

# Forty years of model-based phylogeography

David Bryant and Jamie Kydd

## 1 Introduction

Spatial structure has played a fundamental role in the evolutionary history of most organisms and any attempt at reliable phylogenetic inference needs to take this into account. The real problem is how to do this in practice: how to incorporate sample locations, present and past geography, and the effect of spatial correlations into phylogenetic inference. Integrating geography with phylogeny leads one quickly to a quagmire of difficult modeling and methodological issues, many of which remain unresolved.

It will come of no surprise to computational biologists that one of the first papers to develop methodology for model-based phylogeography is by David Sankoff. His 1972 paper, *Reconstructing the History and Geography of an Evolutionary Tree* [28] was, in many ways, 30-40 years ahead of its time. It describes a stochastic model incorporating *both* genetics and geography and shows how both geography and phylogeny can be estimated simultaneously from data. Sankoff's model captures important features of a dynamic spatial structure without being bogged down in a mass of geographic and environmental data. Curiously, the paper has been seldom cited, and almost entirely only within the field of lexicostatistics. It appears to have been completely missed by the phylogeographers. Our opinion is that the novelty of ideas and models places this work among the classic early papers of model-based phylogeography.

In this chapter we review Sankoff's article and argue that this paper merits revisiting. First we consider the problem of modeling geography, outlining the model in Sankoff's paper and demonstrating some links with related models. We then consider the interaction of genetics and geography. Sankoff introduces an appealing

---

David Bryant  
University of Otago, e-mail: david.bryant@otago.ac.nz

Jamie Kydd  
University of Auckland e-mail: jamie.b.kydd@gmail.com

method for reconstructing phylogeographic patterns, one which has analogues in Markov random field theory. We compare this approach to some contemporary approaches to phylogeography. We argue that Sankoff's approach is genuinely different and that it is not without its advantages.

## 2 Modeling Geography

### 2.1 Background

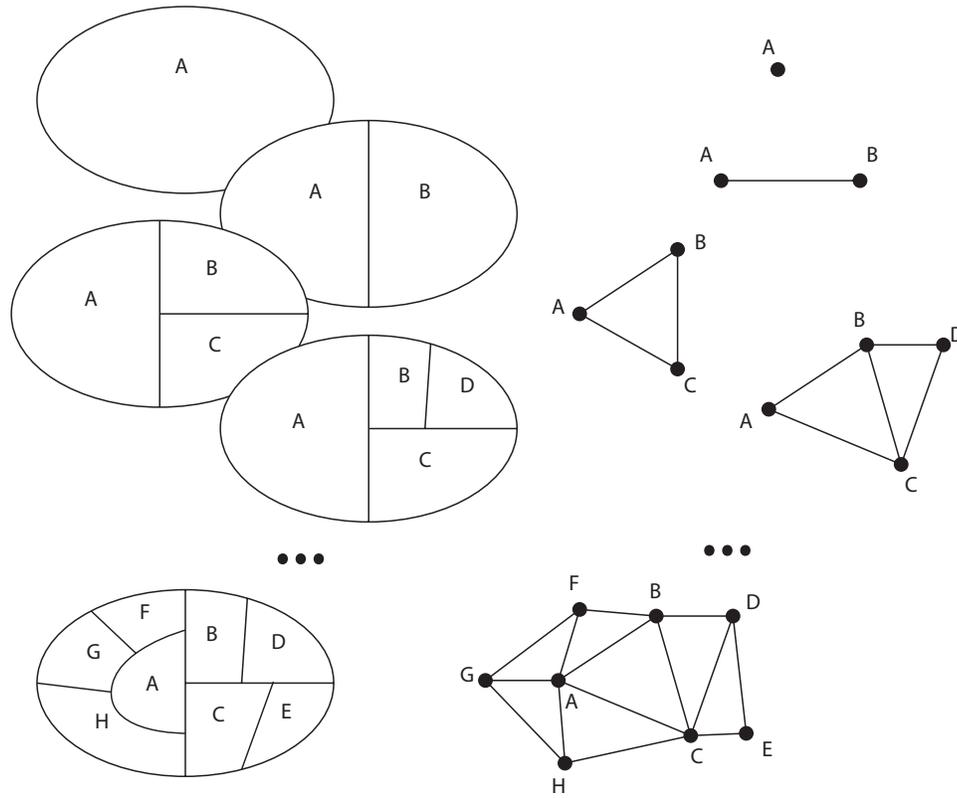
There are two key design decisions to be made in any model integrating geography and genetics [13]. The first is the statistical unit of analysis: does the model describe individuals, family units, villages, populations, species, or something in between. The second is the effect of geography on movement between these groups, or more correctly, on the gene flow between these groups.

