ciudad Universitaria S/N, 04510 Mexico, DF, Mexico

Introduction

Biological research fields increasingly compete with each other not only for financial resources, but also for prominence, as perceived socially by the academic communities involved. For areas of investigation closely linked to the ‘molecular revolution’, claims of contribution to ‘new syntheses’ in biological discourse usually constitute a crucial element in the fate experienced by such enterprises. Multiple evidential sources point to phylogeography as one of the best recent examples within evolutionary biology of a DNA data and bioinformatics-oriented synthetic framework, whose prestige among financing agencies, research journals and local communities of scientists worldwide increases indefatigably. In addition to its appeal to biologists, phylogeography is particularly interesting in epistemological, historical and even sociological terms, because it has recently been the arena of a rather acrimonious controversy on the benefits and limitations that qualitative versus quantitative methods might contribute to inferential processes in evolutionary biology. Drawing on perspectives explored in previous works, in which evolutionary biology debates related to adaptive radiation (Olson and Arroyo-Santos 2009) and ‘pattern cladistics’ (Vergara-Silva 2009) have been analyzed, here we present a first, general overview—to be followed by future in-depth analyses—on the description of the axes of what we formally call ‘the phylogeography debate’. Our intention is to continue inquiry into the nature of scientific controversy in the contemporary biosciences, and also to introduce—with the aid of a basic exposition of the elements at stake in this controversy—the international philosophy of biology community to an area that offers plenty of primary material for philosophical analysis connected to central issues in modern evolutionary biology.

DNA, genealogy and place: the basic goals of phylogeography

American geneticist John C. Avise opened his now famous book on the subject with a clear definition of disciplinary scope: “phylogeography is a field of study concerned with the principles and processes governing the geographic distributions of genealogical lineages, especially those within and among closely related species” (Avise 2000, p. 3).¹ The aim of phylogeography has remained constant since the field’s inception: in the paper that earned Avise his widely recognized status as ‘the founding father of phylogeography’ (Avise et al. 1987; see also Riddle 2008; Avise 2009; Hickerson et al. 2010), it was 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). Avise (2000) presented phylogeography as a sub-discipline of biogeography, but—in congruence with his own

¹ As stated in the back cover of *Phylogeography*, Avise is Professor of Genetics at the University of Georgia, and is a member of the National Academy of Sciences and a Fellow of the American Academy of Arts and Sciences.

specialty and that of many of phylogeography's international devotees and followers—the field has been regarded mainly as a substantial addition to population biology/population genetics. This is because the traditional focus of population-oriented evolutionary biology subdisciplines is allele frequencies, which do not allow the reconstruction of the historical patterns of movement over landscapes at different scales that populations have been through. Phylogenetic techniques based on analysis of DNA sequence variation, however, do recover some aspects of history, and hypotheses regarding how species have moved over time can thus be tested. Thus, phylogeographical studies take approaches developed for inferring relationships between species, usually considered 'phylogeny', and apply them within species; the 'geography' suffix alludes to the analysis of the spatial distribution of genetic variation (Fig. 1).

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 out the samples into groups via similarity criteria (Avice 2000; Clement et al. 2000; Posada and Crandall 2001). 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 this purpose, phylogeographers usually superimpose their branching diagrams of relationships between populations on maps of the provenance of their samples. Such an exercise often highlights genetic differences between populations separated by major geographical barriers. For example, Avice (1992) detailed a case in which a suite of diverse 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 Gulf of Mexico, showed strong divergence in genetic variants between the Gulf and the Atlantic. The author found that, in all of these species, the populations

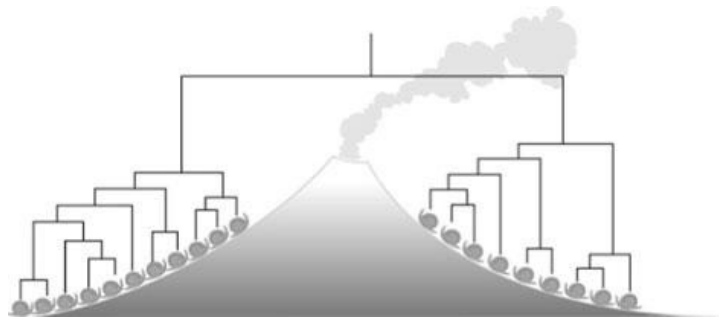


Fig. 1 Phylogeography examines the geographical distribution of genetic variation within species. The *branching lines* represent shared DNA variants between the individual snails sampled. Snails that are separated by fewer branches are more similar to one another and are inferred to have diverged more recently than more distant ones. Here we can see that this species of snail is separated into two great groups or clades, each corresponding to one side of the mountain. This result is congruent with the idea that the mountain represents an important barrier to gene flow in this species, because the DNA variants in each population coalesce such that the members of each population is more closely related to other members of its same population than to members of the other population

inhabiting the Gulf shared more similarities with the other Gulf populations than with those in the Atlantic and vice versa.

In general, the methods favored by phylogeographers involve techniques that group samples based on synapomorphies, that is, similarities that uniquely characterize a given group and are not shared with other groups. This procedure produces a hierarchical branching diagram that has some relationship to the historical branching events between species or between populations of a species. In phylogeography, similar thinking is applied within species.

