Comparing Phylogeographies: Incompatible Geographical Histories in Pathogens' Genomes

Benjamin Singer^{1,*}, Antonello Di Nardo^{2,*}, Jotun Hein³, and Luca Ferretti^{2,4,†}

¹ *Department of Zoology, University of Oxford, UK*

² *The Pirbright Institute, Pirbright, Woking, Surrey, UK*

³ *Department of Statistics, University of Oxford, UK*

⁴ *Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield
Department of Medicine, University of Oxford, UK*

**These authors contributed equally.*

[†] *e-mail: luca.ferretti@bdi.ox.ac.uk, luca.ferretti@gmail.com*

Abstract

Modern phylogeography aims at reconstructing the geographic diffusion of organisms based on their genomic sequences and spatial information. Phylogeographic approaches usually ignore the possibility of recombination, which decouples the evolutionary and geographic histories of different parts of the genome. Genomic regions of recombining or reassorting pathogens often originate and evolve at different times and locations, which characterised their unique spatial histories. Measuring the extent of these differences requires new methods to compare geographic information on phylogenetic trees reconstructed from different parts of the genome. Here we develop for the first time a set of measures of *phylogeographic incompatibility* aimed at detecting differences between geographical histories in terms of distances between phylogeographies. We study the effect of varying demography and recombination on phylogeographic incompatibilities using coalescent simulations. We further apply these measures to the evolutionary history of human and livestock pathogens, either reassorting or recombining, such as the Victoria and Yamagata lineages of influenza B and the O/Ind-2001 foot-and-mouth disease virus strain. Our results reveal diverse geographical paths of diffusion that characterise the origins and evolutionary histories of different viral genes and genomic segments. phylogeography, recombination, viral evolution

1 Introduction

The study of the evolutionary biology and phylodynamics of pathogens has made significant contributions to disease control over the last few decades [Grubaugh et al., 2019, Baele et al., 2018]. Pathogens are dispersed with their human and animal hosts, both of which are seeing increasing mobility as an effect of globalisation and the access of emerging economies to international markets. This condition offers an increasing opportunity for some pathogens to spread rapidly at the global level, influenza being one of the best known examples [Lemey et al., 2014]. The elevated mobility pattern combines with high evolution rates in some pathogens to couple the processes of evolution and spatial diffusion in these organisms. To understand the emergence, adaptation and spread of pathogens, therefore, it is possible to use the methods provided by the field of modern

phylogeography, that jointly infer the evolutionary history (i.e. phylogeny) and the spatial history (i.e. geographic diffusion) of an organism [Dudas et al., 2017].

Phylogeographic methods can make use of the mathematical representation of a phylogeny as a tree, allowing the spatial evolution from the root to the tips to be seen as a diffusion or migration process. This diffusion process proceeds independently on each branch of the tree, with a single origin at the root. The resulting structure greatly simplifies mathematical and computational treatments of systems composed of coupled evolution and migration processes. However, the phylogenetic tree of a set of DNA or protein sequences contains information not only on the relatedness of sequences, but also on their evolutionary history. There exist many different methods for reconstructing such a tree. If genome isolates have been collected from different places, their history could be affected by the geographic structure of the populations, including the rates of migration across different locations [Lemey et al., 2014]. The first step to understanding these effects is the reconstruction of a phylogenetic tree with geographical information attached to the internal nodes, obtained through reconstruction of the spatial movements of lineages in the tree. Such an object corresponds to the phylogeographic history of the given set of sequences. For simplicity, in this paper we will refer to it as a *phylogeography*.

A range of techniques have been developed for the reconstruction of phylogeographies [Faria et al., 2011]. These can be classified according to how the migration process of ancestral lineages is modelled on the tree: either in the form of instantaneous diffusion of discrete traits (such as finite spatial locations), i.e. discrete phylogeography [De Maio et al., 2015, Kühnert et al., 2016, Lemey et al., 2009, Müller et al., 2017]; or as relaxed random walk diffusion in two-dimensional geographic space, i.e. continuous phylogeography [Bouckaert, 2016, Dellicour et al., 2016, Gill et al., 2017, Lemey et al., 2010]. Many of these approaches focus on a single tree, obtained from concatenation based phylogeny or selected sections of a genome, and therefore do not take into account the decoupling of the evolutionary history of different genomic regions due to recombination and/or related processes.

For many species of eukaryotes, as well as for some viruses and bacteria, phylogenies are not tree-like. In the presence of recombination, reassortment or horizontal gene transfer, different regions of the genome might have potentially evolved from unique ancestors. This decouples the evolutionary histories of different genes or regions of the genome, resulting in trees with cross-branch reticulations if the number of recombination events is small, or in phylogenetic networks if recombination is widespread. Recombination leads to organisms that inherit genomic fragments from multiple parent lineages, causing phylogenetic inference methods applied to different parts of the genome to give rise to incompatible trees. By applying a measure of tree distance to these trees, it is possible to make certain claims about the amount of recombination in the ancestral population [Chung et al., 2013], for instance by setting a lower bound on the number of recombination events between loci. Any non-zero tree distance reveals an incompatibility. If the genealogies of different parts of the same genome are incompatible, these incompatibilities can be evidence of recombination, horizontal gene transfer, reassortment, or gene conversion. A variety of incompatibility measures and tree distances exist for phylogenies, most notably the well-known Robinson-Foulds metric and the family known as tree edit (or Levenshtein) distances [Day, 1984], including the subtree prune and regraft (SPR) distance [Semple and Steel, 2003].

Recombination can lead to a decoupling of geographic histories between loci, just as it can lead to a decoupling of phylogenetic histories. The resulting ensemble of spatial trajectories may reveal functional and evolutionary features of different genes. However, phylogeographic inference is highly non-trivial in the presence of reticulations. Reticulations form phylogenetic loops, meaning that two lineages can split at some time in the past, move in geographic space, then converge to the same location and combine into a single lineage. They can share pieces of genetic material with different geographical origins, or copies of material which have reached their current geographical distribution

and evolutionary trajectory in different ways. This means that the spatial evolution of a given pair of lineages can no longer be assumed independent, and some of the advantages of working with diffusion processes on a tree are lost.

The purpose of this paper is to propose and define two mathematical measures of the degree of difference between reconstructed phylogeographies arising from different loci. Only one of these is a true distance in the sense of providing a metric on a vector space, so we will use the more general term ‘incompatibility measure’. A non-zero value of an incompatibility measure between two phylogeographies implies that they are incompatible, and is evidence of both recombination and different patterns of spatial diffusion between loci. Comparing incompatibilities between the phylogeographies of different loci can reveal which regions tend either not to recombine much, or to remain associated during spatial diffusion and those which, conversely, have highly divergent spatial histories. We apply these measures of phylogeographic incompatibility to reconstructed histories of a reassorting human pathogen, the influenza B virus, as well as a recombining animal pathogen, the foot-and-mouth disease virus (FMDV).

2 New Approaches

Any measure of phylogeographic incompatibility will inevitably be related to phylogenetic distances, due to the mathematical relation between phylogeographies and phylogenies. In this section we outline our approach to calculating phylogeographic incompatibilities, starting with a comparison to the phylogenetic case and detailing two possible ways to incorporate geographic information. We then provide an approach that accounts for uncertainties in phylogeographic reconstruction.

2.1 Phylogenetic incompatibilities

Many different measures of tree distance/incompatibility have been developed in the past. In most comparisons, we will use only the two measures most closely related to the phylogeographic incompatibilities introduced in the next sections.

The first is the path distance measure (hereinafter referred to as “path”) introduced by [Steel and Penny, 1993]. This measure compares the heights of the most recent common ancestors (MRCAs) of each pair of leaves for the two trees and computes the norm of their differences. Since most pairwise MRCAs are localised close to the root, the path distance is strongly affected by topological variation of the upper part of the tree.

The second is the natural distance based on the size of the maximum agreement sub-tree (“MAST”) between the two trees [Steel and Warnow, 1993]. The MAST estimate is very large for similar trees, while it is computed using fewer and fewer leaves as the dissimilarity between the trees increases. The MAST distance is dramatically sensitive to changes near the root but it is also generally sensitive to variability across the whole tree. Note that the normalised MAST distance is usually defined as the fraction of tips that do *not* belong to the MAST.