The theoretical foundation for the most commonly used migration models in population genetics is provided by the island models of Wright [32], the stepping stone models [15] and their generalization to arbitrary migration matrices [3]. Under these models, the 'islands' or subpopulations are fixed *a priori*. The rate of migration or gene flow between different islands might be different for each pair, or equal to a constant, or capturing some aspect of the geographic structure. In many cases the choice of migration rates is governed more by mathematical convenience than biological realism.

We note that there are several models which do not break the population up into discrete chunks but instead consider a distribution of individuals in space. These have had far less impact than the island-based models. One important reason for the discrepancy is that it is straight-forward to set up a working discrete population model, whereas continuous space models have hidden difficulties. The continuous space model of Malecot [23] appears, at least superficially, to be quite reasonable and conservative. It is nevertheless internally inconsistent, as demonstrated by Felsenstein [12].

### 2.2 A joint model for phylogeny and geography

In Sankoff's 1972 paper, the (sub)-populations are represented as vertices in a planar graph or, equivalently, as regions in the plane. Migrations occur at a fixed and constant rate between populations connected by an edge. There is, however, a major difference between Sankoff's model and others based on networks on the plane: Sankoff's model is dynamic. The model includes not only the migration patterns between contemporary populations; it also describes how these connections change over time.



**Fig. 1** The splitting process studied by Sankoff. Initially there is a single region, here represented by a single region *A* (on the left) or a single vertex in the adjacency graph (on the right). At each splitting, a region is subdivided by selecting two edges and joining them with a new boundary line. The corresponding adjacency (dual) graphs appear on the right.

Initially, there is one region, corresponding to population at the root of the tree. After each population split (divergence/speciation) the corresponding region is subdivided, thereby adding one new vertex to the dual graph describing adjacencies (Figure 1). This process continues until we obtain one region for every contemporary population. In a way, the model describes a process of successive allopatric speciations.

A splitting is carried out as follows. A region is chosen uniformly at random. Two of the edges bounding that region are picked uniformly, and the region is split in a way that subdivides both edges. In the case that there was only one region, the boundary of that region is subdivided twice and the points of subdivision are joined by a new edge, creating two regions (Figure 1). There is considerable scope for different selection schemes, and Sankoff says as much in his paper.

One splitting process is illustrated in Figure 1. Initially there is a single region  $A$ . This is subdivided to give two regions  $A$  and  $B$ . Then  $B$  is subdivided, giving three regions  $A, B, C$ , and a further splitting of  $B$  gives a map with four regions  $A, B, C, D$ . The adjacencies between these regions are represented by the graph on the right.

One aspect of the model that is not completely clear from the paper is whether the external region could be split. The process would still be well defined, and these splittings could model expansion into new territory which lies outside the boundary of the original region  $A$ . The adjacency graph would be modified to include a specially marked vertex representing the external region, with edges to every region adjacent to the exterior boundary.

In either case, the process captures many aspects of connectivity relating to spatial structure. What is surprising is how much it leaves out. It contains no information about region size, or different environments, or even (beyond adjacency) the shape of the regions. This sparsity of information could turn out to be particularly useful. The model is clearly more believable than, say, models of populations distributed on a torus, or populations with no spatial structure at all. The model is clearly less ‘realistic’ than those incorporating landscape simulators and small-scale geographic niches. However realism in any model is only relative, and a hyper-realistic model is useless if it is not tractable.

### 2.3 *Properties of splitting*

Sankoff’s splitting process produces random planar adjacency graphs. The first natural question is whether there is anything special about the particular graphs produced.

