Measuring the temporal structure in serially sampled phylogenies

Rebecca R. Gray[1], Oliver G. Pybus[2] and Marco Salemi[1]*

[1]*Department of Pathology, Immunology and Laboratory Medicine and Emerging Pathogens Institute, University of Florida, PO Box 100009, 2055 Mowry Rd, Gainesville, FL, USA, 32611; and* [2]*Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, United Kingdom*

Summary

1. Nucleotide sequences sampled at different times (serially sampled sequences) allow researchers to study the rate of evolutionary change and the demographic history of populations. Some phylogenies inferred from serially sampled sequences are described as having strong 'temporal clustering', such that sequences from the same sampling time tend to cluster together and to be the direct ancestors of sequences from the following sampling time. The degree to which phylogenies exhibit these properties is thought to reflect interesting biological processes, such as positive selection or deviation from the molecular clock hypothesis.

2. Here, we introduce the temporal clustering (TC) statistic, which is the first quantitative measure of the degree of topological 'temporal clustering' in a serially sampled phylogeny. The TC statistic represents the expected deviation of an observed phylogeny from the null hypothesis of no temporal clustering, as a proportion of the range of possible values, and can therefore be compared among phylogenies of different sizes.

3. We apply the TC statistic to a range of serially sampled sequence data sets, which represent both rapidly evolving viruses and ancient mitochondrial DNA. In addition, the TC statistic was calculated for phylogenies simulated under a neutral coalescent process.

4. Our results indicate significant TC in many empirical data sets. However, we also find that such clustering is exhibited by trees simulated under a neutral coalescent process; hence, the observation of significant 'temporal clustering' cannot unambiguously indicate the presence of strong positive selection in a population.

5. Quantifying topological structure in this manner will provide new insights into the evolution of measurably evolving populations.

Key-words: ancient DNA, coalescent, dated tips, HCV, HIV-1, molecular clocks

Introduction

Within the past decade, data sets of nucleotide sequences sampled from the same species at evolutionarily distinct points in time have increased in availability. Such data sets (termed heterochronous, serially sampled or measurably evolving) enable researchers to directly estimate the rate of evolution and to reconstruct the past changes in population size and structure (e.g. Drummond *et al.* 2003, 2005). There are two main sources of serially sampled gene sequences: (i) ancient DNA studies, which sample slowly evolving animal sequences over tens or hundreds of thousands of years (e.g. Ho & Gilbert 2010) and (ii) studies of micro-organisms (mostly RNA viruses, but also some DNA viruses and bacteria) that sample rapidly evolving populations over several months, years or decades (e.g. Pybus & Rambaut 2009).

Phylogenies estimated from serially sampled sequence data are shaped by a complex interaction of demographic and selective pressure. Those obtained from genes under strong and continual positive selection are reported to exhibit a 'ladder-like' shape, characterized by (i) phylogenetic asymmetry and (ii) a tendency for sequences sampled at similar times to cluster together (Grenfell *et al.* 2004). As a result, few lineages coexist at any one time point, but there is a rapid turnover of lineages through time. This behaviour is best seen in phylogenies of viral populations under strong selection by the host immune system, for example, HIV-1 viruses sampled throughout the course of chronic infection (e.g. Shankarappa *et al.* 1999) or phylogenies of human influenza A viruses sampled globally

*Correspondence author. E-mail: salemi@pathology.ufl.edu
Correspondence site: http://www.respond2articles.com/MEE/

over several years (e.g. Bush *et al.* 1999). These phylogenies are qualitatively described as having strong 'temporal clustering'.

In contrast, serially sampled phylogenies obtained from genes under weak or no selective pressure are thought to be shaped by other processes, such as demographic history. In these cases, sequences from different sampling times are often intermingled, and there is a greater degree of lineage coexistence through time (Grenfell *et al.* 2004). Examples include phylogenies of HIV-1 and hepatitis-C virus (HCV) subtypes sampled over the last 20–30 years from different hosts (e.g. Gray *et al.* 2009; Magiorkinis *et al.* 2009).

Although comparisons are often made between the two types of phylogenetic structure described earlier, there is at present no way of quantifying the observed degree of 'temporal clustering'. Indeed, even though it is commonly argued that temporal clustering (TC) arises from strong selective pressure (Grenfell *et al.* 2004; Lemey, Rambaut, & Pybus 2006), it is unknown whether such phylogenies could be produced simply by sampling a neutrally evolving population through time. It would be useful, therefore, to have a statistic for quantifying the degree of 'temporal clustering' observed in a given phylogeny reconstructed from serially sampled sequences. There are, of course, methods to test whether serially sampled sequences are evolving according to a molecular clock (Drummond, Pybus, & Rambaut 2003) – that is, whether the genetic distances among sequences correlate with time – and multiple analytic approaches are available that require the assumption of a molecular clock (e.g. Yang *et al.* 2007; Jombart *et al.* 2011). However, these methods do not address the topological issues described earlier, namely that sequences from the same sampling time are expected to cluster together and to be the direct ancestors of sequences from the following sampling time. In other words, although 'clock-like' evolution and 'temporal clustering' may often go hand-in-hand, it is possible for a clock-like phylogeny to have weak TC and for a temporally clustered tree to evolve in a non-clock-like manner.

Statistics used to describe phylogenetic characteristics can be grouped into two classes: those that use branch length information and those that use *only* the phylogenetic topology. Phylogenetic topology statistics are often highly robust because they can provide information independent of branch lengths and the molecular clock assumption. This class of statistics includes measures of tree imbalance/symmetry (Colless 1982; Shao & Sokal 1990; Fusco & Cronk 1995; Agapow & Purvis 2002; Purvis, Katzourakis, & Agapow 2002) and of trait-tip association (Slatkin & Maddison 1989; Wang *et al.* 2001; Salemi *et al.* 2005; Parker, Rambaut, & Pybus 2008). Phylogenies that show strong TC are, by necessity, likely to be asymmetric; however, the converse is not necessarily true. This is because other population processes (e.g. spatial structure) could create an asymmetric phylogeny.