The associations between geography and DNA-based phylogenetic reconstruction are used to make causal inferences regarding the relative role of “recurrent processes”, such as continual genetic interchange, as opposed to historical processes, which are unique events such as the rising of a mountain range. However, the patterns of deployment of genetic variation over local landscapes as well as larger geographical areas are susceptible to a plurality of biological interpretations. As a result, phylogeographers have developed a series of methods to explore these patterns. The development of these techniques could be generally described as a proliferation of bioinformatic softwares with ever increasing degrees of complexity. The contrasting conceptual foundations of the diverse methodologies that have been developed to perform analyses on phylogeographic data sets, have derived in a bitter dispute, creating a division within the discipline that seems difficult to reconcile, not only in practical and theoretical terms, but also sociologically.

We assign the following section(s) to an exposition, devoid of unnecessary mathematical technicisms, of the foundations of the two main methodological stances that in recent years have been at the heart of the debate in phylogeography. We then provide an initial classification of four aspects of epistemological importance that can be detected from our analysis of the corresponding primary literature. To conclude, we provide a general comment—in preparation for subsequent work, in which we will closely analyze such epistemic issues—on the seemingly paradoxical current status of phylogeography: as one of the principal contenders to the title of ‘the next evolutionary synthesis’ that nevertheless has experienced a deep methodological and epistemological division during the majority of its relatively brief existence.

Nested clade phylogeographic analysis

Nested clade phylogeographic analysis (NCPA) was developed by American geneticist Alan Templeton (Templeton 1987, 1998, 2001, 2002, 2006, 2008, 2010a; Templeton et al. 1995; Castelloe and Templeton 1994) and it is often regarded as the most popular phylogeographic method (Petit 2008a). NCPA starts with a haplotype network. A haplotype is a stretch of DNA unique for its combination of base substitutions, base insertions, or deletions. A typical haplotype network diagram is an undirected graph in which the vertices represent samples or inferred samples, and each edge represents a single mutational change that separates two vertices

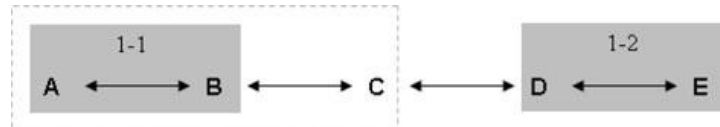


Fig. 2 A haplotype network. *Grey boxes* identify two *one-step* clades labeled 1-1 and 1-2. *Dashed lines* represent a potentially two-step clade formed by haplotype C and clade 1-1. Haplotype C is stranded between the one-step clades 1-1 and 1-2. If haplotype D is not one that was recovered in a sampled population but one inferred to exist between C and E, whereas B was indeed found in a sample, then C is mutationally closer to clade 1-1 thus forming the two-step clade

(Fig. 2).² We use the phrase ‘inferred samples’ because sometimes more than one substitution, insertion, or deletion separates two haplotypes. Numerous methods are available in the specialized literature to estimate haplotype networks (Posada and Crandall 2001; Woolley et al. 2008). For the sake of simplicity, here we will not discuss them in further detail, but will simply assume that a network has been estimated.

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. ‘Clade’ is the term used to denote a phylogenetic grouping that includes an ancestor and all of its descendants. In the context of nested clade analysis, it refers to groups of haplotypes. In the first of a series of steps for finding nested clades, it is necessary to identify the haplotypes that are connected to only one other haplotype (see Templeton 1987; Templeton and Sing 1993). These are referred to as the “tips” of the haplotype network, whereas those connected to more than one node are called “interior” haplotypes. Starting with the tips, all of the haplotypes that are connected by a single mutational step are grouped into “1-step clades”. In Fig. 2, we have outlined 2 one-step clades. In turn, the haplotypes to which they are connected by a further mutational step are regarded as two-step clades and so forth. In some cases, a haplotype may remain unnested. A variety of criteria are used to place these haplotypes, ranging from assessing mutational distances to other haplotypes to only statistical considerations. The mutational steps criterion can be illustrated with an example following Templeton and Sing (1993), Fig. 2 shows a situation in which haplotype C is stranded between the one-step clades 1-1 and 1-2. If haplotype D is not one that was recovered in a sampled population but simply one inferred to exist between C and E, whereas B was indeed found in a sample, then C is mutationally closer to clade 1-1 and would be grouped with it as a two-step clade (dashed lines). When the stranded haplotype is separated by the same number of mutational steps from all possible alternative groupings, then it is joined to the

² The methods favored by Templeton and his followers involve techniques that group samples based on the cladistic (i.e. Hennigian) concept of synapomorphy—that is, similarities that uniquely characterize a given group and are not shared with other groups. This procedure produces a hierarchical branching diagram that has some relationship to the historical branching events between species or between populations of a species. In Templeton’s view of phylogeography, similar thinking is applied within species. Here, we do not explore the complexities associated to old and recent discussions regarding the ‘correct’ interpretation and implementation (in computer programs) of synapomorphy, homology, and related concepts. For an overview of these concepts and some of the dimensions of such debates, see Vergara-Silva (2009) and references therein.

group with the smallest number of observations to maximize statistical power (see also Templeton and Sing 1993; Templeton 2006). Using this iterative process, all clades are classified into all possible levels of nesting.

With the clades classified, the analysis proper can begin. NCPA is based on two distances, the “clade distance” D_c and “nested clade distance” D_n (Templeton et al. 1995). To calculate D_c , it is first necessary to pick a clade and to plot the geographical distribution of its members. The D_c for a given clade is calculated as the mean distance between the localities in which the samples from which the DNA sequences were drawn, weighted by the frequency of each clade at each location. D_c is regarded as an index of how far a given clade has dispersed. Great circle distances are often used but others may be substituted, e.g. distances along waterways for aquatic organisms. Because D_c is an average distance weighted by the density of the collections at each site, clades that are both widespread and abundant at all sites will have D_c values that are higher than those for clades that are widespread only by virtue of, for example, a single individual collected far from an otherwise restricted range.