To answer this, we start by looking not at the graph, but at the configurations of regions (or cells) produced by subdivision. Observe that the faces of the adjacency graph correspond to places where more *at least* three regions meet at a point, and that every time this happens, *exactly* three regions meet at that point. As a consequence all of the faces of the adjacency graph, except perhaps the external face, are triangular. It is not hard to show that this will always be the case, provided the number  $n$  of regions is at least three.

In the second version of the splitting model we permit splittings of the external vertex. In this case, all faces of the adjacency graph will be triangular, so that the adjacency graph (as a drawing) is a planar triangulation. There has been a great deal of work on these triangulations due to their application in surface visualization [14], finite element methods [5] and spatial data analysis [22]. The splitting operation of Sankoff corresponds exactly to *vertex splitting*; the reverse operation is called *edge contraction*.

An edge is said to be *contractible* if contracting that edge produces a valid triangulation. It can be easily shown that an edge  $e$  is contractible if and only if

- $e$  does not lie on any triangle of the graph which is not a face of the graph; and
- The triangulation is not  $K_4$  embedded on a sphere.

Steinitz and Rademacher [30] proved in 1934 that every triangulation of the sphere (assuming  $n \geq 4$ ) can be converted into  $K_4$  by a sequence of (valid) edge contractions. As a consequence, we have that the graphs produced by this version of the splitting process are exactly the triangulations: planar graphs with triangular faces.

We now turn to the version of the splitting process where splittings of the external region are *not* allowed. This change makes the analysis a little more complicated. The adjacency graph is still planar, and every face except the exterior face is triangular. The problem of finding valid edge contractions in this case is known as *polygon reduction*, a problem with applications to 3D graphics in the gaming industry [24]. Conditions for a valid edge contraction in this instance were established by [14]:

- $e$  does not lie on any non-facial triangle; and
- the triangulation is not  $K_3$  embedded in the plane.

An analogue of Steinitz and Rademacher's result for this version follows from Theorem 3 of [14] (which establishes that the triangle is a *subdivision* of a triangulation) and Lemma 4 of [14] (which shows that a subdivision of these triangulations can be obtained by edge contractions). Hence the adjacency graphs produced by this version of Sankoff's process are exactly the plane graphs with all triangular faces except the exterior face (also called *simplicial surfaces* [14]).

There are many more avenues for mathematical investigation here. The splitting model generates a distribution on triangulations; what is that distribution? Is it possible to compute, in polynomial time, the probability of a given planar triangulation? This question has a similar flavor to analyses of processes generating random trees [29].

### 3 Modeling Genetics

#### 3.1 The model

The genetic model described in Sankoff's paper is essentially the *infinitely many alleles* model, first introduced by [15]. It assumes a set  $\Gamma$  of genetic sites, and at each site the populations (regions) have a particular state (allele). Every mutation produces a new and unique type, and the only information available is whether two individuals carry the same type of allele at a site. There is a further assumption that there is no variation *within* each region (polymorphism). This is akin to assuming that the population sizes are small so that any mutant is quickly lost or fixed. A similar assumption is made by [27].

There are three distinct processes contributing to gene dynamics in the model. First, mutation: for each region, mutations occur at a fixed rate, and each mutation creates a new and distinct genetic type. Second, migration: at a fixed rate the type of a region is transferred from one of the adjacent regions. Random processes of this form are examples of *Markov random fields*, and appear in numerous guises in fields ranging from statistical physics to epidemiology. Kindermann and Snell

[16] credit the first work on Markov random fields to mathematicians in the former Soviet Union (e.g. [10]). Their appearance in the West coincided roughly with the publication of Sankoff's paper.

Of course the connection with Markov random fields only applies *between* splitting times, when the adjacency graph remains constant. This third process, splitting, adds a fairly novel twist to the analysis (compare the 'splitting operator' of [31]).

For the moment, consider the dynamics of an individual site. The relevant state information at a time  $t$  is then just the partition of the populations into types. We can analyze the model as a Markov chain with a state space equal to the set of partitions. A mutation takes a population and puts it into a class by itself. A migration (borrowing) transfers a population from one class to another. A splitting duplicates an element contained in one of the classes.

### 3.2 *Dynamic similarity*

The approach taken by Sankoff is to bypass computations over the space of partitions, and instead concentrate on the dynamics for pairs of populations. The resulting calculations are still exact and take account of all populations simultaneously, they just don't capture all of the higher-order dependencies between populations.