A widely used statistic of tip-trait association is the parsimony score (PS), which is the number of state changes under the most parsimonious reconstruction (MPR) of ancestral states. The PS forms the basis of the Slatkin–Maddison test, which compares the PS of an observed phylogeny to a null distribution of that score. The null distribution represents the PS values expected when taxa states are allocated randomly to taxa (Slatkin & Maddison 1989). However, PS cannot be directly compared between phylogenies of different sizes and is inappropriate as a measure of TC because it allows temporally impossible state changes.

Here, we introduce the TC statistic, which is designed to quantify the degree of topological temporal structure in a phylogeny estimated from serially sampled sequences. Our aim here is to use the TC statistic as a descriptive statistic of phylogenetic topology; hence, it can be applied to any tree without regard to sampling or to population genetic assumptions. We provide a randomization-based statistical test for the presence of significant temporal structure using the TC statistic. We test the performance of the TC statistic using simulation and apply the test to a variety of simulated and empirical data sets, including alignments of rapidly evolving viruses (HIV-1 and HCV) and of bison ancient mitochondrial (mt)DNA sequences. Lastly, we discuss possible future applications of the TC statistic, particularly its potential as a test of neutral evolution.

Methods

TEMPORAL CLUSTERING STATISTIC

Suppose we have a phylogeny representing the shared ancestry of N taxa sampled at t different time points ($t \geq 3$). Let n_k be the number of sequences sampled at time point k. The index $k = 1, 2, ..., t$ reflects the order of empirical sampling times, such that $k = 1$ is the earliest time point, and $N = n_1 + n_2 + n_3 + \cdots n_t$. The sampling times should be in units appropriate to the organism and the data under investigation (e.g. millennia, years, months, etc.). If insufficient time separates two or more different samples, then they can be pooled into a single group. The most appropriate grouping strategy for any given data set will depend on (i) the durations between sampling times, (ii) the rate of sequence evolution and (iii) sequence length (see Drummond *et al.* (2003) for consideration of this issue).

Next, a one-character data matrix is generated by assigning to each taxon an integer value that equals the index (k) of the sampling time of that taxon. The character state of each internal node in the tree is then inferred by finding the MPR of the character using the Fitch algorithm (Fitch 1971). However, it is necessary to constrain changes in character state so that they are irreversible; i.e., a sequence from a later time point cannot be ancestral to a sequence sampled at an earlier time point. To impose this irreversibility criterion, we used an irreversible cost matrix ($\mathbf{W}$):

$$\mathbf{W} = \begin{pmatrix} 0 & 1 & 2 & \cdots & t-1 \\ \infty & 0 & 1 & \cdots & t-2 \\ \infty & \infty & 0 & \cdots & \cdots \\ \infty & \infty & \cdots & 0 & 1 \\ \infty & \infty & \cdots & \infty & 0 \end{pmatrix}$$

In practice, $\mathbf{W}$ assigns a weight (w_{ij}) to each change in the sample index from i to j (where $j \geq i$) in the tree so that $w_{ij} = j-i$. The final tree score (S) is the sum of all of the state changes across all branches. The minimum number of steps (Min) expected for a perfectly temporally structured tree (Fig. 1a) is
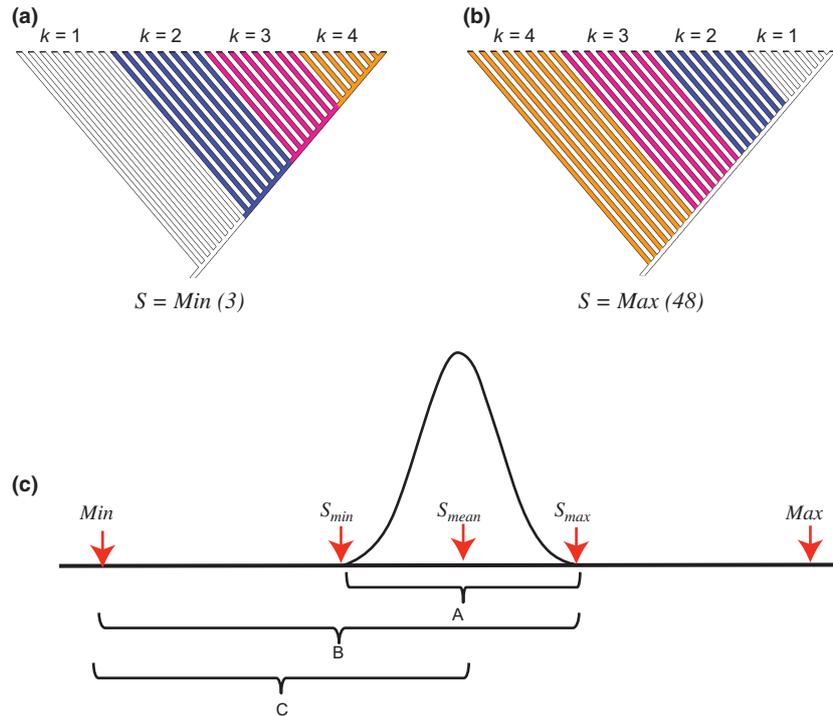
Fig. 1. (a) and (b) Schematic phylogenies representing the tree topologies with the lowest (a) and highest (b) possible tree score (S) values. Terminal branches are coloured according to the time point at which the hypothetical sample was taken (white = 1st, blue = 2nd, pink = 3rd, orange = 4th). Internal branches are coloured according to the most parsimonious reconstruction of the ancestral state, according to the irreversible matrix used to calculate the temporal clustering (TC) statistic (see text). (c) A number line representing the parameter space of the tree score (S) value. The null distribution is shown as a bell curve with maximum value (S_{max}), minimum value (S_{min}) and mean (S_{mean}) indicated. The absolute Minimum (Min) and Maximum (Max) tree scores are also indicated. The interval A denotes range of the null distribution. The interval B denotes the range of possible values (sample space). The interval C denotes the possible values for the numerator of the TC statistic.