We also consider four other distance measures defined on phylogenies (without a geographical element):

- Robinson-Foulds metric (“RF”) [Robinson and Foulds, 1981]
- Robinson-Foulds metric weighted by branch lengths (“WRF”) [Robinson and Foulds, 1981]
- subtree-prune-and-regraft edit distance (“SPR”) [De Oliveira Martins et al., 2014, Allen and Steel, 2001]

- Kuhner-Felsenstein branch score distance (“KF”) [Kuhner and Felsenstein, 1994]

Each of these incompatibility measures defines a proper metric in the space of rooted binary trees with branches of finite length. However, all these distances capture different ways in which phylogenies can differ. Given the high-dimensional and heterogeneous nature of phylogenetic spaces, it is not surprising that there is little agreement among the patterns of distances, as illustrated in the Appendix. As an example, trees differing by the topology of the upper half (near the root) tend to have a high MAST distance and an comparatively even larger path distance; the opposite is true for trees differing in the topology of the lower half (near the tips), which tend to be close in terms of path distance but differ sensibly in terms of MAST distance.

Throughout this paper, we used the version of these measures adapted to rooted binary trees. These measures are currently implemented in the R packages *APE* [Paradis and Schliep, 2018] and *phangorn* [Schliep, 2010].

2.2 Phylogeographic space

The first step in building natural methods of comparison for phylogeographies is to define what a phylogeography is and to situate it in a well-defined mathematical space. The incompatibility measures should facilitate the definition of distance measures over this space. Such a space needs to be able to represent both phylogenetic and geographical information.

We define a phylogeography as a tree with geographical discrete information assigned to each node of the tree. For practical purposes, inferred phylogenies are represented by rooted phylogenetic trees with branch lengths and labelled tips. Geographical space can be represented either as a network of discrete locations or as a continuous surface. Our approach covers both cases. The question then becomes how to combine these spaces, and how to define a joint measure on the resulting combination. These two questions are not independent, as different spaces require different approaches for comparing phylogeographies.

We present a simple definition of phylogeographic space. The most general space is denoted by $\mathcal{PG}_n$, with integer n , and its elements are rooted binary trees with n labelled leaves and with locations assigned to all nodes (both leaves and internal nodes). The illustrations in figure 1 are of phylogeographies of this kind. This space corresponds to the Cartesian product of the space of bifurcating rooted trees with n leaves $\mathcal{T}_n$ and the Cartesian power of the geographic space $\mathcal{G}$ to the number of nodes, i.e. $\mathcal{PG}_n = \mathcal{T}_n \times \mathcal{G}^{2n-1}$. In practice, we will work on subspaces of $\mathcal{PG}_n$ with a given vector of tip locations $\mathbf{g}$, since the locations of the leaves are usually known and given as input into phylogeographic analyses. For a given set of locations $\mathbf{g}$, the corresponding phylogeographic subspace $\mathcal{PG}_{\mathbf{g}}$ is isomorphic to $\mathcal{T}_n \times \mathcal{G}^{n-1}$. In the following, when discussing $\mathcal{PG}_{\mathbf{g}}$ for a specific (although implicit) choice of $\mathbf{g}$, we will refer to it simply as $\mathcal{PG}$.

2.3 Pairwise MRCA Incompatibilities

The geographic space $\mathcal{G}$ has a metric $d_{\mathcal{G}}(g, g')$ already defined on it. We can use this distance to define a purely geographic measure of incompatibilities on $\mathcal{PG}$. However, this requires a pairwise comparison of spatial locations. Since locations are assigned to nodes on different trees, we must solve the matching of nodes across trees with different shapes. The simplest approach is to map the MRCAs for pairs of leaves, i.e. compare nodes which are the MRCA of the same pair of leaves on different trees.

The first step to performing such a comparison for a pair of trees with n leaves is to project the phylogeography PG to a $\binom{n}{2}$ -ary vector $\mathbf{g}_{MRC A}(PG)$, where the element $g_{MRC A(l_1, l_2)}$ is the location of the MRCA of the tips l_1 and l_2 . Such vectors span a subspace of the $\binom{n}{2}$ -ary Cartesian power of

$\mathcal{G}$. The L_1 product metric is a natural metric for this space and therefore it can be used to compute geographic incompatibilities between phylogeographies.

The incompatibility measure between two phylogeographies PG_1 and PG_2 is then defined as

$$I_{MRC A}(PG_1, PG_2) = \frac{\sum_{\{l_1, l_2\}} d_{\mathcal{G}}(g_{MRC A(l_1, l_2)}(PG_1), g_{MRC A(l_1, l_2)}(PG_2))}{\binom{n}{2}}, \quad (1)$$

where $\{l_1, l_2\}$ range over the set of pairs of leaves, $g_{MRC A(l_1, l_2)}(PG_k)$ is the location assigned to the MRCA of leaves l_1 and l_2 in the phylogeography PG_k , and $d_{\mathcal{G}}$ is the metric on the geographic network. The normalisation factor $\binom{n}{2}$ ensures that the minimum and maximum possible values of the measure are 0 and $\max_{g, g'} d_{\mathcal{G}}(g, g')$, respectively.

For a simple example of how to calculate $I_{MRC A}$, consider the phylogeographies in Figure 1a. Take A , B , and C to refer to the arbitrary locations Argentina, Bolivia and Chile, and place these countries on a network where there is a pairwise distance of 1 between each pair, since they all share borders. The MRCA vectors are therefore $g_{MRC A}(PG_1) = (B, A, A, A, A, A)$ for the phylogeography on the left, and $g_{MRC A}(PG_2) = (A, A, A, B, B, B)$ for the phylogeography on the right. So the total incompatibility is $I_{MRC A}(PG_1, PG_2) = \frac{4d_{\mathcal{G}}(A, B)}{6} = 2/3$. Here, the incompatibility comes about due to four comparisons between Argentina and Bolivia, and the only MRCA paths that do not contribute to the incompatibility are between l_1 and l_3 , and between l_1 and l_4 , which meet at the root in both phylogeographies.

$I_{MRC A}$ is especially sensitive to the locations assigned to nodes with many descendants, the strongest example of this being the root of the tree, which can contribute to half or more of the comparisons made in calculating the incompatibility. This sensitivity depends in part on the tree shape, with the root contributing more in symmetric trees than in asymmetric ones. This is because leaf pairs are more likely to find their MRCA at the root when the tree is more symmetric, and more likely to find their MRCA on the trunk of the tree when the tree is less symmetric.

Note that $I_{MRC A}$ is not a distance on $\mathcal{PG}$, since trees with a different topology may still have no incompatibilities. However, this issue is easily solved. In fact, any positive linear combination of $I_{MRC A}$ and a tree distance on $\mathcal{T}_n$ is a proper metric on $\mathcal{PG}$.

2.4 Maximum Agreement Sub-Phylogeography

To incorporate both phylogenetic and geographic differences into a metric in a unified way, an alternative approach is to extend an existing metric on $\mathcal{T}_n$ into $\mathcal{PG}_n$. One of the simplest measures of phylogenetic incompatibility is based on the notion of a maximum agreement subtree, which we will here extend to phylogeographies.

As other phylogenetic distances, this accounts for the similarity between different trees based on shared subtrees. There already exists a fairly comprehensive body of literature on finding MASTs between pairs of trees [Hein et al., 1996, Steel, 2016, Semple and Steel, 2003]. An agreement subtree between the trees T_1 and T_2 with labelled leaves is a tree S , with labelled leaves, such that the subtrees induced on T_1 and T_2 by the set of leaves corresponding to those in S are both identical to S . The MAST is the largest such subtree. Figure 1b provides an example of a MAST.

We can use the MAST to define a metric over $\mathcal{T}_n$. To find the distance between T_1 and T_2 , we start to find $MAST(T_1, T_2)$, and then write $d_{MAST}(T_1, T_2) = n - |MAST(T_1, T_2)|$. Here, $|T|$ is the number of leaves on T , and $n = |T_1| = |T_2|$. Thus, for the trees in the example in Figure 1b, $d_{MAST}(T_1, T_2) = 1$, since the trees are of size five and the MAST is of size four.