Let  $\mathbf{X}_t, \mathbf{Y}_t$  be two populations at some time  $t$  in the past, where  $t$  ranges from  $t = 0$  (the present) to  $t = -T$  (time of the first splitting). Let  $s(\mathbf{X}_t, \mathbf{Y}_t)$  denote the proportion of sites at which  $\mathbf{X}_t$  and  $\mathbf{Y}_t$  share the same state. Sankoff's calculations assume that the set  $\Gamma$  of sites is large enough that  $s(\mathbf{X}_t, \mathbf{Y}_t)$  coincides with the *probability* that the populations have the same type at a *particular* site, or equivalently that  $s(\mathbf{X}_t, \mathbf{Y}_t)$  is the *expected* proportion of sites at which the two populations have the same state.

A system of differential equations can be derived for  $s(\mathbf{X}_t, \mathbf{Y}_t)$  by means of a case-by-case analysis. For the moment, just consider time periods between population splits, so that the only processes to analyse are mutations and migrations. For each region  $X$ , let  $N_X$  denote the set of neighboring vertices (excluding  $X$  itself) and let  $k(\mathbf{X}) = |N_X|$ . Let  $r$  denote the rate of mutation (loss) of an allele at a single locus and  $a$  the rate at which a region  $\mathbf{X}$  adopts a type (borrows) from one of its  $k(\mathbf{X})$  neighbors.

We then have

$$\frac{ds(\mathbf{X}_t, \mathbf{Y}_t)}{dt} = -2rs(\mathbf{X}_t, \mathbf{Y}_t) \quad (1)$$

$$+ (1 - s(\mathbf{X}_t, \mathbf{Y}_t))a(1/k(\mathbf{X}) + 1/k(\mathbf{Y})) \quad (2)$$

$$+ \frac{a}{k(\mathbf{X}) - 1} \sum_{\mathbf{Z} \in N_{\mathbf{X}} - \{\mathbf{Y}\}} (1 - s(\mathbf{X}_t, \mathbf{Y}_t))s(\mathbf{Y}_t, \mathbf{Z}_t) - s(\mathbf{X}_t, \mathbf{Y}_t)(1 - s(\mathbf{Y}_t, \mathbf{Z}_t)) \quad (3)$$

$$+ \frac{a}{k(\mathbf{Y}) - 1} \sum_{\mathbf{Z} \in N_{\mathbf{Y}} - \{\mathbf{X}\}} (1 - s(\mathbf{X}_t, \mathbf{Y}_t))s(\mathbf{X}_t, \mathbf{Z}_t) - s(\mathbf{X}_t, \mathbf{Y}_t)(1 - s(\mathbf{X}_t, \mathbf{Z}_t)). \quad (4)$$

Here (1) corresponds to the loss of identity following mutation in  $X$  or  $Y$ ; (2) follows from the gain in identity when  $\mathbf{X}$  or  $\mathbf{Y}$  obtain a state from each other; (3) corresponds to the event when a state is transferred from a neighbor of  $\mathbf{X}$  to  $\mathbf{X}$  which either restores or removes identity<sup>1</sup>; while (4) is the symmetric case for  $\mathbf{Y}$ . Unaware of Sankoff's work, Bryant re-derived analogous equations in [6], a paper modeling network breaking in the Polynesian languages.

Sankoff showed that if we are provided with the adjacency graph at time 0 (the present) as well as the quantities  $s(\mathbf{X}_0, \mathbf{Y}_0)$  for all  $\mathbf{X}, \mathbf{Y}$ , the entire history of splittings can be reconstructed. The proof works by induction on the number of populations. Let  $\mathbf{U}, \mathbf{V}$  be the regions created in the most recent split, and suppose that this occurred at time  $\tau$ . We then have

- $s(\mathbf{U}_\tau, \mathbf{V}_\tau) = 1$ ;
- $s(\mathbf{X}_\tau, \mathbf{Y}_\tau) < 1$  for all  $X, Y$  not resulting from a split at time  $\tau$ ;
- $s(\mathbf{X}_t, \mathbf{Y}_t) < 1$  for all  $t > \tau$  and all  $X, Y$  including  $\mathbf{U}, \mathbf{V}$ .