$$\text{Min} = t - 1$$

On the other hand, the maximum number (Max) of steps expected for the least temporally clustered tree (Fig. 1b) is

$$\text{Max} = \sum_{j=1}^{t} n_j w_{1j}$$

The observed tree score (S_{obs}) can be compared to a null distribution obtained by randomizing the character states of the tips many times (while the phylogeny itself is held constant), and calculating the tree score S_r for each randomized replicate. A null distribution of S_r values is thereby obtained (let S_{max}, S_{min} and S_{mean} indicate the maximum, minimum and mean score of randomized replicates, respectively; see Fig. 1c). S_{obs} is considered significant if it is less than the critical value of the null distribution (S_{min}); i.e. if 1000 replicates are performed and $S_{obs} < S_{min}$, then S_{obs} is significant at the $P < 0.001$ level.

Although S_{obs} can be used to statistically test for the presence of temporal clustering, it cannot be compared directly between two different trees because the absolute value of S_{obs} will depend on the number of taxa (N) and number of time points (t). To resolve this problem, we introduce the TC statistic, which is calculated as follows:

$$\text{TC} = \frac{\ln(S_{mean}) - \ln(S_{obs})}{\ln(S_{max}) - \ln(\text{Min})}$$

The numerator represents the deviation from the null hypothesis (i.e. no temporal clustering), while the denominator represents the

range of possible values for the given N and t. If the numerator is negative, then TC is set to zero. Note that TC is defined between $0 \leq \text{TC} < 1$ (where TC = 0 indicates complete absence of temporal clustering). S_{max} is used rather than Max, as S values larger than those in the null distribution are highly unlikely in empirical or simulated data sets (data not shown). The TC statistic can thus be interpreted as 'the expected deviation of the observed topology from the null hypothesis of no temporal clustering, as a proportion of the range of possible values' and can therefore be compared among data sets.

PERFORMANCE OF THE TC STATISTIC

We first investigated how the sensitivity of the TC statistic might change as the number of taxa (N) and time points (t) changes. To quantify this, we performed two sets of simulations, and for each, we calculated the width of the null distribution as a proportion of total available parameter space:

$$\frac{S_{max} - S_{min}}{S_{max} - \text{Min}}$$

In the first set of simulations, the TC statistic was tested on phylogenies with different degrees of symmetry, as measured using the modified I' statistic for imbalance (Purvis, Katzourakis, & Agapow 2002) (Fig. 2a–d). Various numbers of taxa (N = 4, 8, 16, 32, 64, 128, 512) were used to simulate each tree topology, with the number of sampling times held constant (t = 4) and an equal number of taxa assigned to each time point. For each simulated data set, 1000 replicates were used to calculate the null distribution of the tree score.

In the second set of simulations, the number of sampling times (*t*) was varied. We investigated 64-taxa data sets with *t* = 4, 8 and 16, and 100-taxa data sets with *t* = 4, 5, 10, 20 and 25, respectively. Only the most asymmetric phylogenetic topology was used (i.e. that in Fig. 2d) to keep the topology constant and comparable between simulations. As before, for each simulated data set, 1000 replicates were used to calculate the null distribution of the tree score. All analyses were conducted using MacClade (Maddison & Maddison 2003).

EMPIRICAL DATA SETS

The TC statistic was calculated for fourteen empirical data sets. Nine data sets comprised intra-host HIV-1 sequences from chronically infected patients (Shankarappa *et al.* 1999). The time between first and last sampling time points for each patient ranged from 5·83 to 10·92 years (see Table 1). Sequences were assigned sampling index values according to their month of sampling.
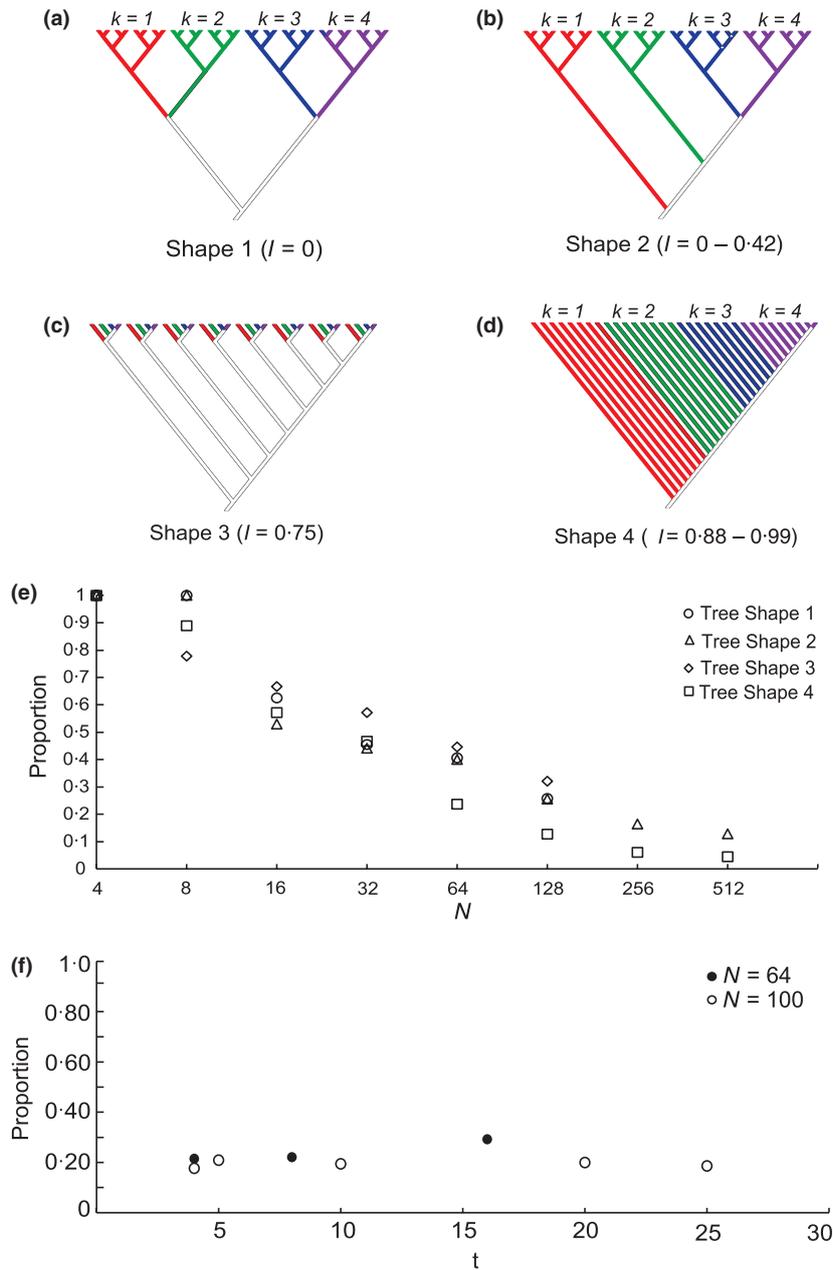
Fig. 2. The effect of varying tree shape, number of taxa and number of time points on the null distribution of the temporal clustering statistic. (a–d) Four hypothetical topologies, with varying degrees of asymmetry, each containing 32 taxa sampled at 4 time points. Values in parentheses represent the range of tree asymmetry scores (the I statistic) for trees with *N* = 4–128 taxa. Branches are coloured according to the indexed time of sampling, *k* = 1 (earliest) to 4 (latest). (e) Plot showing the proportion of tree score parameter space occupied by the null distribution (see text) as a function of the number of taxa (with the number of time points held constant). The proportion was calculated for each of the four topologies (denoted by symbols). (f) Plot showing the proportion of parameter space occupied by the null distribution as a function of the number of time points (for trees containing 64 or 100 taxa).