Now we extend this distance to a metric on $\mathcal{PG}$. First, we define a maximum agreement sub-phylogeography (MASPG). Take two phylogeographies PG_1 and PG_2 , with locations assigned to

(a) Two phylogeography diagrams illustrating the MRCA-based incompatibility measure. Paths to MRCA of pairs of leaves with conflicting locations are highlighted in various colours, and the nodes causing the incompatibility are likewise highlighted. There are four conflicting MRCAs here, so the unweighted incompatibility is 4.

(b) Two phylogeography diagrams illustrating the difference between the MAST (in yellow) and the MASPG (in green). The MASPG includes internal node locations, and the internal nodes that contribute to the MASPG are highlighted in green. Due to the requirement of matching internal node locations, the MASPG has size 3, while the MAST, which does not have this constraint, has size 4.

Figure 1: An illustrations of the two measures of phylogeographic incompatibility presented in this paper: (a) pairwise MRCA incompatibility, (b) Maximum Agreement PhyloGeography (MASPG).

each node. Let us define an agreement sub-phylogeography as a sub-phylogeography without unary nodes such that (i) the corresponding subtree is an agreement subtree and (ii) for each node in the subtree, the location assigned to the node matches the locations on the corresponding subtrees in both PG_1 and PG_2 . That is, the induced subtrees must be identical not only in topology but also in the geographical assignment of the nodes. The MASPG is defined as the largest such sub-phylogeography.

We can then define the phylogeographic incompatibility between two phylogeographies PG_1 and PG_2 , both with leaf number $n = |PG_1| = |PG_2|$, as

$$I_{MASPG}(PG_1, PG_2) = 1 - |MASPG(PG_1, PG_2)|/n, \quad (2)$$

where $|T|$ is the number of leaves on the tree T . With this normalisation, the minimum and maximum possible values of I_{MASPG} are 0 and 1, respectively. For the phylogeographies in figure 1b the MASPG is of size three and the phylogeographies are of size five, so the incompatibility is 0.4. I_{MASPG} is a metric on $\mathcal{PG}_n$, as shown in Appendix 1.

2.5 Incompatibility between distributions of phylogeographies

In practical applications, phylogeographies are not known with absolute precision, but they are inferred from noisy and biased data (e.g. biased sampling of subpopulations). In fact, Bayesian approaches to phylogeographic inference actually return a posterior distribution of phylogeographies. For this reason, we consider the more general case of comparisons between two *distributions* of phylogeographies. In such a case, the objects of comparison are two random variables PG_1 and PG_2 derived from different loci along a genome. These variables are likely to be highly correlated even if their posterior (marginal) distributions are inferred independently, due to linkage between loci and the fixed assignment of tip locations from sample metadata. This causes strong biases if the expected value of $I_{MRC A}$ is computed assuming the two distributions as independent.

The simplest and most conservative choice is to compute the incompatibility as a Wasserstein metric (or earth mover's distance) between the distributions, using one of the incompatibility measures proposed above as cost function. This choice reduces the likelihood of spurious incompatibilities due to correlations between the inferred phylogeographies at different loci.

Hence, we can use $I_{MRC A}$ or I_{MASPG} as a cost function for differences between phylogeographies, yielding a definition of incompatibility between PG_1 and PG_2 as the first Wasserstein metric on the corresponding distributions. In practice, we estimate this by taking N samples from each posterior, pairing these samples in all possible $N!$ ways (e.g. s_1 with s'_1 , s_2 with s'_2 etc), computing the mean of the incompatibilities $\sum_{i=1}^N I(s_i, s'_i)/N$ between all pairs of samples from the two distributions, then minimising this quantity over all choices of pairings.

3 Results

3.1 Effect of evolutionary parameters on phylogeographic incompatibilities

To understand the impact of recombination and migration rates on phylogeographic incompatibilities, we generated random phylogeographies from structured coalescent models with three populations/locations. We simulated both coalescent trees and node locations of 20 sequences from each

Figure 2: Effect of evolutionary parameters on phylogeographic incompatibilities, using 100 simulations at each calculated pair of parameter values. The plots show smoothed surfaces of each phylogeographic incompatibility measure, labelled with the corresponding network structure. The network structures are illustrated with locations in which the population can originate coloured in yellow.

location, all sampled at the present time. Simulations were performed by varying population-scaled recombination rate ρ and migration rate μ , as well as migration patterns.

Incompatibilities are generated by different mechanisms for the patterns of migration represented e.g. in the models in Figure 2. In model (iii), there is only one possible chain of geographical transmissions for each lineage, but incompatibilities are generated by the different pace of lineages along this chain. These differences originate from neutral (stochastic) or selective differences in the patterns of coalescence or migration of the lineages. In model (iv), multiple patterns of transmission represent another important contribution to phylogeographic incompatibilities, with lineages following different routes between the source and the final location. In model (i), an additional source of incompatibility is provided by bidirectional spread, with lineages bouncing back and forth or wandering between distant locations.

Figure 2 reveals that the incompatibility between simulated trees increases with both recombination rate and migration rate, at least for low/moderate rates of migration. This is to be expected, since increased recombination implies more different trees, while low migration leads to less spatial transitions along branches and therefore more similar geographic histories.

In most models, changes in recombination rates have a much stronger impact on phylogeographic incompatibilities than changes in migration rates. This can be explained by the fact that recombination is necessary to actually observe a difference in migration histories. Without any topological variation of the tree, the phylogeographic reconstruction would not differ and any change in migration patterns would be undetectable. In fact, incompatibilities are very weak for recombination rates $\rho < 1$ and increase steeply for larger values of ρ . Incompatibilities tend to be weaker as well for low migration rates $\mu \ll 1$.

As outlined above, migration patterns have a major impact on incompatibilities, depending on the migration rate and the specific measure of incompatibility. Incompatibilities tend to be small if there is a single source population (Figure 2-ii), since this imposes strong constraints on the migration histories and on the origin of the ancestors of the sample. Scenarios with longer transmission chains (Figure 2-iii) or more complex patterns of transmission (Figure 2-iv) lead to a slight increase in differences among phylogeographies and a more complex dependence on migration rates. Finally, in scenarios with multiple bidirectional migration patterns, migration histories can be wildly different, hence incompatibilities tend to be much larger (Figure 2-i) and to increase monotonically but weakly with migration rates.

In some scenarios, different measures of incompatibilities show strikingly different trends. This is due to the fact that they capture different components of the signal. As an example, widespread differences in recent geographical history would give rise to a much stronger signal on MASPG than on MRCA incompatibilities.

3.2 Phylogeography of the O/ME-SA/Ind-2001 FMDV lineage epidemics

Foot-and-mouth disease is a viral disease affecting wild and domesticated cloven-hoofed animals [Alexandersen et al., 2003] and it is one of the most economically relevant diseases of livestock worldwide [Knight-Jones and Rushton, 2013]. It is caused by the FMDV, a positive-sense single-stranded RNA virus belonging to the family *Picornaviridae*. This virus has a relatively short (of about 8kb) but highly variable linear genome coding for 12 proteins with different functions [Jackson et al., 2003]. It provides an interesting case study for phylogeographic incompatibilities, since it has a non-trivial geographic structure, being spread out across host populations worldwide [Brito et al., 2017, Di Nardo et al., 2011], and a high recombination rate, especially between structural and non-structural proteins [Carrillo et al., 2005, Jackson et al., 2007].

Whole-genome sequences from a recently analysed FMDV lineage (the O/ME-SA/Ind-2001)

were provided by the World Reference Laboratory for FMD (WRLFMD) at the Pirbright Institute [Bachanek-Bankowska et al., 2018]. Phylogeographic analyses were performed on most loci along the genome, as well as on the WGS. The phylogenetic and phylogeographic incompatibilities are illustrated in Figure 3.

The two measures of phylogeographic incompatibility are only partially related. However, this is true for usual phylogenetic incompatibilities as well, as can be seen comparing the left and right panel of Figure 3.