The D_n of a clade is calculated as the average distance between the center of distribution of the clade and the center of the clade within which it is nested. D_n is thus regarded as providing an index of how far a given clade has dispersed with respect to the clades with which it makes up a higher-order clade. Whereas D_c describes the current extent of a haplotype, D_n provides an idea of the movement of a haplotype with respect to its ancestor. NCPA makes inferences regarding source and descendant populations by invoking the assumption that tip haplotypes are likely to be younger than those that are connected to many other haplotypes deep in the tree (Castelloe and Templeton 1994).

The subsequent step asks whether or not haplotypes are distributed randomly across the landscape. Using a computer simulation, the haplotypes or clades nested within a given clade are randomly permuted across the geographical locations in which samples were collected, maintaining constant the number of haplotypes and sample sizes per geographical locality. D_c and D_n are recalculated at each permutation, generating a null distribution against which the observed values are examined. In this process, low D_n values that are due to very low sample sizes are distinguished from those with large ones because the ones that are represented by one or a few samples do not achieve statistical significance. The main inferences regarding biogeographical processes are generated in nested clade analysis by comparing the observed and null distributions. Significant deviations from the null expectation of no association between haplotypes and geography is taken as showing that the haplotypes are distributed non-randomly in space and significantly large or small clade and nested clade distances are identified. Given a model of recurrent gene flow with isolation by distance, the D_n and D_c should both show similar patterns of statistical significance (both should either be significantly large or small). If the D_c for a given clade is significantly large whereas the D_n is not, then we infer that the clade has expanded its range. If the D_n is larger than expected, then we infer that a shift in (the center of) distribution has occurred. Templeton et al. (1995) provide an inference key for interpreting results that forms the basis for all nested clade analysis studies. The key is a 21-question algorithm that leads the user

step by step to infer the biological meaning of NCPA's results. For example, the first two steps of the inference key ask:

1. Are all clades within the nesting clade found in separate areas with no overlap?
 - NO—Go to step 2.
 - YES—Go to step 19.
2. Is at least one of the following conditions satisfied?
 - a. The Dc's for one or more tips are significantly small and the Dc's for one or more of the interiors are significantly large or non-significant.
 - b. The Dc's for one or more tips are significantly small or non-significant and the Dc's for some but not all of the interiors are significantly small.
 - c. The Dc's for one or more interiors are significantly large and the Dc's for the tips are either significantly small or non-significant.
 - d. The I-T Dc is significantly large.
 - NO—Go to step 11.
 - YES—Go to step 3.
 - Tip/Interior Status Cannot be Determined—Inconclusive Outcome.

The key has been continually updated, with the most current version made available at a web site (http://darwin.uvigo.es/download/geodisKey_06Jan11.pdf). Computer programs that automate parts or all of the process have also been made available (Posada et al. 2000; Panchal and Beaumont 2007).

An example of an NCPA study is Templeton et al. (1995). They found that the tiger salamander *Ambystoma tigrinum* of the central US is divided into two genetically highly differentiated clades. Because loci will tend to coalesce in interbreeding populations, it is likely that these two clades became genetically differentiated as a result of an event that separated them. Translated into NCPA terms, Dn was much larger than expected when comparing the two great clades. However, the Dc of each clade was not significantly larger than that in the null model. Many combinations of Dn and Dc possible for different positions across a large gene genealogy. It is the interpretation of these patterns that forms the basis for the debate surrounding NCPA. We now turn to the main alternative methodology.

Approximate Bayesian computation

A group of increasingly influential phylogeographers reject NCPA in favor of a family of Bayesian-based statistical models, the most popular of which is called approximate Bayesian computation (ABC; see, for instance, Tavaré et al. 1997; Beaumont et al. 2002; Sunnåker et al. 2013). To test phylogeographic hypotheses with an ABC model, the phylogeographic researcher proposes a small set of hypotheses that might account for the geographical distribution of genetic variation observed at present. 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.³ 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 results support or contradict a given hypothesis. In the context of ABC this amounts, for example, to finding the simulated posterior probability distribution (this is, the probability of observing the parameter of interest conditional on the available data) that best approximates the observed data⁴ (for a more detailed discussion of Bayesian models see Sunnåker et al. 2013; Beaumont et al. 2010; Nielsen and Beaumont 2009; Beaumont and Rannala 2004 or Huelsenbeck et al. 2001). For example, it might be thought that the populations of a given species were separated into two clusters either by the rising of a mountain range in the Pleistocene or the drying of the climate in the Miocene. After ABC modeling, the more likely of these two scenarios could be calculated based on their relative fits to the distribution generated by the modeling procedure.