The case of multiple splittings at exactly the same time can be dealt with by picking one pair arbitrarily.

The values  $s(\mathbf{X}_t, \mathbf{Y}_t)$  can be computed for  $t \geq \tau$  by solving the initial value problem (1)–(4) with initial values  $s(\mathbf{X}_0, \mathbf{Y}_0)$ . In this way,  $\mathbf{U}, \mathbf{V}$  and  $\tau$  can be identified as well as all of the values  $s(\mathbf{X}_\tau, \mathbf{Y}_\tau)$ . Replacing  $\mathbf{U}$  and  $\mathbf{V}$  by a single population we continue to obtain the second most recent splitting, and so on.

The analysis in Sankoff's paper assumes that the rate of mutation  $r$  and the rate of adoptions  $a$  are known. It is clear that we cannot identify both parameters given just the similarity values  $s(\mathbf{X}_0, \mathbf{Y}_0)$ : if we scale the rates and times simultaneously we can obtain identical similarity values. In fact, the situation is even more difficult. It was shown in [19] that in some cases one of these rates cannot be identified even though the other is known.

Sankoff makes the convenient (and acknowledged) assumption that the probabilities of identity  $s(\mathbf{X}_t, \mathbf{Y}_t)$  are known without error. Any serious application of the approach to real data will require some degree of uncertainty quantification. In a paper on Polynesian languages [6], Bryant used parametric bootstrapping to estimate variance in parameter estimates. Unfortunately, parametric bootstrapping is compu-

<sup>1</sup> A typo in Sankoff's original version of (3) was pointed out, rather excitedly, by [11].

tationally inefficient, and it can be problematic when faced with substantial model error on top of sampling error.

### 3.3 Multiway similarities

One way to potentially address the problem of parameter estimation and uncertainty quantification is to follow the lead of Markov random field theory (e.g. [16, pg 76]) and compute probabilities of identity for not just pairs, but triples and larger sets of regions.

Let  $\mathcal{R}_t$  denote the set of regions present at time  $t$ . For  $\mathcal{X}_t \subseteq \mathcal{R}_t$  we let  $s(\mathcal{X}_t)$  denote the probability that all of the regions  $\mathbf{X}_t \in \mathcal{X}_t$  have the same state at a site. Hence  $s(\{\mathbf{X}_t, \mathbf{Y}_t\}) = s(\mathbf{X}_t, \mathbf{Y}_t)$  for all pairs  $\mathbf{X}_t, \mathbf{Y}_t$ . We note that if we define

$$\delta(\mathcal{X}) = 1 - s(\mathcal{X})$$

for all subsets  $\mathcal{X}$  then (for generic  $t$ ) the function  $\delta$  satisfies the properties of a *diversity* [7], that is,  $\delta(\mathcal{X}) \geq 0$ ,  $\delta(\mathcal{X}) = 0$  if and only if  $|\mathcal{X}| = 1$  and

$$\delta(\mathcal{X} \cup \mathcal{Y}) \leq \delta(\mathcal{X} \cup \mathcal{Z}) + \delta(\mathcal{Y} \cup \mathcal{Z})$$

whenever  $\mathcal{Z} \neq \emptyset$ . This gives us access to a small, but growing, set of tools and theorems to aid analysis and computation.

Our main observation here though is that Equations (1)–(4) translate quite elegantly to this new context. Instead of computing probabilities for pairs of regions, we compute  $s(\mathcal{X}_t)$  for a set of regions  $\mathcal{X}_t$ , this being the probability that *all* regions in the set have the same type of allele at time  $t$ . It is now a straightforward matter to derive the differential equations for the probabilities  $s(\mathcal{X}_t)$ .

$$\frac{ds(\mathcal{X}_t)}{dt} = -|\mathcal{X}_t|rs(\mathcal{X}_t) \quad (5)$$

$$+ (1 - s(\mathcal{X}_t)) \sum_{\mathbf{X} \in \mathcal{X}} \sum_{\mathbf{Y} \in N_{\mathbf{X}}} a \frac{1}{k(\mathbf{X})} s((\mathcal{X}_t \setminus \{\mathbf{X}\}) \cup \{\mathbf{Y}\}) \quad (6)$$

$$s(\mathcal{X}_t) \sum_{\mathbf{X} \in \mathcal{X}} \sum_{\mathbf{Y} \in N_{\mathbf{X}} - \mathcal{X}} a \frac{1}{k(\mathbf{X})} (1 - s((\mathcal{X}_t \setminus \{\mathbf{X}\}) \cup \{\mathbf{Y}\})). \quad (7)$$