Four further data sets comprised HCV sequences sampled from different individuals, obtained from the Broad Institute data base (http://www.broadinstitute.org/annotation/viral/HCV/Home.html). Because sequences from recent years were over-represented in this data base, a maximum of five sequences from each year were randomly selected and retained. The four data sets represent two HCV gene regions (E1E2 and NS5a) for two different HCV genotypes (1a and 1b). The time between first and last sampling time points was 19 years for all four data sets. Sequences were assigned sampling index values according to their year of sampling (see Tables S1 and S2 for accession numbers).

A final data set comprising ancient mtDNA sequences from bison was analysed (kindly provided by B. Shapiro). The time between first and last sampling time points was *c.* 55 000 years. Sequences were assigned sampling index values according to 5000 year intervals (based on the mutation rate in Shapiro *et al.* (2004)).

NEUTRAL COALESCENT SIMULATIONS

To explore how the TC values obtained from empirical data sets varied from those expected under neutral, clock-like evolution, we simulated serially sampled coalescent trees (Rodrigo & Felsenstein 1999) under four specified demographic models: constant population size, exponential growth, logistic growth and sinusoidal population size change. In each case, the sample sizes and sequence dates exactly followed those of the HIV-1 and HCV empirical data sets. Values of the population size parameter $N_e\tau$ (effective population size*generation length) were chosen to be consistent with published estimates for each data set (HIV-1: $N_e\tau = 10$; HCV-1: $N_e\tau = 100$; mtDNA: $N_e\tau = 1000$). Custom scripts were used to perform the simulations (available from the authors on request).

BAYESIAN PHYLOGENETIC INFERENCE

For each of the fourteen empirical data sets, the posterior distribution of trees was inferred using MrBayes (Ronquist & Huelsenbeck 2003). The general time reversible nucleotide substitution model with gamma-distributed rate variation among sites was used, and no molecular clock was assumed. Two independent chains of 3×10^7 states were analysed, with samples taken every 3×10^4 states. Chain mixing was determined by the standard deviation of the chains being < 0.01. For the HCV data sets, an outgroup sequence from a different subtype of HCV genotype 1 was used. The HIV-1 and bison data sets

Table 1. Sampling schemes of the data sets

Patient	Number of time points	Range of sampling[a]
Shankarappa pt. 1	15	8·5
Shankarappa pt. 2	14	10·08
Shankarappa pt. 3	10	7·66
Shankarappa pt. 5	16	6·5
Shankarappa pt. 6	10	5·83
Shankarappa pt. 7	9	6·5
Shankarappa pt. 8	13	7·66
Shankarappa pt. 9	12	10·92
Shankarappa pt. 11	6	8·08
HCV 1a	18	19
HCV 1b	15	19
Ancient mtDNA bison	11	55 182

HCV, hepatitis-C virus.
[a]Sampling time given in years.

were rooted on the branch that resulted in the highest correlation of root-to-tip genetic distances against time (as performed by Path-O-Gen, available from http://tree.bio.ed.ac.uk/software/pathogen/). The posterior tree distribution for each data set was subsequently thinned to 200 phylogenies for computational tractability, and TC was calculated for each phylogeny as described earlier.

Results

PERFORMANCE OF THE TC STATISTIC

The TC statistic measures the position and degree of topological clustering of sequences sampled at different times in a serially sampled tree. It can be interpreted as the expected deviation of an observed topology from the null hypothesis of no temporal clustering, as a proportion of the range of possible values.

It is expected that as *N* increases, the difference between the mean score of the null distribution (S_{mean}) and the minimum score under perfect temporal structure (Min) will increase, because S_{mean} is a sum across branches. However, it is less clear whether the same holds true for the variance of the null distribution, and it is unknown how the phylogenetic topology will affect the null distribution of TC.