Our results show that the portions of the genome coding for the capsid structure (comprising of the structural proteins from VP1 to VP4) have highly coupled histories, both phylogenetically and geographically. This fits with the well-known fact that this region exhibits a relatively low recombination rate [Carrillo et al., 2005, Jackson et al., 2007]. The non-structural genes at the 3' end of the genome (i.e. 2C, 3A, 3C, and 3D) also share similar phylogenetic and geographic histories. The majority of incompatibilities are located along the tree trunk of the main sublineage, as illustrated in Figure 3: they are mostly related to differences in reconstructed virus movements within the Indian subcontinent, as well as variability in the timing of transmission of different genomic regions to the Middle-East.

Some genomic regions share similar phylogenies, but clearly distinct phylogeographies. This is the case e.g. for the 5' UTR versus 2B-3A and for 2B versus 2C-3A, according to the MRCA measures estimated. In fact, topological similarity does not always reflect compatible geographical histories, and this relation is not reflected in the phylogenetic distances: for example, the WGS reconstruction is phylogenetically similar to the one from the L-protease, but the geographical histories are quite different.

It is important to note that the phylogeography created using the WGS does not provide a good summary of the phylogeographies of the individual genomic regions, and in fact it only resembles the phylogeographies from the genome segments comprising of the 2C and 3A proteins, which tend to be quite dissimilar to those of the rest of the sequences. This could be due to the effect produced on the WGS topology by recombinant viruses (i.e. the O/BAR/15/2015 and O/BHU/3/2016) [Bachanek-Bankowska et al., 2018], which provide the strongest contribution to incompatibilities (as is evident from the widest branch in Figure 3).

It is therefore important in FMDV evolutionary studies to take account of geographic as well as phylogenetic differences when comparing reconstructed evolutionary histories of distinct genomic regions outside of the capsid.

3.3 Spatial evolution of genomic segments of Influenza B

Influenza B virus is a segmented virus in the family *Orthomyxoviridae*. It is a major human pathogen, although it causes a lesser threat to public health than the related influenza A virus. Influenza B virus has a short positive-sense single-stranded RNA genome (of about 15 kb) organised in eight linear segments [Bouvier and Palese, 2008]. Reassortment of these segments plays a major role in the evolution of the virus, hence influenza B represents an interesting case study for phylogeographic incompatibilities in reassorting pathogens. Two main lineages (denoted as Victoria and Yamagata) are recognised based on differences in the hemagglutinin protein. It has been previously shown how these lineages evolved under different dynamics [Langat et al., 2017].

We analysed a set of 242 sequences (122 of the Victoria lineage, and 120 of the Yamagata) from a recent study of worldwide evolution of these lineages [Bedford et al., 2015]. These sequences include 5 genes (HA, NA, PB1 and PB2, and NS1 protein) located on different segments and therefore reassorting freely among them. Phylogenetic and phylogeographic incompatibilities among these segments are illustrated in Figure 4 for both lineages. We also reconstructed and compared the

Figure 3: Phylogenetic and phylogeographic incompatibilities estimated for the FMDV O/ME-SA/Ind-2001 epidemics. A panel - Comparisons of various incompatibility measures between phylogeographies reconstructed from segments of the FMDV O/ME-SA/Ind-2001 genome. In the plot on the left, I_{MRCA} (upper triangle) is compared to I_{path} (lower triangle). In the plot on the right, I_{MASPG} (upper triangle) is compared to I_{MAST} (lower triangle). B panel - Illustration of the incompatibilities along the MCC tree. The width of each branch represents the total amount of phylogeographic MRCA incompatibilities between all trees built from individual genomic regions and the whole-genome tree. Pie charts represent the distribution of MRCA incompatibilities per genomic region (left) and per location (right) on the corresponding nodes. Branch colors on the right represent the phylogeography reconstructed from the whole-genome sequence alignment.

phylogeographic history of the joint sequences of all these genes (8.6 kb).

Despite free reassortment between all segments, the structure of incompatibilities follows some clear patterns. Phylogeographic incompatibilities between segments are stronger for the Victoria strain, while for Yamagata these are comparable to the ones of a recombining virus such as FMDV (see Figure 6). This is partly a result of the large impact of reassortment on the phylogenetic trees of the Victoria lineage.

There is a strong mismatch between the evolutionary histories of the two proteins that are the most important in the immune response against the virus (i.e. of HA and NA). The phylogenetic histories of HA and NA are highly divergent in the Victoria lineage, but the difference between their geographical histories is less extreme. On the other hand, the geographical spread of HA and NA in the Yamagata lineage followed very different routes, despite intermediate levels of reassortment. The geographical history of NA in this lineage is, in fact, markedly different from all other genes.

Clear patterns of incompatibility differ between the two lineages, as shown also in Figure 5. PB1 and PB2 phylogeographies are closely related for Yamagata but very different for Victoria. HA and NA phylogeographies are related to NS1 for Victoria, but the three genes present very different phylogeographies for Yamagata. Hence, the Victoria and Yamagata lineages differ not only in the tree topology, but also in the way different genes are spread geographically. For the Victoria lineage, different geographical histories in the basal lineages of the tree correspond to different movements among East Asian and Australasian regions, whilst in more recent years viruses are reconstructed to spread in different ways between Asia and Europe, with the trunk more concentrated in Asia for NA and HA and in Europe for the other segments. For the Yamagata lineage, whilst the clade 2 (i.e. B/Massachusetts/02/2012 clade) shows little evidence of incompatibility between geographical histories, the spatial evolution of different genomic segments along the trunk of clade 3 (i.e. B/Wisconsin/1/2010 clade) clearly follows different routes of epidemics across Europe, America and East Asia/Australia.

Also for the Influenza B case, the phylogeographic reconstruction from all genes together is not a reliable description of the evolutionary histories of individual genes. For both lineages, the overall phylogeny and phylogeography resembles mostly the one of PB2 and it is strongly incompatible with the geographic history of the NS1, hemagglutinin (in the Victoria lineage) and of the neuraminidase (in the Yamagata one).

4 Discussion

We have here presented two measures of phylogeographic concordance or incompatibility, $I_{MRC A}$ and I_{MASPG} . The first is derived from comparison of the geographical status of all the MRCAs of pairs of leaves. The second is derived from maximum agreement sub-phylogeographies. Both of these measures can be used to define metrics on the space of phylogeographies.

These incompatibility measures can be applied to any phylogeography, and more generally to any phylogeny where each tip has been assigned a geographical discrete location or any other continuous/discrete “trait” independent of the sequence. A typical example of such “traits” could be the host species for pathogens, or environmental variables for species living in a varied range of environments. The natural assignment method is parsimony as pioneered by Maddison, or Bayesian approaches such as BEAST [Suchard et al., 2018], but our measures are independent of the reconstruction method used and on the specific trait. Our approach can also compare multiple trees inferred from different origins, e.g. phylogenetic vs phylolinguistic trees, provided that the tips have the same geographical labels (representing e.g. different populations).

Figure 4: Comparisons of various incompatibility measures between phylogeographies reconstructed from segments of the influenza B genome for both the Victoria and Yamagata lineages. On the the left, I_{MRCA} (upper triangle) is compared to I_{path} (lower triangle) for the Victoria (upper plot) and Yamagata (lower plot) strains. On the right, I_{MASPG} (upper triangle) is compared to I_{MAST} (lower triangle) for the Victoria (upper plot) and Yamagata (lower plot) strains.

Figure 5: Illustration of the incompatibilities along the MCC tree for the Influenza B Victoria and Yamagata lineages epidemics. The width of each branch represents the total amount of phylogeographic MRCA incompatibilities between all trees built from individual genomic regions and the whole-genome tree. Pie charts represent the distribution of MRCA incompatibilities per genomic region (left) and per location (right) on the corresponding nodes. Branch colors on the right represent the phylogeography reconstructed from the joint sequences alignment of all the genes investigated (~ 8.6 kb).

Figure 6: Illustrations of the correlations between the measures investigated in this paper. (a) Correlations between incompatibility measures. The upper triangle shows the Pearson correlation, and the lower triangle shows the Spearman correlation. (b-d) Scatter plots illustrating the correlations between incompatibility measures for the three viruses whose phylogeographies were investigated in this paper.