Here, (5) captures the rate at which mutations occur within  $\mathcal{X}_t$ ; (6) captures the rate by which a migration removes a dissimilar element from within  $\mathcal{X}_t$ , thereby making all regions have the same allele type; (7) captures the rate at which migrations from outside  $\mathcal{X}_t$  introduce non-identical allele types.

A convenient property of (5)–(7) is that the equation for  $\frac{ds(\mathcal{X})}{dt}$  involves only variables  $s(\mathcal{X}')$  with  $|\mathcal{X}'| \leq |\mathcal{X}|$ . Extending Sankoff's approach to triples or quadruples of regions will not generate an exponential explosion in complexity.

## 4 Alternative methods for analysis

### 4.1 *The structured coalescent*

Given that 40 years have passed since Sankoff's model was published, we might expect dramatic progress in the tools we can bring to the analysis. Here we briefly consider the range of modern approaches which we might use to carry out inference with the splitting model.

Perhaps the biggest methodological breakthrough in population genetics is due to *coalescent theory*, originally published by Kingman [17] but greatly advanced by a large number of mathematicians and statisticians. Let  $S$  denote the sequence data,  $G$  the geography (adjacency graph and splittings) and  $T$  the genealogical tree for a particular site. The tree is affected by both splittings and borrowings. Using the structured coalescent (e.g. [25]) one obtains a distribution  $P(T|G)$  of the tree given the adjacency graph and splittings. Following a Bayesian analysis, one uses Monte Carlo algorithms to simulate values from the joint posterior distribution

$$\begin{aligned} P(G, T|S) &\propto P(S|G, T)P(G, T) \\ &= P(S|T)P(T|G)P(G). \end{aligned}$$

The posterior distribution  $P(G|S)$  follows directly.

One open problem is the calculation of  $P(T|G)$ : given a splitting process and the resulting adjacency graph, what is the distribution for a genealogical tree. If the rate of borrowing or migration is low, then  $P(T|G)$  will primarily reflect the tree formed from the splittings themselves. As the rate of borrowing increases, the distribution will become more diffuse.

It may be useful to combine the multi-way similarity method of the previous section with the structured coalescent. Consider a subset  $\mathcal{X}$  of regions and let  $L(\mathcal{X})$  be the length of a gene tree with tips corresponding to these regions. The probability of all regions sharing the same trait is the probability of no mutation along the length of this tree, or  $e^{-L(\mathcal{X})}$ . Hence for a random gene tree,

$$s(\mathcal{X}) = E[e^{-L(\mathcal{X})}].$$

This connection could, for example, be used to check convergence when sampling gene trees.

### 4.2 *Stochastic diffusion methods*

The stochastic diffusion strategy [20, 21] is to rearrange the conditional probabilities, to give

$$P(G, T|S) \propto P(S|G, T)P(G, T) \quad (8)$$

$$= P(S|T)P(G|T)P(T). \quad (9)$$

The difference is that we now have to calculate  $P(G|T)$  instead of  $P(T|G)$ . This method has been widely used, including several high profile applications (e.g. [4, 26]). The accompanying software produces beautiful graphics depicting phylogenies on maps.

In itself, (9) is, of course, completely correct, and is a simple consequence of conditional probabilities. The problem is the computation of  $P(G|T)$ . In [20] it is assumed that computing  $P(G|T)$  is a simple matter of adapting standard algorithms for phylogenetic characters. Kühnert et al. [18] dub these ‘mugration’ models since they analyze migrations using models designed for mutations.

They, and a large number of earlier likelihood-based analysis, have made a critical error in their probability calculations. When they calculate the probability of the geographic locations on the tree, they assume that the same conditional independence underlying likelihood calculations for genetic data also applies for geographic characters. However by conditioning on a tree they are making an implicit assumption that all lineages survive to the present.