The phylogeography debate

Templeton's proposal of NCPA as an analytical tool in phylogeography was not initially meant to form the basis of a 'phylogeographic school of thought' against opposite stances, in analogy to the systematic biology/phylogenetics-related methodological conflicts of previous times (Felsenstein 2004; Ebach and Williams 2007; Vergara-Silva 2009). The current conflict between NCPA and ABC only became clear in 2002, when American geneticists and phylogeographers L. Lacey Knowles and Wayne P. Maddison published a criticism against NCPA. For Knowles and Maddison NCPA was a "primarily non-statistical" method to reconstruct phylogeographic history, through which "inferences about particular biological processes (are made) without explicit reference to stochastically derived expectations" (Knowles and Maddison 2002, p. 2623). In the same article, Knowles and Maddison also advocated for "the use and development of phylogeographic methods that make both explicit statistical links between process, prediction and test (like the coalescent-based population genetic models) and consider a diverse array of processes and histories (like NCA)" (Knowles and Maddison 2002, p. 2624. Note that these authors drop one letter from the NCPA acronym). The latter mention of

³ 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".

⁴ Briefly, Bayesian models make estimates of uncertain events using probability statements conditional on available information and an underlying model. At the core of Bayesian models lies the concept of *conditionalization* that, informally speaking, represents the notion that your beliefs or previous knowledge on something can change *conditional* on learning new and relevant things. Bayesian vocabulary divides previous knowledge or belief as *prior probability*, and the updated belief as *posterior probability*.

NC(P)A is, however, not used by these authors to establish a commonality of interests as their main objective, but to differentiate their approach from Templeton's program (see Knowles 2004, 2008, 2009; Nielsen and Beaumont 2009; Panchal and Beaumont 2010, among others).

From the standpoint of Knowles and Maddison the 'phylogeography debate' is seen as a controversy regarding statistics, in which what is at stake is whether or not NCPA is a suitable method to infer evolutionary scenarios. Knowles and Maddison are emphatic regarding the central role of statistics. For example, the term they chose to identify their stance "statistical phylogeography" (Knowles and Maddison 2002, p. 2624), a tag that has been preserved on many subsequent publications in the controversy. In what follows we quote extracts of the sometimes vitriolic interchange between Templeton and Knowles, Maddison and other detractors of NCPA. Our interest is to illustrate the nature of the dispute between ABC and NCPA, and how the debate is in fact a series of interrelated controversies, and not a unitary dispute, centered on a single dimension of theory and practice. To be sure, it is easy to assume that the controversy is centered only on the statistical limitations of both methods. The strictly statistical aspects of the controversy are undeniable. However, we argue that the debate, taken as a whole, actually deals with distinct issues, each reflecting the disciplinary vision each group advocates for phylogeography. In the remaining of this section, we comment on the exclusively statistical aspect of the controversy, and the move on to describe and classify other facets of the debate.

Statistics proper in the phylogeography debate

Two issues can be distinguished on the purely statistical side of the controversy. Firstly, there is a strictly statistical aspect related to the problems of the statistical methods per se. Secondly, there is a controversy dealing with *interpretation*, meaning differences regarding how the results of statistical analyses should be connected to actual features of individual organisms. In the paper that started the controversy, Knowles and Maddison address both issues, raising concerns about the use of statistical error in NCPA's inferences, and about the role played by statistical testing and its interpretation.

Focusing on the statistical controversy proper, Knowles and Maddison complain that single-locus NCPA ignores the randomness associated with the coalescent process and the accumulation of mutations. Knowles and Maddison ask "[D]oes NC[P]A then fill the role desired of a flexible, statistical inference procedure?" (p. 2631). To answer their question they tested how accurately NC[P]A could reconstruct a known history of allopatric divergence to conclude that "[T]his small study does illustrate one of the primary problems with NC[P]A, which is that it does not adequately consider the stochastic nature of the coalescent process when making historical estimates" (p. 2632). Knowles and Maddison (2002) charge that NCPA "does not attempt to distinguish statistically among alternative interpretations, nor does it provide an estimate of the uncertainty in its conclusions. Thus, for any interpretation derived from Templeton's inference key, we do not know 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).

Further studies by the same authors (and others with similar views) charge NCPA with overinterpreting the data. These authors charge that ‘stochastic variance’ - which in this context refers to the degree to which gene trees coincide or fail to coincide with population history- can lead to well supported but erroneous results (Hey and Machado 2003; Knowles 2004). Knowles (2004) is particularly emphatic on this point and criticizes NCPA for taking “a gene tree more or less at face value (that is, they do not explicitly consider the stochasticity of genetic processes)” (p. 4). Knowles goes on to note that, as a result, “Tests of phylogeographical hypotheses therefore can be sensitive to which and how many parameters are included in the model depending on the extent that (a) the models accurately represent the history and (b) the alternative hypotheses can be distinguished statistically with the available data”. Finally, computer simulations made by Knowles and Maddison in 2002, and by Panchal and Beaumont in 2007 and 2010, found that NCPA had a high-rate of false positive results, that is, NCPA was prone to inferring a significant event where none had occurred.⁵

The ‘interpretation controversy’ has dealt primarily with the separation of statistical testing and biological interpretation versus the position that statistical tests can be designed so as to have ineluctable biological significance. Knowles and Maddison (2002) charged that in NCPA “the inference key goes far beyond rejection of this null model, and selects a particular process to which the pattern is attributed. This selection of a particular process, over others, is not done statistically. The fact that one can conclude statistically that there must be a process generating a pattern does not say what process in particular generated that pattern” (p. 2630). Followers of ABC in general have insisted that their approach inextricably ties together testing and interpretation and that this is the correct way to proceed (see for example Beaumont et al. 2010; Knowles 2004, 2008). In the words of Knowles (2004, p. 5) “the actual design of a study follows from the specific questions being addressed. This contrasts with the tradition of first collecting data, and then asking what the data mean (e.g. Avise 1998; Templeton et al. 1995).”