The two measures we propose are correlated but by no means identical. In fact, they are sensitive to different aspects of phylogeographic incompatibilities. We present these two particular measures in detail due to their naturalness and ease of computational implementation. There are however alternative possibilities, including extensions of tree distances other than the path and MAST distances. One promising such generalisation would be to include geographical information in Subtree-Prune-and-Regraft edit distances [De Oliveira Martins et al., 2014, Allen and Steel, 2001]. However, such a generalisation requires many arbitrary decisions in how internal node states are treated, and its values are challenging to compute. In practice, $I_{MRC A}$ is both simple and sensitive enough to be the primary choice for phylogeographic incompatibilities.

There are obvious extensions to the present methods, such as allowing incomplete geographical metadata for the samples, constraints on the geographical assignments of internal nodes, or weighting geographical incompatibilities by their location on the phylogeny.

One important limitation of our approach is that we consider distinct phylogeographies, instead of a phylogeographic network or a geographically labelled ARG. This simplification is rooted in the current approaches to phylogeographic inference. In fact, both phylogenetic trees and phylogeographies are usually inferred from sequence data and geographical metadata; most often, inference is performed separately for each locus, resulting in posterior distributions with unknown correlations between loci. To account for uncertainties in the inference, we defined a measure of incompatibility over both posterior distributions of phylogeographies which is conservative with respect to potential correlations.

Eventually, a full stochastic model of sequence evolution, birth-death, transmission, recombination, and geography should be developed for joint inference at multiple loci. This is both conceptually and computationally hard, but in no way impossible. As a example, the model of [Dialdestoro et al., 2016] has these components except transmission and geography. Including transmission and geography is technically doable but the details of the implementation are challenging and it is not clear how it would scale to large data sets. Even analysing only intra-patient sequences, the model by [Dialdestoro et al., 2016] pushed the computational limits. Recent developments enable the inference of ARGs as collection of related trees for large genomes and large numbers of sequences [Kelleher et al., 2018], but it is unclear how the model can be extended to accommodate uncertainties and trait reconstruction. However, even a complete statistical model will not make the present work redundant, as any approach to statistical phylogeography could be faced with the need to define some kind of incompatibility between phylogeographies. In its simplicity, the approach outlined here is the first one to enable a systematic comparison and to provide a definition of distance between reconstructed phylogeographic histories.

5 Materials and Methods

5.1 Coalescent simulations

To investigate the dependence of $I_{MRC A}$ on evolutionary parameters, we implemented a simulator of genealogies with migration and recombination. This simulator generates ancestral recombination graphs based on structured coalescent simulations in MASTER 6.1.1 [Vaughan and Drummond, 2013], implemented in BEAST2 2.6.1 [Bouckaert et al., 2019]. The simulation uses an island model of population structure with three populations of 20 individuals each and different patterns of migration between all populations (as illustrated in Figure 2). The simulations output geographically labelled ancestral recombination graphs (ARGs) for pairs of loci. We converted the ARGs into pairs of phylogeographies, one for each locus, and calculated their pairwise $I_{MRC A}$ for each set of parameters.

The mean $I_{MRC A}$ for each parameter combination is shown in Figure 2.

5.2 Phylogeographic reconstruction

In order to test our method to real evolutionary scenarios, we compiled two different data sets of genome sequences from influenza B and foot-and-mouth disease (FMD) viral infections, which affect both human and livestock species. Reassortment events and recombination patterns have been described for both the viral infections along with their dynamics of geographical transitions [Carrillo et al., 2005, Jackson et al., 2007, Dudas et al., 2014, Bedford et al., 2015]. For influenza B virus we selected 122 and 120 unique genome sequences of, respectively, the Victoria (Vic) and Yamagata (Yam) lineages [Bedford et al., 2015, Langat et al., 2017], characterising five out of the eight gene segments and encoding: the polymerase basic subunits 1 and 2 (PB1 and PB2), the hemagglutinin (HA) and neuraminidase (NA) glycoproteins, and the non-structural protein 1 (NS1) [Bouvier and Palese, 2008]. For FMD we selected 74 whole-genome sequences (WGSs) characterising the O/ME-SA/Ind-2001 lineage [Bachanek-Bankowska et al., 2018], extracting separate alignments for each of the four structural (VP1 to VP4) and six non-structural (2A to 2C, and 3A to 3D) proteins, along with the leader polypeptide (Lpro) and the 5' untranslated region (5'UTR) alignments [Jackson et al., 2003].

Distinct time-resolved phylogeographic trees were inferred for each of the influenza B genes and FMD virus (FMDV) proteins using BEAST 1.10.4 [Suchard et al., 2018]. Virus evolution was modelled by defining: the substitution process by the HKY [Hasegawa et al., 1985] and GTR+Γ4 [Tavaré, 1986] models respectively for influenza B and FMD; a log-normal relaxed molecular clock across branches [Drummond et al., 2006]; a flexible Bayesian Skyride coalescent prior on trees [Minin et al., 2008]. Phylogeographic patterns of lineage movement across geographic locations (defining discrete traits as countries for FMD, and as United Nations geoscheme regions for influenza B) were reconstructed using a discrete-state continuous-time Markov chain (CTMC) process, and assuming a non-reversible transition model with a Bayesian stochastic search variable selection (BSSVS) procedure [Lemey et al., 2009]. All priors were left at their default values. The Markov chain Monte Carlo (MCMC) chains were run for 200 million iterations, sampling trees every 20000 states. Inference were based on the resulting 9000 trees obtained after discarding the initial 10% of the chain as burn-in. Phylogeographic incompatibilities were assessed in the posterior phylogeographic space of a subset of 100 trees uniformly sampled from the posterior set of reconstructed phylogenies, using an approximation of the Wasserstein metric as described above.

6 Acknowledgements

We thank Nick Knowles and colleagues at the WRLFMD for useful discussions. The Pirbright Institute receives grant-aided support from the Biotechnology and Biological Sciences Research Council (BBSRC) of the United Kingdom (projects BBS/E/I/00007035, BBS/E/I/00007036 and BBS/E/I/00007039). This work was supported by funding from the BBSRC, grant number BBS/M011224/1. ADN acknowledges support from the Department for Environment, Food and Rural Affairs (Defra), United Kingdom, through funding from the SE2943 and SE2944 research grants.

References

- [Alexandersen et al., 2003] Alexandersen, S., Zhang, Z., Donaldson, A., and Garland, A. (2003). The pathogenesis and diagnosis of foot-and-mouth disease. *Journal of Comparative Pathology*, 129(1):1 – 36.
- [Allen and Steel, 2001] Allen, B. L. and Steel, M. (2001). Subtree transfer operations and their induced metrics on evolutionary trees. *Annals of combinatorics*, 5(1):1–15.
- [Arzt et al., 2011a] Arzt, J., Baxt, B., Grubman, M. J., Jackson, T., Juleff, N., Rhyan, J., Rieder, E., Waters, R., and Rodriguez, L. L. (2011a). The Pathogenesis of Foot-and-Mouth Disease II: Viral Pathways in Swine, Small Ruminants, and Wildlife; Myotropism, Chronic Syndromes, and Molecular Virus-Host Interactions. *Transboundary and Emerging Diseases*, 58(4):305–326.
- [Arzt et al., 2011b] Arzt, J., Juleff, N., Zhang, Z., and Rodriguez, L. L. (2011b). The Pathogenesis of Foot-and-Mouth Disease I: Viral Pathways in Cattle. *Transboundary and Emerging Diseases*, 58(4):291–304.
- [Avice, 2000] Avice, J. C. (2000). *Phylogeography : the history and formation of species*. Harvard University Press.
- [Bachanek-Bankowska et al., 2018] Bachanek-Bankowska, K., Di Nardo, A., Wadsworth, J., Mioulet, V., Pezzoni, G., Grazioli, S., Brocchi, E., Kafle, S. C., Hettiarachchi, R., Kumarawadu, P. L., Eldaghayes, I. M., Dayhum, A. S., Meenowa, D., Sghaier, S., Madani, H., Abouchoaib, N., Hoang, B. H., Vu, P. P., Dukpa, K., Gurung, R. B., Tenzin, S., Wernery, U., Panthumart, A., Seeyo, K. B., Linchongsubongkoch, W., Relmy, A., Bakkali-Kassimi, L., Scherbakov, A., King, D. P., and Knowles, N. J. (2018). Reconstructing the evolutionary history of pandemic foot-and-mouth disease viruses: the impact of recombination within the emerging o/me-sa/ind-2001 lineage. *Scientific Reports*, 8(1):14693.
- [Baele et al., 2018] Baele, G., Dellicour, S., Suchard, M., Lemey, P., and Vrancken, B. (2018). Recent advances in computational phylodynamics. *Current Opinion in Virology*, 31:24–32.
- [Barthélemy, 2011] Barthélemy, M. (2011). Spatial Networks. *Physics Reports*, (499):1–101.
- [Bedford et al., 2015] Bedford, T., Riley, S., Barr, I. G., Broor, S., Chadha, M., Cox, N. J., Daniels, R. S., Gunasekaran, C. P., Hurt, A. C., Kelso, A., Klimov, A., Lewis, N. S., Li, X., McCauley, J. W., Odagiri, T., Potdar, V., Rambaut, A., Shu, Y., Skepner, E., Smith, D. J., Suchard, M. A., Tashiro, M., Wang, D., Xu, X., Lemey, P., and Russell, C. A. (2015). Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature*, 523:217.
- [Bouckaert, 2016] Bouckaert, R. (2016). Phylogeography by diffusion on a sphere: whole world phylogeography. *PeerJ*, 4:e2406.
- [Bouckaert et al., 2019] Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., du Plessis, L., Poppinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M. A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T., and Drummond, A. J. (2019). Beast 2.5: An advanced software platform for bayesian evolutionary analysis. *PLOS Computational Biology*, 15(4):1–28.
- [Bouvier and Palese, 2008] Bouvier, N. M. and Palese, P. (2008). The biology of influenza viruses. *Vaccine*, 26:D49 – D53. Influenza Vaccines: Research, Development and Public Health Challenges.