To investigate these issues, we simulated phylogenies with different numbers of taxa (*N*), different numbers of time points (*t*) and different tree asymmetries (Fig. 2a–d). For each scenario, the proportion of available parameter space occupied by the null distribution was calculated. As shown in Fig. 2e, when the number of taxa was small (*N* = 4), the null distribution stretched across the entire parameter space, for all topologies. As the number of taxa increased modestly (to *N* = 16), the proportion decreased rapidly to *c.* 0·6. The proportion continued to decrease as *N* increased to < 0·1 when *N* = 512. Thus, the null distribution occupied a decreasingly small range of total parameter space as log(*N*) increased, reflecting an increase in potential sensitivity and power to reject the null hypothesis. The effect of phylogenetic topology on the proportion was noticeable: the asymmetric tree resulted in comparatively tighter null distributions than the other topologies, particularly for *N* > 32. These results also confirmed our prior belief that the appropriate null test for the TC statistic involves randomizing sample indices (the *k* values) among taxa, not randomizing the phylogenetic topology.

Next, we assessed the effect of increasing the number of sampling times (*t*) while *N* was held constant (Fig. 2f). As *t* increased, there was a very slight increase in the proportion of parameter space covered by the null distribution, but this was less apparent when larger sample sizes were used (*N* = 100). This demonstrated that even in extreme cases (such as incorporating only four samples from each of 25 sampling times), the potential power of the TC statistic should remain practically high.

Taken together, these results indicated that, as *N* increased, the precision and potential power of the TC statistic increased and that the statistic may be weak when applied to small data sets (*N* < 20).

EMPIRICAL DATA SETS

The TC statistic was calculated for a number of previously published empirical data sets. Specifically, for each data set, the mean TC statistic of 200 trees sub-sampled from the posterior distribution of trees was calculated. For the nine within-patient HIV-1 data sets (Shankarappa *et al.* 1999), the degree of TC was high in all cases but differed significantly among patients (Fig. 3a). For all patients, the null hypothesis could be rejected for all 200 phylogenies investigated (i.e. S_{obs} was less than S_{min} for all tip randomizations, $P < 0.001$). All these data sets had mean TC statistics between 0·3 and 0·7 with narrow confidence intervals (i.e. there was little variation in TC among trees in the posterior distribution) (see Fig. 3a). A representative rooted maximum clade credibility (MCC) tree is shown for p5 (Fig. 3b). The ladder-like phylogenetic topology is consistent with significant temporal clustering.

The mean TC statistic was also calculated for phylogenies estimated from sequences of the E1E2 and NS5A genomic regions of HCV subtypes 1a and 1b (Fig. 3c). The null hypothesis of no TC (i.e. $S_{obs} < S_{min}$) could not be rejected for any of the HCV-1b phylogenies. For the HCV-1a data set, the null hypothesis was rejected for only *c.* 25% of the phylogenies investigated (E1E2: 44/200; NS5a: 46/200). The TC values were all very low, indicating little TC in these data sets. A representative MCC tree is shown for HCV-1a E1E2 (Fig. 3d). This shows a lack of temporal structure, with sequences from all time points intermixed.

For the ancient bison mtDNA data set (Shapiro *et al.* 2004), the mean TC value (0·19) was intermediate between the HIV-1 and HCV-1a and 1b data sets, and S_{obs} was significantly lower than the null distribution for all 200 phylogenies investigated (Fig. 3e). A low level of temporal structure was evident in the MCC tree (Fig. 3f).
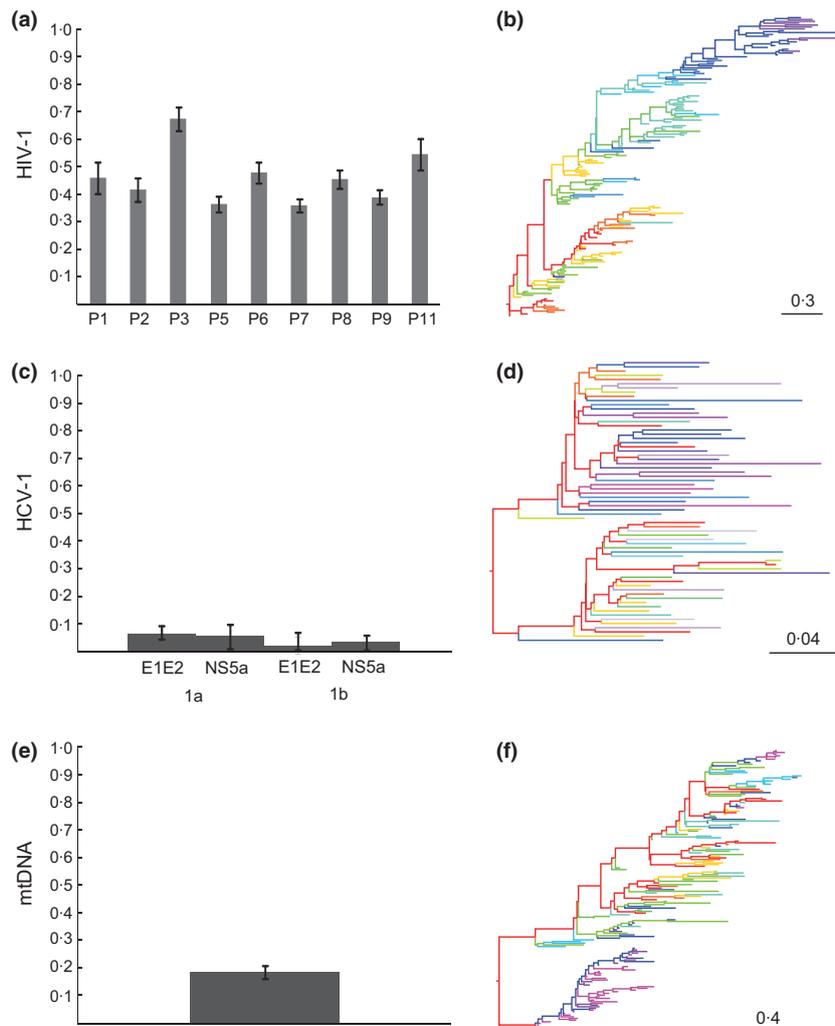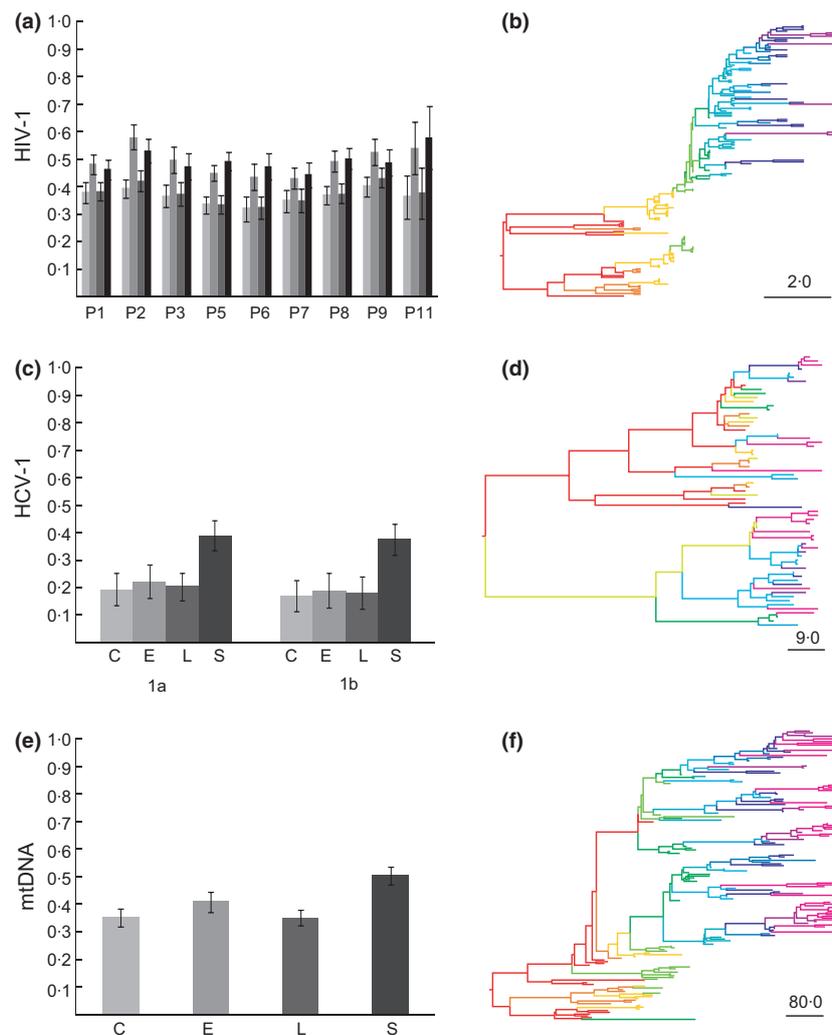