Templeton has replied that the process advocated by Knowles, which according to her makes statistical tests and hypothesis testing one and the same, also separates these elements, as in NCPA and as in any other study using statistics. Moreover, he says that Knowles’s approach relies on an inference key just as much as NCPA, by specifying which genetic patterns should be associated with which historical events and recurrent processes. The following passage from Templeton (2004) is representative of this position: “Both the phylogeographical methods advocated by Knowles and Maddison (2002) and NCPA distinguish among alternative

⁵ On the statistical proper front, Templeton has accused ABC of producing posterior probabilities that are not true probabilities, of dealing with overlapping models (thus producing incorrect mathematical estimates), of using subjective estimates in the form of prior probabilities, and of not exhausting the hypothesis space falling into what is known as ‘the problem of the bad lot’ in the Bayesian literature meaning that, because all such models must choose the best candidate from any given set of options, if the lot includes only bad choices (for example, all options are false), the method will choose the better candidate nonetheless (see Templeton 2008, 2009a, b, 2010a, b).

interpretations by finding a statistic or set of statistics that deviate significantly from some well-defined model coupled with an interpretative key. The main difference between these approaches is that the interpretative key is applied a priori and implicitly in statistical phylogeography sensu Knowles and Maddison (2002), whereas it is applied a posteriori and explicitly by Templeton et al. (1995).” (p. 798; see also Templeton 2009a).

Because phylogeographers of both persuasions have spent a great deal of time discussing the pros and cons of statistical modeling approaches, the phylogeography debate might seem to be simply a discussion over methodologies. For example, in a recent review from the primary biological literature, where the existence of a ‘phylogeographic controversy’ is explicitly acknowledged, Bloomquist and collaborators state that the debate “has focused around two general points: (1) does single-locus and multi-locus NCPA have an inherently high false-positive rate and does this preclude its use? and (2) do model-based methods or NCPA provide a more appropriate way to analyze phylogeographic data?” (Bloomquist et al. 2010, 627). It is important to note that if the NCPA controversy were simply a methodological dispute, then it would be resolved when both methods were rigorously tested—i.e. by showing that one is more reliable than the other. However, the controversy cannot be settled by the results generated by the very methodologies in dispute, because the notions regarding what would constitute proof and the process to generate this proof, differ between the two groups in the controversy. NCPA proponents lean toward a sort of deductivism and supporters of ABC models 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 or no’ decisions, which are given biological meaning in the context of the inference key. ABC models instead compare the computer modeling of different evolutionary scenarios, and among these uses statistical evaluations to decide which seems the likeliest among the alternatives modeled. 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 in evolutionary biology.

On another reading, the debate might seem to be also the continuation of the long controversy between frequentist and Bayesian interpretations of probability in a phylogeographic context (e.g. Aldrich 2008; Aldrich 2002; Howie 2002; Efron 1986). However, competing groups have never said that either null-hypothesis testing or Bayesian modeling are wrong in any statistical sense. Templeton actually advocates using both approaches; in fact, he recently said that “[A]lthough I have been misrepresented as being antiBayesian, I am not even anti-ABC” (Templeton 2010b, 489; Templeton 2009a).

Our analysis shows that the debate actually has nothing to do with ‘good versus bad’ statistical implementations, but about contrasting visions of what phylogeography ought to be. In this sense, quarreling parties use tags such as ‘statistical phylogeography’ to make general claims about their practice, methodologies, and scientific values. For instance, statistically oriented phylogeographer R my Petit once said that this debate might be properly described as a discussion between “two different schools of thought (...), the phylogeographic school, which relies on graphical representations of intraspecific

phylogenetic trees or networks, and the mathematical one, which relies on mathematical models of population genetics” (Petit 2008a, p. 516. Note the use of the “school of thought” tag by a participant in the debate). Petit circumscribed his commentary to the reliance on trees, crucial to NCPA, but more deeply his qualms had to do with the value of graphical versus mathematical representations for statistical inferences. In our view, Petit’s comment has deeper roots that translate into how competing groups understand their research field and the scientific tradition it belongs to. In a significant gesture from a philosophy of science perspective, Templeton, the herald of one of the schools in dispute, has repeatedly appealed to Popper and falsificationism to defend NCPA. In doing so, he continues a long tradition in evolutionary biology to rely on Popper as a metatheoretical source of justification. In contrast, the statistico-mathematical school appeals to what they call model-based inferences, and insist upon the exploitation of technological advances to demonstrate the virtues of comparing complex models.

Four epistemic aspects of the phylogeography debate (separate from strictly statistical issues)

The strictly statistical aspects of the controversy are undeniable. In the previous section, we mentioned some of the numerous differences both parties have on this front. However, we argue that the debate, taken as a whole, actually deals with other issues deriving from the disciplinary vision each group advocates for phylogeography. Next, we review four issues central to the controversy.

What counts as a good inference?