If you consider an ancestral lineage, the rate of migration to a given island is one thing, the rate of migration *conditional on survival to the present* is another. If the islands (or regions) are small then any lineage is likely to be either lost or fixed. Hence if one lineage in the tree occupies a particular island it is highly unlikely that any other lineage will occupy the island at the same time. This conflict breaks down the conditional independence that is so critical for the calculation of phylogenetic likelihoods. Furthermore, ignoring these conflicts is essentially equivalent to ignoring the interplay between geographic structure and drift, assuming infinite effective population sizes within each region.

### 4.3 Approximate Bayesian Computation

Approximate Bayesian computation (ABC) [2, 8, 9] has grown immensely in popularity, partly because it (at first glance) doesn’t require much specialist mathematical knowledge to implement and partly because it is possible to set up analyses for extremely complex models fairly simply.

The general idea is to use simulations in place of likelihood calculations, and it can be proven that, with sufficient iterations and sufficient summary statistics, the approach provably converges to the correct posterior distribution. The approach has a simplicity and transparency which makes it especially attractive. There are, unfortunately, serious issues in higher dimensional space, though recent techniques have a great deal of potential [1]. Until these issues are resolved, we suspect that we will not have the computational power to do more than infer, roughly, even the smallest splitting history.

## 5 Future work

We see two principal methodological advances made in Sankoff's work. The first is the idea to model adjacencies not by a single graph, but by a sequence of vertex splittings generating a graph. The model uses an abstraction of the spacial component, in the sense that the vertices in the graph are not given specific geographic locations. The potential advantage of this could be tractable inferential methods which still capture aspects of geography essential to phylogenetics.

The second advance is the method for computing pairwise similarities. [3] had investigated this problem for a general migration matrix, but only derived approximate results. Sankoff's approach computes exact probabilities incorporating all regions simultaneously. We have shown that this extends beyond pairwise comparisons, and that the approach has not been superseded by modern developments, though many practical computational and statistical problems remain.

## References

1. Mark A. Beaumont, Rasmus Nielsen, Christian Robert, Jody Hey, Oscar Gaggiotti, Lacey Knowles, Arnaud Estoup, Mahesh Panchal, Jukka Corander, and Mike Hickerson. In defence of model-based inference in phylogeography. *Molecular Ecology*, 19(3):436–446, 2010.
2. Mark A. Beaumont, Wenyang Zhang, and David J. Balding. Approximate bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.
3. Walter F. Bodmer and Luigi L. Cavalli-Sforza. A migration matrix model for the study of random genetic drift. *Genetics*, 59(4):565–592, August 1968. PMID: 5708302 PMCID: PMC1212023.
4. Remco Bouckaert, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard, and Quentin D. Atkinson. Mapping the origins and expansion of the indo-european language family. *Science*, 337(6097):957–960, 2012.
5. Susanne C. Brenner and Ridgway Scott. *The Mathematical Theory of Finite Element Methods*. Springer, 2008.
6. David Bryant. Radiation and network breaking in polynesian linguistics. In Peter Forster and Colin Renfrew, editors, *Phylogenetic methods and the prehistory of languages*, pages 111–118. McDonald Institute for Archeological Research, 2006.
7. David Bryant and Paul F. Tupper. Hyperconvexity and tight-span theory for diversities. *Advances in Mathematics*, 231(6):3172–3198, 2012.
8. Jean-Marie Cornuet, Filipe Santos, Mark A. Beaumont, Christian P. Robert, Jean-Michel Marin, David J. Balding, Thomas Guillemaud, and Arnaud Estoup. Inferring population history with DIY ABC: a user-friendly approach to approximate bayesian computation. *Bioinformatics*, 24(23):2713–2719, 2008.
9. Katalin Csilléry, Michael GB Blum, Oscar E. Gaggiotti, and Olivier Fran\ccois. Approximate bayesian computation (ABC) in practice. *Trends in ecology & evolution*, 25(7):410–418, 2010.
10. Roland L'vovich Dobrushin. Gibbsian random fields for lattice systems with pairwise interactions. *Functional Analysis and its applications*, 2(4):292–301, 1968.
11. Sheila Embleton. Lexicostatistical tree reconstruction incorporating borrowing. *Toronto Working Papers in Linguistics*, 2, 1981.
12. Joseph Felsenstein. A pain in the torus: Some difficulties with models of isolation by distance. *The American Naturalist*, 109(967):359–368, May 1975. ArticleType: research-article / Full publication date: May - Jun., 1975 / Copyright © 1975 The University of Chicago.