- [Brito et al., 2017] Brito, B. P., Rodriguez, L. L., Hammond, J. M., Pinto, J., and Perez, A. M. (2017). Review of the global distribution of foot-and-mouth disease virus from 2007 to 2014. *Transboundary and Emerging Diseases*, 64(2):316–332.
- [Carrillo et al., 2005] Carrillo, C., Tulman, E. R., Delhon, G., Lu, Z., Carreno, A., Vagnozzi, A., Kutish, G. F., and Rock, D. L. (2005). Comparative genomics of foot-and-mouth disease virus. *Journal of Virology*, 79(10):6487–6504.
- [Chung et al., 2013] Chung, Y., Perna, N. T., and Ane, C. (2013). Computing the Joint Distribution of Tree Shape and Tree Distance for Gene Tree Inference and Recombination Detection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 10(5):1–1.
- [Day, 1984] Day, W. H. (1984). Properties of levenshtein metrics on sequences. *Bulletin of Mathematical Biology*, 46(2):327–332.
- [De Maio et al., 2015] De Maio, N., Wu, C.-H., O’Reilly, K. M., and Wilson, D. (2015). New routes to phylogeography: A bayesian structured coalescent approximation. *PLOS Genetics*, 11(8):1–22.
- [De Oliveira Martins et al., 2014] De Oliveira Martins, L., Mallo, D., and Posada, D. (2014). A bayesian supertree model for genome-wide species tree reconstruction. *Systematic biology*, 65(3):397–416.
- [Dellicour et al., 2016] Dellicour, S., Rose, R., Faria, N. R., Lemey, P., and Pybus, O. G. (2016). SERAPHIM: studying environmental rasters and phylogenetically informed movements. *Bioinformatics*, 32(20):3204–3206.
- [Di Nardo et al., 2011] Di Nardo, A., Knowles, N. J., and Paton, D. P. (2011). Combining livestock trade patterns with phylogenetics to help understand the spread of foot and mouth disease in sub-saharan africa, the middle east and southeast asia. *Scientific and Technical Review*, 30(1):63–85.
- [Dialdestoro et al., 2016] Dialdestoro, K., Sibbesen, J. A., Maretty, L., Raghvani, J., Gall, A., Kellam, P., Pybus, O. G., Hein, J., and Jenkins, P. A. (2016). Coalescent Inference Using Serially Sampled, High-Throughput Sequencing Data from Intrahost HIV Infection. *Genetics*, 202(4):1449–72.
- [Drummond et al., 2006] Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLOS Biology*, 4(5).
- [Dudas et al., 2014] Dudas, G., Bedford, T., Lycett, S., and Rambaut, A. (2014). Reassortment between Influenza B Lineages and the Emergence of a Coadapted PB1–PB2–HA Gene Complex. *Molecular Biology and Evolution*, 32(1):162–172.
- [Dudas et al., 2017] Dudas, G., Carvalho, L. M., Bedford, T., Tatem, A. J., Baele, G., Faria, N. R., Park, D. J., Ladner, J. T., Arias, A., Asogun, D., Bielejec, F., Caddy, S. L., Cotten, M., D’Ambrozio, J., Dellicour, S., Di Caro, A., DiClaro, J. W., Duraffour, S., Elmore, M. J., Fakoli, L. S., Faye, O., Gilbert, M. L., Gevao, S. M., Gire, S., Gladden-Young, A., Gnirke, A., Goba, A., Grant, D. S., Haagmans, B. L., Hiscox, J. A., Jah, U., Kugelman, J. R., Liu, D., Lu, J., Malboeuf, C. M., Mate, S., Matthews, D. A., Matranga, C. B., Meredith, L. W., Qu, J., Quick, J., Pas, S. D., Phan, M. V. T., Pollakis, G., Reusken, C. B., Sanchez-Lockhart, M., Schaffner, S. F., Schieffelin, J. S., Sealfon, R. S., Simon-Loriere, E., Smits, S. L., Stoecker, K., Thorne, L., Tobin, E. A., Vandi, M. A., Watson, S. J., West, K., Whitmer, S., Wiley, M. R., Winnicki, S. M., Wohl, S., Wölfel, R., Yozwiak, N. L., Andersen, K. G., Blyden, S. O., Bolay, F., Carroll, M. W., Dahn, B., Diallo, B., Formenty, P., Fraser, C., Gao, G. F., Garry, R. F., Goodfellow, I., Günther, S.,