As mentioned before, Templeton (2009a) has presented NCPA as being a Popperian method. Because the NCPA inference key builds its conclusions based on a series of ‘yes or no’ answers, Templeton asserts that it can be considered as fitting in a falsificationist framework. In the inference key, a given pattern of significance between D_c and D_n among nested clades should be the signature of a given historical or recurrent process. At each step, the results are either congruent with what would be expected if a given process had occurred, or are incongruent. In the latter case, Templeton considers that the hypothesis of a given process is falsified, and falsification is ‘strong’ inference (in the sense of Platt 1964; see also Chamberlin 1897; Beard and Kushmerick 2009). In contrast, Templeton presents the modeling approach as ‘weak’ inference. It is presented as weak because, rather than rejecting hypotheses because the data do not meet the criteria specified, goodness-of-fit metrics are used to choose the scenario that best fits the data. The problem is that even among several models that do not reflect the population history well, one will likely fit the data better than others. In this regard, Templeton asserts that “when the hypothesis space is not exhaustively covered, testing the relative merits among a set of hypotheses results in weak inference. A serious deficiency of weak inference occurs when all of the hypotheses being compared are false. It is still possible that one of them fits the data much better than the alternatives and as a result could have a high relative probability” (Templeton 2009a, p. 320).

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, an epistemological stance traditionally presented in evolutionary biology as the antithesis to Popper and indeed to science (e.g. Mayr 1982). Beaumont and colleagues are emphatic on this point: “without a properly randomized experiment, causal explanations are necessarily weak because they are potentially confounded with unobserved effects. Since they are based on observational data, phylogeographic studies are not amenable to randomized interventions and therefore all phylogeographic inference methods, including NCPA, lead to ‘weak scientific inference’ in the sense that it does not arise from planned scientific experiments. Popper was fiercely opposed to inductivism, whereby facts are gathered and then general laws identified. In this regard, rather than being a Popperian falsification method, NCPA can in fact be viewed as an anti-Popperian inductivist approach (Beaumont and Panchal 2008), since a story is built out of the patterns in the data” (Beaumont et al. 2010, p. 438).

In a 2010 paper, Templeton presented another twist to the inference side of the controversy charging that NCPA is a coherent method. By this, he means that “the statistics or probabilities used to measure the goodness of fit of the models obey the constraints imposed by formal logic” (Templeton 2010c, p. 6376). Templeton assumes here that any hypothesis is actually a group of interconnected hypotheses of different generality. In such a situation, the probability of any interconnected hypothesis must be less than the probability of the general model. Whenever this condition is violated, the model is incoherent. Needless to say, under this perspective ABC is incoherent. The ‘coherent-incoherence’ paper has received answers from the Bayesian community that felt Templeton mounted a “broad attack on the foundations of Bayesian statistical methods—rather than on the purely numerical technique called approximate Bayesian computation (ABC)—using incorrect arguments and selective references taken out of context” (Berger et al. 2010, p. E15, Beaumont et al. 2010). This controversy can thus be seen as an extension of presenting Popperianism as the correct inference stance. The root of Templeton’s distinction on coherence and incoherence can be traced back to Popper’s *Logic of Scientific Discovery* and his discussion on the logical form of probability statements. (see Popper 1959, sections 66–68, see also Mayo and Spanos 2010 for a similar defense of Popperianism).

Automation and performance

A major component of the debate regarding NCPA has focused on the results of automated studies as a means of validating a given methodology. Three such studies have been conducted to evaluate NCPA, one by Knowles and Maddison (2002), and two more by Beaumont and Panchal (2008, 2010). These studies asserted that NCPA has very high rates of false positives and that as a result the method should be abandoned. This conclusion has been taken up by some subsequent authors—most notably by Petit (2008a, b), who has declared the simulation studies to have delivered the “coup de grace” to NCPA. Templeton has responded to criticisms raised by these studies (Templeton 2004), but the true conflict, again, has not been

about the results per se, but has been related instead to differences regarding the value of automated quantitative processes. For example, Panchal and Beaumont (2010, p. 17) claim that “having tested the NCPA method under a variety of settings, we have demonstrated that the method is at best unreliable. The NCPA procedure continues to be strongly defended (Templeton 2008, 2009a, b) and is unusual in that it has been particularly difficult to test because it has relied on lengthy manual procedures, as outlined in Panchal and Beaumont (2007) “and if it continues to be applied in future, it needs to be developed as a unitary, algorithmically defined, procedure that can be straightforwardly tested and compared by independent researchers. Without such development, we would urge extreme caution in its future use”. Previously, these same authors had 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?” (Beaumont and Panchal 2008, p. 2564). In short, Beaumont and Panchal’s major concern has to do with the manual procedures of NCPA. Knowles and Maddison (2002, p. 2630) also raise similar concerns, when they state that that “unfortunately, the fact that a key step of NCA—the inference key—must currently be executed by hand makes testing the method extremely labour-intensive”.

What method better reproduces known results?

Templeton has often explained his position by criticizing the work of Fagundes and collaborators on human evolution (Templeton 2008, 2009a, 2010a). Fagundes et al. used ABC to contrast three different models of human evolution: (1) modern humans originating in Africa migrate and replace totally other *Homo* species; (2) interbreeding occurred; and (3) gene flow took place between populations of *Homo sapiens* living in Eurasia and Africa. In Templeton’s account, Fagundes et al. concluded that option (1) is the most probable. These authors noted that “although our best supported model certainly does not represent the exact history of modern humans, we show here that it is much better supported by a random set of neutral loci than any other models involving interbreeding with other *Homo* species” (Fagundes et al. 2007, p. 17619). According to Templeton, the problem with the Fagundes et al. study is that option (3) is contrary to the prevailing notion of isolation by distance between humans living in Eurasia and Africa. Whereas Fagundes et al. consider this scenario as a viable option, they do not include other options that according to Templeton have been heavily corroborated such as the Acheulean out-of-Africa expansion (Templeton 2009a). Furthermore, “it is patent that the parameter values chosen by Fagundes et al. are strongly discrepant with the empirical data on autosomal coalescent times” (Templeton 2009a, p. 323). Finally, previous results obtained by Eswaran et al. (2005) arrived at exactly the opposite conclusion and refuted the out-of-Africa replacement model. Eswaran et al., like Fagundes et al., used a Bayesian simulation but considered the option of isolation by distance (the opposite to Fagundes et al.’s third option). For Templeton, this case exemplifies the ambiguity in model-based methods.