13. Gilles Guillot, Raphael Leblois, Aurelie Coulon, and Alain C. Frantz. Statistical methods in spatial genetics. *Molecular Ecology*, 18(23):4734–4756, 2009.
14. Hugues Hoppe, Tony DeRose, Tom Duchamp, John McDonald, and Werner Stuetzle. Mesh optimization. 1993.
15. Motoo Kimura and James F. Crow. The number of alleles that can be maintained in a finite population. *Genetics*, 49(4):725, 1964.
16. Ross Kindermann and James Laurie Snell. *Markov random fields and their applications*. American Mathematical Society Providence, RI, 1980.
17. John FC Kingman. The coalescent. *Stochastic processes and their applications*, 13(3):235–248, 1982.
18. Denise Kühnert, Chieh-Hsi Wu, and Alexei J. Drummond. Phylogenetic and epidemic modeling of rapidly evolving infectious diseases. *Infection, Genetics and Evolution*, 11(8):1825–1841, 2011.
19. Jamie Kydd. *The effect of horizontal Transfer and borrowing on phylogenetic signal*. Honours dissertation, University of Auckland, Auckland, 2007.
20. Philippe Lemey, Andrew Rambaut, Alexei J. Drummond, and Marc A. Suchard. Bayesian phylogeography finds its roots. *PLoS computational biology*, 5(9):e1000520, 2009.
21. Philippe Lemey, Andrew Rambaut, John J. Welch, and Marc A. Suchard. Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution*, 27(8):1877–1885, 2010.
22. Jesús A. De Loera, Jörg Rambau, and Francisco Santos. *Triangulations: Structures for Algorithms and Applications*. Springer, 2010.
23. Gustave Malécot. *The Mathematics of Heredity*. Freeman, 1970.
24. Stan Melax. A simple, fast, and effective polygon reduction algorithm. *Game Developer*, 5(11):44–49, 1998.
25. M. Notohara. The coalescent and the genealogical process in geographically structured population. *Journal of Mathematical Biology*, 29(1):59–75, 1990.
26. Chinyere K. Okoro, Robert A. Kingsley, Thomas R. Connor, Simon R. Harris, Christopher M. Parry, Manar N. Al-Mashhadani, Samuel Kariuki, Chisomo L. Msefula, Melita A. Gordon, Elizabeth de Pinna, John Wain, Robert S. Heyderman, Stephen Obaro, Pedro L. Alonso, Inacio Mandomando, Calman A. MacLennan, Milagritos D. Tapia, Myron M. Levine, Sharon M. Tennant, Julian Parkhill, and Gordon Dougan. Intracontinental spread of human invasive salmonella typhimurium pathovariants in sub-saharan africa. *Nature Genetics*, 44(11):1215–1221, 2012.
27. J. J. Robledo-Arnuncio and F. Rousset. Isolation by distance in a continuous population under stochastic demographic fluctuations. *Journal of Evolutionary Biology*, 23(1):53–71, January 2010.
28. David Sankoff. Reconstructing the history and geography of an evolutionary tree. *The American Mathematical Monthly*, 79(6):596–603, June 1972. ArticleType: research-article / Full publication date: Jun. - Jul., 1972 / Copyright © 1972 Mathematical Association of America.
29. Mike Steel and Andy McKenzie. Properties of phylogenetic trees generated by yule-type speciation models. *Mathematical biosciences*, 170(1):91, 2001.
30. E. Steinitz and H. Rademacher. *Vorlesungen über die Theorie der Polyeder*. Springer, Berlin, 1934.
31. J. G. Sumner and P. D. Jarvis. Entanglement invariants and phylogenetic branching. *Journal of mathematical biology*, 51(1):18–36, 2005.
32. Sewall Wright. Isolation by distance. *Genetics*, 28(2):114–138, March 1943. PMID: 17247074 PMID: PMC1209196.