- Happi, C. T., Holmes, E. C., Kargbo, B., Keita, S., Kellam, P., Koopmans, M. P. G., Kuhn, J. H., Loman, N. J., Magassouba, N., Naidoo, D., Nichol, S. T., Nyenswah, T., Palacios, G., Pybus, O. G., Sabeti, P. C., Sall, A., Ströher, U., Wurie, I., Suchard, M. A., Lemey, P., and Rambaut, A. (2017). Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature*, 544(7650):309–315.
- [Edwards et al., 2016] Edwards, S. V., Potter, S., Schmitt, C. J., Bragg, J. G., and Moritz, C. (2016). Reticulation, divergence, and the phylogeography-phylogenetics continuum. *Proceedings of the National Academy of Sciences of the United States of America*, 113(29):8025–32.
- [Faria et al., 2018] Faria, N. R., Kraemer, M. U. G., Hill, S. C., Goes de Jesus, J., Aguiar, R. S., Iani, F. C. M., Xavier, J., Quick, J., du Plessis, L., Dellicour, S., Thézè, J., Carvalho, R. D. O., Baele, G., Wu, C.-H., Silveira, P. P., Arruda, M. B., Pereira, M. A., Pereira, G. C., Lourenço, J., Obolski, U., Abade, L., Vasylyeva, T. I., Giovanetti, M., Yi, D., Weiss, D. J., Wint, G. R. W., Shearer, F. M., Funk, S., Nikolay, B., Fonseca, V., Adelino, T. E. R., Oliveira, M. A. A., Silva, M. V. F., Sacchetto, L., Figueiredo, P. O., Rezende, I. M., Mello, E. M., Said, R. F. C., Santos, D. A., Ferraz, M. L., Brito, M. G., Santana, L. F., Menezes, M. T., Brindeiro, R. M., Tanuri, A., dos Santos, F. C. P., Cunha, M. S., Nogueira, J. S., Rocco, I. M., da Costa, A. C., Komninakis, S. C. V., Azevedo, V., Chieppe, A. O., Araujo, E. S. M., Mendonça, M. C. L., dos Santos, C. C., dos Santos, C. D., Mares-Guia, A. M., Nogueira, R. M. R., Sequeira, P. C., Abreu, R. G., Garcia, M. H. O., Abreu, A. L., Okumoto, O., Kroon, E. G., de Albuquerque, C. F. C., Lewandowski, K., Pullan, S. T., Carroll, M., de Oliveira, T., Sabino, E. C., Souza, R. P., Suchard, M. A., Lemey, P., Trindade, G. S., Drumond, B. P., Filippis, A. M. B., Loman, N. J., Cauchemez, S., Alcantara, L. C. J., and Pybus, O. G. (2018). Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science*, 361(6405):894–899.
- [Faria et al., 2011] Faria, N. R., Suchard, M. A., Rambaut, A., and Lemey, P. (2011). Toward a quantitative understanding of viral phylogeography. *Current Opinion in Virology*, 1(5):423–429.
- [Gernhard, 2008] Gernhard, T. (2008). The conditioned reconstructed process. *Journal of Theoretical Biology*, 253(4):769–778.
- [Gill et al., 2017] Gill, M. S., Tung Ho, L. S., Baele, G., Lemey, P., and Suchard, M. A. (2017). A Relaxed Directional Random Walk Model for Phylogenetic Trait Evolution. *Systematic Biology*, 66(3):299–319.
- [Goddard et al., 1994] Goddard, W., Kubicka, E., Kubicki, G., and McMorris, F. (1994). The agreement metric for labeled binary trees. *Mathematical Biosciences*, 123(2):215–226.
- [Grubaugh et al., 2019] Grubaugh, N., Ladner, J., Lemey, P., Pybus, O., Rambaut, A., Holmes, E., and K.G., A. (2019). Tracking virus outbreaks in the twenty-first century. *Nature Microbiology*, 4(1):10–19.
- [Hasegawa et al., 1985] Hasegawa, M., Kishino, H., and Yano, T.-a. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial dna. *Journal of Molecular Evolution*, 22(2):160–174.
- [Hein et al., 1996] Hein, J., Jiang, T., Wang, L., and Zhang, K. (1996). On the complexity of comparing evolutionary trees. *Discrete Applied Mathematics*, 71(1-3):153–169.

- [Jackson et al., 2007] Jackson, A. L., O'Neill, H., Maree, F., Blignaut, B., Carrillo, C., Rodriguez, L., and Haydon, D. T. (2007). Mosaic structure of foot-and-mouth disease virus genomes. *Journal of General Virology*, 88(2):487–492.
- [Jackson et al., 2003] Jackson, T., King, A. M., Stuart, D. I., and Fry, E. (2003). Structure and receptor binding. *Virus Research*, 91(1):33 – 46. Foot-and-Mouth Disease.
- [Jain et al., 2018] Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., Snutch, T. P., Tee, L., Paten, B., Phillippy, A. M., Simpson, J. T., Loman, N. J., and Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*.
- [Kelleher et al., 2018] Kelleher, J., Wong, Y., Albers, P. K., Wohns, A. W., and McVean, G. (2018). Inferring the ancestry of everyone. *bioRxiv*.
- [Knight-Jones and Rushton, 2013] Knight-Jones, T. J. D. and Rushton, J. (2013). The economic impacts of foot and mouth disease - what are they, how big are they and where do they occur? *Preventive veterinary medicine*, 112(3-4):161–73.
- [Knowles, 2009] Knowles, L. L. (2009). Statistical phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, 40:593–612.
- [Kuhner and Felsenstein, 1994] Kuhner, M. K. and Felsenstein, J. (1994). A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular biology and evolution*, 11(3):459–468.
- [Kühnert et al., 2016] Kühnert, D., Stadler, T., Vaughan, T. G., and Drummond, A. J. (2016). Phylodynamics with Migration: A Computational Framework to Quantify Population Structure from Genomic Data. *Molecular Biology and Evolution*, 33(8):2102–2116.
- [Langat et al., 2017] Langat, P., Raghvani, J., Dudas, G., Bowden, T. A., Edwards, S., Gall, A., Bedford, T., Rambaut, A., Daniels, R. S., Russell, C. A., Pybus, O. G., McCauley, J., Kellam, P., and Watson, S. J. (2017). Genome-wide evolutionary dynamics of influenza b viruses on a global scale. *PLOS Pathogens*, 13(12):1–26.
- [Lanier and Knowles, 2012] Lanier, H. C. and Knowles, L. L. (2012). Is Recombination a Problem for Species-Tree Analyses? *Systematic Biology*, 61(4):691–701.
- [Lemey et al., 2014] Lemey, P., Rambaut, A., Bedford, T., Faria, N., Bielejec, F., Baele, G., Russell, C. A., Smith, D. J., Pybus, O. G., Brockmann, D., and Suchard, M. A. (2014). Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza h3n2. *PLOS Pathogens*, 10(2):1–10.
- [Lemey et al., 2009] Lemey, P., Rambaut, A., Drummond, A. J., and Suchard, M. A. (2009). Bayesian phylogeography finds its roots. *PLOS Computational Biology*, 5(9):1–16.
- [Lemey et al., 2010] Lemey, P., Rambaut, A., Welch, J. J., and Suchard, M. A. (2010). Phylogeography Takes a Relaxed Random Walk in Continuous Space and Time. *Molecular Biology and Evolution*, 27(8):1877–1885.

- [Minin et al., 2008] Minin, V. N., Bloomquist, E. W., and Suchard, M. A. (2008). Smooth Skyride through a Rough Skyline: Bayesian Coalescent-Based Inference of Population Dynamics. *Molecular Biology and Evolution*, 25(7):1459–1471.
- [Müller et al., 2017] Müller, N. F., Rasmussen, D. A., and Stadler, T. (2017). The Structured Coalescent and Its Approximations. *Molecular Biology and Evolution*, 34(11):2970–2981.
- [Paradis and Schliep, 2018] Paradis, E. and Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in r. *Bioinformatics*, 35(3):526–528.
- [Quick et al., 2016] Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J. H. J., Becker-Ziaja, B., Boettcher, J. P., Cabeza-Cabrero, M., Camino-Sánchez, Á., Carter, L. L., Doerrbecker, J., Enkirch, T., Dorival, I. G., Hetzelt, N., Hinzmann, J., Holm, T., Kafetzopoulou, L. E., Koropogui, M., Kosgey, A., Kuisma, E., Logue, C. H., Mazzarelli, A., Meisel, S., Mertens, M., Michel, J., Ngabo, D., Nitzsche, K., Pallasch, E., Patrono, L. V., Portmann, J., Repits, J. G., Rickett, N. Y., Sachse, A., Singethan, K., Vitoriano, I., Yemanaberhan, R. L., Zekeng, E. G., Racine, T., Bello, A., Sall, A. A., Faye, O., Faye, O., Magassouba, N., Williams, C. V., Amburgey, V., Winona, L., Davis, E., Gerlach, J., Washington, F., Monteil, V., Jourdain, M., Bererd, M., Camara, A., Somlare, H., Camara, A., Gerard, M., Bado, G., Baillet, B., Delaune, D., Nebie, K. Y., Diarra, A., Savane, Y., Pallawo, R. B., Gutierrez, G. J., Milhano, N., Roger, I., Williams, C. J., Yattara, F., Lewandowski, K., Taylor, J., Rachwal, P., J. Turner, D., Pollakis, G., Hiscox, J. A., Matthews, D. A., Shea, M. K. O., Johnston, A. M., Wilson, D., Hutley, E., Smit, E., Di Caro, A., Wölfel, R., Stoecker, K., Fleischmann, E., Gabriel, M., Weller, S. A., Koivogui, L., Diallo, B., Keita, S., Rambaut, A., Formenty, P., Günther, S., and Carroll, M. W. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530(7589):228–232.
- [Rasmussen, 2016] Rasmussen, N. S. (2016). *Algorithms for Computing Maximum Agreement Subtrees*. PhD thesis, Aarhus Universitet, Datalogisk Institut.
- [Robinson and Foulds, 1981] Robinson, D. F. and Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical biosciences*, 53(1-2):131–147.
- [Schliep, 2010] Schliep, K. P. (2010). phangorn: phylogenetic analysis in r. *Bioinformatics*, 27(4):592–593.
- [Semple and Steel, 2003] Semple, C. and Steel, M. A. (2003). *Phylogenetics*. Oxford University Press.
- [Slatkin and Maddison, 1989] Slatkin, M. and Maddison, W. P. (1989). A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics*, 123(3):603–613.
- [Steel, 2016] Steel, M. (2016). Pulling trees apart and putting trees together. In *Phylogeny: Discrete and Random Processes in Evolution*, chapter 4, pages 63–86.
- [Steel and Warnow, 1993] Steel, M. and Warnow, T. (1993). Kaikoura tree theorems: Computing the maximum agreement subtree. *Information Processing Letters*, 48(2):77–82.
- [Steel and Penny, 1993] Steel, M. A. and Penny, D. (1993). Distributions of tree comparison metrics—some new results. *Systematic biology*, 42(2):126–141.