Authority: “Why does a method that fails continue to be used?”

In a 2008 paper published in the journal *Evolution*, Knowles asked the question reproduced above, prompted by the ‘terribly high’ error-rate of NCPA. The rationale is quite simple: if it has been shown by three independent groups that NCPA has such high error-rate, how come it is still as popular as ever? Knowles’ conclusion is that its popularity is due to what it promises to do, and not to what it can actually do. If correct, then all of NCPA’s popularity is based solely on Templeton’s authority.

Appeals to authority are conspicuous in this debate. The best case at hand is Beaumont et al. (2010), a letter signed by 22 authors that expresses concerns regarding NCPA and their support of ABC models. 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 made by Beaumont and Panchal (2008, p. 2563) who note that “there is a disagreement between Templeton (2004, 2008), who suggests the method works well, and three independent groups (Knowles and Maddison 2002; Petit and Grivet 2002; Panchal and 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”. Another approach is to appeal to the authority of a single prominent figure, such as Templeton’s aforementioned appeals to Popper, or 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, Knowles cites Felsenstein (2004)—one of the leading developers of phylogenetic methods. Given his prominence in systematic and evolutionary biology, Felsenstein’s verdict against NCPA would be weighty. Surprisingly, though, we could only locate one reference to NCPA in Felsenstein’s book (2004, 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 et al. 2000; Posada et al. 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’ citation of Felsenstein as an attempt to borrow from his authority to shore up her position. Finally, Rémy Petit (2008a, p. 1404) appeals to the authority of institutions, when he concludes that “the results of Panchal and Beaumont (2007) convince me that reputable journals should (1) discourage the use of the NCPA method for single locus data sets (...), and (2) 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.

Conclusions: phylogeography, controversies, and epistemic issues in neo-synthetic evolutionary biology

It is interesting to ponder that members of the statistico-mathematical camp in phylogeography have discordant views on the subject of research publication in their field. As mentioned in the closing section of this brief presentation of the main elements at play in ‘the phylogeography debate’, while authors such as Petit (2008a, b) explicitly recommend that ‘reputable journals’ should not publish NCPA-based phylogeographic studies, others—e.g. Knowles (2009, p. 596)—celebrate the “impressive trajectory (of) increase in the number of phylogeographic studies over the past 20 years”. It is undeniable that current model-based phylogeography is a burgeoning field, and it will be up to practicing biologists to decide if, ultimately, the implicit historiography of success that their most visible members are constructing—see, for instance, (Avice 2000, 2009; Knowles 2009; Hickerson et al. 2010)—is or should be a reflection of the type of phylogeography that should be deemed ‘good science’. From a metascientific point of view, however, phylogeography already provides an extraordinary case study that can illuminate investigations on the trajectories followed by theories, practices, and opposing research communities when an evolutionary biology controversy is ‘solved’.

Not many of the relatively young conglomerates of practices and theories in evolutionary biology claim to be “usher(ing) a new phylogenetic paradigm” (Knowles 2009, p. 603); not many of them⁶ have witnessed such polarized internal debates, either. In this context, we claim that an important element of the status of ‘novelty’ in phylogeography is anchored in the increasing dependence that both practical and theoretical components, apparently specific to the field, have on methodological strategies developed outside evolutionary biology. From our perspective, this ‘borrowed epistemic credibility’ is an aspect that might prove crucial to understand the presumed resolution of the phylogeography debate, and is therefore an issue that we will tackle in subsequent metascientific analyses on this prestigious evolutionary biology research area.

References

- Aldrich J (2002) How likelihood and identification Went Bayesian. *Int Stat* 70:79–98
- Aldrich J (2008) R. A. Fisher on Bayes and Bayes’ Theorem. *Bayesian Anal* 3:161–170
- Avice 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
- Avice JC (1998) The history and purview of phylogeography: a personal reflection. *Mol Ecol* 7:371–379
- Avice JC (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge
- Avice JC (2009) Phylogeography: retrospect and prospect. *J Biogeogr* 36:3–15

⁶ A rather evident example in this regard is evolutionary developmental biology (evo-devo), a research area also of recent inception in which the adoption of methods usually proceeds peacefully and with the agreement of large portions of the research communities involved. Evo-devo has sometimes been the arena of controversy, but this situation is usually restricted to cases in which the subdiscipline is compared to the (now old) ‘modern synthesis’ (e.g. Laubichler 2010 and Minelli 2010).