Fig. 3. Temporal clustering (TC) values for the empirical data sets (a) HIV-1 intra-patient data sets (c) hepatitis-C virus (HCV) subtype 1a and 1b data sets (e) bison ancient mtDNA data set. TC values were calculated as described in the main text, for each of 200 topologies randomly sampled from the posterior distribution of trees, given the data. The TC values are plotted on the *y*-axis, with the height of the bar indicating the average TC and the error bars representing ±2 SD among the 200 topologies. (b), (d) and (f) Representative rooted maximum clade credibility (MCC) trees for the HIV-1, HCV and mtDNA data sets, respectively (p5 for HIV-1; E1E2 subtype 1a for HCV). Branch lengths are units of substitutions/site (see scale bar). Branches are coloured using a rainbow spectrum according to the actual (or reconstructed) time point, where red = earliest time point and purple = latest time point.

NEUTRAL COALESCENT SIMULATIONS

We simulated neutrally evolving data sets under the hypothesis of a strict molecular clock using the serially sampled coalescent process (Rodrigo & Felsenstein 1999). Four demographic models were used (see Methods). For each empirical data set and demographic model, we simulated 100 coalescent phylogenies using the temporal sampling scheme of the corresponding empirical data set. The mean TC statistic was then calculated for each set of 100 simulated phylogenies.

The TC scores were moderate for all data sets simulated under the HIV-1 sampling scheme, ranging from 0·3 to 0·6 (Fig. 4a), with corresponding phylogenies exhibiting a strong ladder-like shape (Fig. 4b). The confidence intervals of TC

were larger for P11, correctly reflecting the smaller number of time points ($t = 6$) in this data set. Several important points arise from these simulation results. First, the TC statistic varied significantly among demographic models on the same data set. When sequences are all sampled at the same time, the neutral coalescent process is, by definition, independent of the phylogenetic topology. However, the act of sampling sequences through time imposes constraints on the topology; hence, the topology (and TC) is dependent on the neutral serially sampled coalescent process. Second, the null hypothesis of no TC could be rejected for all the simulated HIV phylogenies. This demonstrates that the property of 'temporal clustering' is not solely the result of strong positive selection: neutrally evolving serially sampled populations are also capable of exhibiting this

Fig. 4. Temporal clustering (TC) values for data sets simulated under a neutral coalescent process. Simulations were performed under four demographic models: constant population size (C), exponential growth (E), logistic growth (L) and sinusoidal change (S). In each case, trees were simulated under the sampling times of the corresponding empirical data set: (a) 9 intra-host HIV-1 data sets, (c) hepatitis-C virus (HCV) subtype 1a and 1b data sets (e) bison ancient mtDNA data set. For each data set/model combination, 100 coalescent trees were simulated, and the TC value of each tree was calculated as described in the main text. The TC values are plotted on the *y*-axis, with the height of the bar indicating the average TC and the error bars representing ± 2 SD among the 100 simulations. (b), (d) and (f) Representative rooted maximum clade credibility (MCC) trees for the simulated HIV-1, HCV and mtDNA data sets, respectively (p5 for HIV-1; E1E2 subtype 1a for HCV). Branch lengths are units of years (see scale bar). Branches are coloured using a rainbow spectrum according to the actual (or reconstructed) time point, where red = earliest time point and purple = latest time point.

behaviour. Third, for many data sets (e.g. P3), the empirical TC statistic was significantly greater than that under any coalescent demographic model (see Discussion).