- [Suchard et al., 2018] Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, 4(1).
- [Tavaré, 1986] Tavaré, S. (1986). Some probabilistic and statistical problems on the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, 17:57–86.
- [Vaughan and Drummond, 2013] Vaughan, T. G. and Drummond, A. J. (2013). A Stochastic Simulator of Birth-Death Master Equations with Application to Phylodynamics. *Molecular Biology and Evolution*, 30(6):1480–1493.

A Proofs that incompatibility measures are distances on $\mathcal{PG}$

I_{MASPG} is a metric on $\mathcal{PG}_n$. It is non-negative, and satisfies identity of indiscernibles and symmetry. The proof of the triangle inequality for I_{MASPG} is as follows.

Let us define $S_{i,j} = MASPG(T_i, T_j)$. We require that, for three phylogeographies T_1 , T_2 , and T_3 ,

$$I_{MASPG}(T_1, T_3) \leq I_{MASPG}(T_1, T_2) + I_{MASPG}(T_2, T_3). \quad (3)$$

This is equivalent to requiring that

$$n - |S_{1,3}| \leq n - |S_{1,2}| + n - |S_{2,3}|, \quad (4)$$

i.e.

$$|S_{1,3}| \geq |S_{1,2}| + |S_{2,3}| - n. \quad (5)$$

Let us note that any agreement sub-phylogeography between $S_{1,2}$ and $S_{2,3}$ must be a sub-phylogeography of all three trees, and so is a sub-phylogeography of $S_{1,3}$. Let U be the largest geographically agreeing sub-phylogeography for all pairs of trees in $\{T_1, T_2, T_3\}$. Now for each pair of trees the MASPG is either identical to U or is some larger agreement sub-phylogeography, so

$$|S_{i,j}| \geq |U| \quad (6)$$

for all $i, j \in \{1, 2, 3\}$. Now, if n is the number of leaves, then $S_{1,2}$ and $S_{2,3}$ cannot share fewer than $|S_{1,2}| + |S_{2,3}| - n$ leaves, since the union of the sets of leaves on the two sub-phylogeographies cannot be larger than n . Therefore

$$|U| \geq |S_{1,2}| + |S_{2,3}| - n. \quad (7)$$

From (6) we get $|S_{1,3}| \geq |U|$, and substituting this into (7) yields (5), thus proving that the triangle inequality holds and that I_{MASPG} is a metric.

B MASPG algorithm

A phylogeography, P_a , is defined by a tree T_a and a mapping f_G from nodes of T_a to some geographic space $\mathcal{G}$. The algorithm for the MASPG of two phylogeographies is based on a modification of the Goddard algorithm for the MAST [Goddard et al., 1994]. The modification is based on this lemma, following [Rasmussen, 2016].

Lemma 1. *Let P_a be a phylogeography with tree T_a rooted at node a with the children b and c , and let P_w have a tree T_w rooted at w with children x and y . Let $f_G(a) = A$, and $f_G(w) = W$. We now define $\#(P_a, P_w)$ as the size of the maximum agreement sub-phylogeography (MASPG) of P_a and P_w . If $A = W$, $\#(P_a, P_w)$ is given by the maximum of the following six numbers: $\#(P_b, P_x) + \#(P_c, P_y)$, $\#(P_b, P_y) + \#(P_c, P_x)$, $\#(P_a, P_x)$, $\#(P_a, P_y)$, $\#(P_b, P_w)$, $\#(P_c, P_w)$. If $A \neq W$, $\#(P_a, P_w)$ is given by the maximum of $\#(P_a, P_x)$, $\#(P_a, P_y)$, $\#(P_b, P_w)$, and $\#(P_c, P_w)$; i.e. the last four numbers in the list above.*

Proof. This is a proof by case over the possible sub-phylogeographies contributing to the agreement sub-phylogeography A_{aw} for P_a and P_w .

Case 1: Only P_b of P_a contributes to A_{aw} . If this is the case, then the agreement sub-phylogeography must be an agreement sub-phylogeography of P_b and P_w , and its size is equal to $\#(P_b, P_w)$.

Case 2: Only P_c of P_a contributes to A_{aw} . If this is the case, then the agreement sub-phylogeography must be an agreement sub-phylogeography of P_c and P_w , and its size is equal to $\#(P_c, P_w)$.

Figure B1: An illustration of the central claim in Case 5 of the proof by case of Lemma 1. The figures show that a child of r cannot contain a leaf b' from P_b and a leaf c' from P_c at the same time, since the sub-phylogeographies induced by these leaves and a leaf from a child of r on A_r and P_a must differ topologically.

Case 3: Only P_x of P_w contributes to A_{aw} . If this is the case, then the agreement sub-phylogeography must be an agreement sub-phylogeography of P_x and P_a , and its size is equal to $\#(P_a, P_x)$.

Case 4: Only P_y of P_w contributes to A_{aw} . If this is the case, then the agreement sub-phylogeography must be an agreement sub-phylogeography of P_y and P_a , and its size is equal to $\#(P_a, P_y)$.

Case 5: All children b , c , x , and y contribute to the MASPG. It follows that a and w are part of the MASPG, and so $A = W$. Consider the ASPG A_r , rooted in r with children s and t . We claim that a child of r cannot contain a leaf b' from P_b and a leaf c' from P_c at the same time. If this is the case, we only need to find the MASPG size of child pair permutations across P_a and P_w : $\max\{\#(P_b, P_x) + \#(P_c, P_y), \#(P_b, P_y) + \#(P_c, P_x)\}$. The proof of this claim by contradiction is presented below.

Consider the ASPG A_r , rooted in r with children s and t . Suppose that A_s contains a leaf b' from P_b and a leaf c' from P_c . Let f be any leaf from A_t . Assume that f is from P_b (the proof for the other case is equivalent). Then the sub-phylogeography induced by b' , f , and c' must differ for A_r and P_a , as illustrated in Figure 1. $\square$

The algorithm to calculate I_{MASPG} can then be constructed by iterating the process of finding the maximum of the sizes of various agreement sub-phylogeographies, working from the leaves upwards.

Figure B2: MCC phylogeographies of the O/ME-SA/Ind-2001 epidemic reconstructed for the whole-genome sequence and each of the FMDV protein.

Figure B3: MCC phylogeographies of the Influenza B Victoria epidemics reconstructed for the whole-genome sequence and each of the segment.

Figure B4: MCC phylogeographies of the Influenza B Yamagata epidemics reconstructed for the whole-genome sequence and each of the segment.

Figure B5: Effect of evolutionary parameters on the variance of phylogeographic incompatibilities, using 100 simulations at each calculated pair of parameter values. The plots show smoothed surfaces of the variance each phylogeographic incompatibility measure, labelled with the corresponding network structure. The network structures are illustrated with locations in which the population can originate coloured yellow.

Figure B6: Phylogenetic incompatibilities across genomic regions in FMDV

Figure B7: Phylogenetic incompatibilities across genomic regions in Influenza B Victoria

Figure B8: Phylogenetic incompatibilities across genomic regions in Influenza B Yamagata