- Avice 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. *Annu Rev Ecol Syst* 18:489–522
- Beard DA, Kushmerick MJ (2009) Strong inference for systems biology. *PLoS Comput Biol* 5(8):e1000459
- Beaumont M, Panchal M (2008) On the validity of nested clade phylogeographical analysis. *Mol Ecol* 17:2563–2565
- Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nat Rev Genet* 5:251–261
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035
- 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
- Berger JO, Fienberg SE, Raftery AE, Robert CP (2010) Incoherent phylogeographic inference. *Proc Natl Acad Sci USA* 107(41):E157
- Bloomquist EW, Lemey P, Suchard MA (2010) Three roads diverged? Routes to phylogeographic inference. *Trends Ecol Evol* 25:626–632
- Castelloe J, Templeton RA (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. *Mol Phylogenet Evol* 3:102–113
- Chamberlin TC (1897) The method of multiple working hypotheses. *J Geol* 5:837
- Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. *Mol Ecol* 9:1657–1659
- Ebach MC, Williams DM (2007) An outline of the foundations of systematics and biogeography. *Hist Philos Life Sci* 29:83–92
- Efron B (1986) Why isn't everyone a Bayesian? *Am Stat* 40:1–5
- Eswaran V, Harpending H, Rogers AR (2005) Genomics refutes an exclusively African origin of humans. *J Hum Evo* 49:1–18
- Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L (2007) Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci* 104:17614–17619
- Felsenstein J (2004) *Inferring phylogenies*. Sinauer Associates, Sunderland, Mass
- Hey JCA, Machado CA (2003) The study of structured populations—new hope for a difficult and divided science. *Nat Rev Genet* 4:535–543
- 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 Avice, 2000. *Mol Phylogenet Evol* 54:291–301
- Howie D (2002) *Interpreting probability: controversies and developments in the early twentieth century*. Cambridge University Press, New York
- Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–2314
- 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 (2009) Statistical phylogeography. *Annu Rev Ecol Syst* 40:593–612
- Knowles LL, Maddison WP (2002) Statistical phylogeography. *Mol Ecol* 11:2623–2635
- Laubichler MD (2010) Evolutionary developmental biology offers a significant challenge to the Neo-Darwinian paradigm. In: Ayala F, Arp R (eds) *Contemporary debates in philosophy of biology*. Wiley-Blackwell, Singapore, pp 199–212
- Mayo D, Spanos A (eds) (2010) *Error and inference: recent exchanges on experimental reasoning, reliability and the objectivity and rationality of science*. Cambridge University Press, Cambridge
- Mayr E (1982) *The growth of biological thought: diversity, evolution and inheritance*. Harvard University Press, Cambridge
- Minelli A (2010) Evolutionary developmental biology does not offer a significant challenge to the Neo-Darwinian paradigm. In: Ayala F, Arp R (eds) *Contemporary debates in philosophy of biology*. Wiley-Blackwell, Singapore, pp 213–226
- Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Mol Ecol* 18:1034–1047
- Olson ME, Arroyo-Santos A (2009) Thinking in continua: beyond the “adaptive radiation” metaphor. *BioEssays* 31:1337–1346

- Panchal M, Beaumont MA (2007) The automation of nested clade phylogeographical analysis. *Evolution* 61:1466–1480
- Panchal M, Beaumont MA (2010) Evaluating nested clade phylogeographic analysis under models of restricted gene flow. *Syst Biol* 59:415–432
- Petit RJ (2008a) The coup de grâce for nested clade phylogeographic analysis? *Mol Ecol* 17:516–518
- Petit RJ (2008b) 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
- Platt JR (1964) Strong inference. *Science* 146:347–353
- Popper KR (1959) *The logic of scientific discovery*. Hutchinson, London
- Posada D, Crandall KA (2001) Intraspecific phylogenetics: trees grafting into networks. *Trends Ecol Evol* 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
- Riddle BR (2008) What is modern biogeography without phylogeography? *J Biogeog* 36:1–2
- Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C (2013) Approximate Bayesian computation. *PLoS Comput Biol* 9(1):e1002803
- Tavare S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. *Genetics* 145:505–518
- Templeton AR (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 AR, Sing CF, Kessling A, Humphries S (1988) A cladistic analysis of phenotype associations with haplotypes inferred from restriction endonuclease mapping 2. The analysis of natural-populations. *Genetics* 120:1145–1154
- Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Mol Ecol* 7:381–390
- Templeton AR (2001) Using phylogeographic analyses of gene trees to test species status and processes. *Mol Ecol* 10:779–791
- Templeton AR (2002) “Optimal” randomization strategies when testing the existence of a phylogeographic structure: a reply to Petit and Grivet. *Genetics* 161:473–475
- Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Mol Ecol* 13:789–809
- Templeton AR (2006) *Population genetics and microevolutionary theory*. Wiley-Liss, NJ
- Templeton AR (2008) Nested clade analysis: an extensively validated method for strong phylogeographic inference. *Mol Ecol* 17:516–518
- Templeton AR (2009a) Statistical hypothesis testing in intraspecific phylogeography: nested clade phylogeographical analysis vs. approximate Bayesian computation. *Mol Ecol* 18:319–331
- Templeton AR (2009b) Why does a method that fails continue to be used: the answer. *Evolution* 63:807–812
- Templeton AR (2010a) Coalescent-based, maximum likelihood inference in phylogeography. *Mol Ecol* 19:431–435
- Templeton AR (2010b) Correcting approximate Bayesian computation. *Trends Ecol Evol* 25(9):488–489, author reply 490–491
- Templeton AR (2010c) Coherent and incoherent inference in phylogeography and human evolution. *Proc Natl Acad Sci* 107(14):6376–6381
- Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics* 134:659–669
- Templeton AR, Routman E, Phillips CA (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
- Vergara-Silva F (2009) Pattern cladistics and the ‘realism-antirealism debate’ in the philosophy of biology. *Acta Biotheor* 57:269–294
- 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