The same strategy was also used to simulate trees according to the sampling times of the HCV-1a and 1b data sets (Fig. 4c). TC values for the HCV simulations were lower (0·2–0·4) than those for the HIV-1 simulations, which reflect differences in the sampling schemes of these two data sets. The temporal range of sequence sampling times (proportional to the total age of the tree) is much higher for the HIV data sets than for the HCV data sets. Again, all phylogenies resulted in a significant TC index for 100% of simulations. A representative MCC phylogeny from these simulations displayed modest temporal structure (Fig. 4d).

For the bison mtDNA simulations (Fig. 4e), the TC values were approximately equivalent to those obtained for the HIV-1 simulations, and the corresponding phylogeny was ladder-like, although with several contemporaneous lineages (Fig. 4f). Like the HIV-1 simulations, the range of sampling times for the bison mtDNA simulations is large, with sampled sequences placed reasonably close to the root of the tree. Thus, TC appears to increase as the range of sampling times (as a proportion of tree length) increased, again demonstrating that significant 'temporal clustering' can arise from temporal sampling in the absence of positive selection.

In all data sets, the TC values of phylogenies generated under the constant size and logistic growth population models were almost always significantly lower than those generated under a model of sinusoidal population size change (TC values simulated under exponential growth were intermediate). This makes intuitive sense: the sinusoidal model represents repeated genetic bottlenecks through time, which would remove coexisting lineages and therefore increase temporal structure. Thus, selectively, neutral demographic change can also generate significant temporal clustering.

Discussion

The use of serially sampled sequences can provide insights into evolutionary and demographic processes that are unobtainable from genetic samples taken from one time point. The shape of the sample phylogeny is thought to reflect population genetic forces, such as selection, population subdivision and demographic change (Grenfell *et al.* 2004). The TC statistic introduced here is the first quantitative measure of the 'temporal clustering' of a tree topology. 'Temporally clustered' phylogenies inferred without the constraint of a molecular clock are characterized by several properties, including (i) sequences collected earliest tend to be closer to the root, (ii) sequences from one time point tend to be directly ancestral to those from the next time point and (iii) sequences sampled at the same time tend to cluster with each other. Violations of these expectations indicate a deviation from a perfectly temporally clustered tree and can provide insight into the population under investigation.

Importantly, our results demonstrate that strong and significant TC is expected under neutral evolution simply as a result

of sampling sequences serially through time (Fig. 4). The observation of significant TC does not, therefore, uniquely demonstrate the action of strong or recurrent positive selection. Furthermore, we found that different demographic histories can result in significantly different levels of topological clustering, whereas this behaviour cannot occur if sequences are all sampled at the same time (in the absence of population sub-division). Repeated genetic bottlenecks (as represented here by sinusoidal changes in effective population size) give rise to phylogenetic topologies that are significantly more temporally clustered.

It is useful to consider what factors might lead an empirical phylogeny to have a higher or lower TC value than that seen under neutral coalescent simulation. The TC scores of both the HCV and bison mtDNA empirical phylogenies are lower than under any of the coalescent simulations. There are two likely explanations. First, the simulated data sets were generated according to a strict molecular clock, whereas both the HCV and bison mtDNA sequences are known to exhibit significant evolutionary rate variation among lineages (e.g. Pybus *et al.* 2009). This rate variation will lower the degree of topological temporal structure in the empirical phylogenies. Second, both these data sets are characterized by population subdivision (Shapiro *et al.* 2004), which has the effect of promoting lineage coexistence and therefore reducing the empirical TC value relative to the simulated value.

In comparison, the empirical and simulated TC values for the intra-patient HIV-1 data sets are quite similar (Figs 3a and 4a), even though these data sets do show rate variation among lineages (Lemey, Rambaut, & Pybus 2006). However, some of the empirical HIV-1 data sets, particularly P3, have *higher* empirical TC scores than under most, or all, of the coalescent simulations. Because the simulations are already perfectly clock-like, this *must* reflect the action of a process that creates stronger temporal clustering. One plausible candidate is positive selection, which purges coexisting lineages even more strongly than genetic drift under sinusoidal population size change. Indeed, the HIV-1 data sets used here are known to be under strong and recurrent positive selection (Williamson 2003). Intriguingly, the data set with the fastest estimated rate of molecular adaptation (P3; Williamson 2003) is also that with the highest TC value. In summary, therefore, positive selection *can* increase phylogenetic temporal clustering, but is not unique in being able to do so; both the range of sampling dates and the demography of the sampled population can also significantly increase TC.

The TC statistic shares some similarities with the commonly used Slatkin–Maddison test (Slatkin & Maddison 1989), with several important distinctions. First, the TC statistic treats sampling times as ordered character states, not unordered trait labels. Second, the use of an irreversible matrix enforces a unidirectional time course, such that later-sampled sequences cannot give rise to earlier sequences. This essential biological assumption also reduces the state space of ancestral configurations and avoids many equally parsimonious reconstructions. Third, in the implementation described here, the TC statistic incorporates phylogenetic uncertainty by averaging over a set

of trees from a posterior distribution (although TC can also be applied to individual phylogenies, or to a set of bootstrap trees). Finally, the TC statistic is rescaled between zero and one, so it can be directly compared between data sets of different sizes. While theoretically applicable to trees inferred under a molecular clock assumption, TC would likely be biased upwards (as seen in the simulated data sets for HCV and bison mtDNA).

The TC statistic is flexible and can be adapted for testing various user-specified hypotheses. For example, the assignment of sampling indices to taxa can be defined by the user to best suit the organism under study. The irreversible matrix could also be altered so that rather than sampling times being represented by consecutive integers, a weighted matrix could be employed in which the cost between two sampling times is weighted by the actual amount of time separating them (although the results described here were robust to the weighting scheme; data not shown). Finally, our results have shown that TC could have uses beyond the quantification of temporal clustering, in particular as a tool to investigate the interplay between the demographic and selective processes that shape observed phylogenies. Although this study is limited to the use of TC as a measure of phylogenetic topology, our results suggest that TC statistic could potentially be developed into a quantitative measure of positive selection.

Acknowledgements

References

Agapow, P. & Purvis, A. (2002) Power of eight tree shape statistics to detect nonrandom diversification: a comparison by simulation of two models of cladogenesis. *Systematic Biology*, **51**, 866–872.

Bush, R., Bender, C., Subbarao, K., Cox, N. & Fitch, W. (1999) Predicting the evolution of human influenza A. *Science*, **286**, 1921–1925.

Colless, D.H. (1982) Review of *Phylogenetics: the theory and practice of phylogenetic systematics. Systematic Zoology*, **31**, 100–104.

Drummond, A., Pybus, O.G. & Rambaut, A. (2003) Inference of viral evolutionary rates from molecular sequences. *Advances in Parasitology*, **54**, 331–358.

Drummond, A.J., Pybus, O.G., Rambaut, A., Forsberg, R. & Rodrigo, A.G. (2003) Measurably evolving populations. *Trends in Ecology and Evolution*, **18**, 481–488.

Drummond, A.J., Rambaut, A., Shapiro, B. & Pybus, O.G. (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.

Fitch, W. (1971) Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Zoology*, **20**, 401–416.

Fusco, G. & Cronk, Q.C.B. (1995) A new method for evaluating the shape of phylogenies. *Journal of Theoretical Biology*, **175**, 235–243.

Gray, R.R., Tatem, A.J., Lamers, S., Hou, W., Laeyendecker, O., Serwadda, D. *et al.* (2009) Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. *AIDS*, **23**, F9–F17.

Grenfell, B.T., Pybus, O.G., Gog, J.R., Wood, J.L., Daly, J.M., Mumford, J.A. & Holmes, E.C. (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, **303**, 327–332.

Ho, S.Y. & Gilbert, M.T. (2010) Ancient mitogenomics. *Mitochondrion*, **10**, 1–11.

Jombart, T., Eggo, R.M., Dodd, P.J. & Balloux, F. (2011) Reconstructing disease outbreaks from genetic data: a graph approach. *Heredity*, **106**, 383–390.

Lemey, P., Rambaut, A. & Pybus, O. (2006) HIV evolutionary dynamics within and among hosts. *AIDS Reviews*, **8**, 125–140.

Maddison, W.P. & Maddison, D.R. (2003) *MacClade Version 4: Analysis of Phylogeny and Character Evolution*. Sinauer Associates, Sunderland, Massachusetts.

Magiorkinis, G., Magiorkinis, E., Paraskevis, D., Ho, S., Shapiro, B., Pybus, O., Allain, J. & Hatzakis, A. (2009) The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. *PLoS Medicine*, **6**, e1000198.

Parker, J., Rambaut, A. & Pybus, O. (2008) Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infection, Genetics and Evolution*, **8**, 239–246.

Purvis, A., Katzourakis, A. & Agapow, P. (2002) Evaluating phylogenetic tree shape: two modifications to Fusco & Cronk's method. *Journal of Theoretical Biology*, **214**, 99–103.

Pybus, O. & Rambaut, A. (2009) Evolutionary analysis of the dynamics of viral infectious disease. *Nature Reviews Genetics*, **10**, 540–550.

Pybus, O., Barnes, E., Taggart, R., Lemey, P., Markov, P., Rasachak, B. *et al.* (2009) Genetic history of hepatitis C virus in East Asia. *Journal of Virology*, **83**, 1071–1082.

Rodrigo, A. & Felsenstein, J. (1999) Coalescent approaches to HIV-1 population genetics. *Molecular Evolution of HIV* (ed K. Crandall), pp. 233–274. Johns Hopkins University Press, Baltimore, Maryland.

Ronquist, F. & Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Salemi, M., Lamers, S.L., Yu, S., de Oliveira, T., Fitch, W.M. & McGrath, M.S. (2005) Phylodynamic analysis of human immunodeficiency virus type 1 in distinct brain compartments provides a model for the neuropathogenesis of AIDS. *Journal of Virology*, **79**, 11343–11352.

Shankarappa, R., Margolick, J.B., Gange, S.J., Rodrigo, A.G., Upchurch, D., Farzadegan, H. *et al.* (1999) Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *Journal of Virology*, **73**, 10489–10502.

Shao, K. & Sokal, R.R. (1990) Tree balance. *Systematic Zoology*, **39**, 266–276.

Shapiro, B., Drummond, A.J., Rambaut, A., Wilson, M.C., Matheus, P.E., Sher, A.V. *et al.* (2004) Rise and fall of the Beringian steppe bison. *Science*, **306**, 1561–1565.

Slatkin, M. & Maddison, W.P. (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics*, **123**, 603–613.

Wang, T.H., Donaldson, Y.K., Brettle, R.P., Bell, J.E. & Simmonds, P. (2001) Identification of shared populations of human immunodeficiency virus type 1 infecting microglia and tissue macrophages outside the central nervous system. *Journal of Virology*, **75**, 11686–11699.

Williamson, S. (2003) Adaptation in the env gene of HIV-1 and evolutionary theories of disease progression. *Molecular Biology and Evolution*, **20**, 1318–1325.

Yang, Z., O'Brien, J.D., Zheng, X., Zhu, H.Q. & She, Z.S. (2007) Tree and rate estimation by local evaluation of heterochronous nucleotide data. *Bioinformatics*, **23**, 169–176.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Table S1. Accession numbers and year of sampling for HCV-1a sequences.

Table S2. Accession numbers and year of sampling for HCV-1b sequences.

Data S1. Tutorial for calculating the TC statistic using Macclade.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organised for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.