Published in final edited form as:

Comput Biol Med. 2015 March 1; 0: 1–13. doi:10.1016/j.compbiomed.2014.12.013.

Cladograms with Path to Event (ClaPTE): A novel algorithm to detect associations between genotypes or phenotypes using phylogenies

Samuel K Handelman^{a,b,*}, Jacob M. Aaronson^c, Michal Seweryn^b, Igor Voronkin^c, Jesse J. Kwiek^d, Wolfgang Sadee^a, Joseph S. Verducci^e, and Daniel A. Janies^f

Samuel K Handelman: handelman.9@osu.edu

^aDepartment of Pharmacology, Ohio State University College of Medicine, 5072 Graves Hall, 333 West 10th Avenue, Columbus, OH 43210, United States

^bMathematical Biosciences Institute, The Ohio State University, Jennings Hall 3rd Floor, 1735 Neil Avenue, Columbus, OH 43210, United States

^cDepartment of Biomedical Informatics, Ohio State University College of Medicine, 3190 Graves Hall, 333 West 10th Avenue, Columbus, OH 43210, United States

^dDepartment of Microbial Infection & Immunity and Department of Microbiology, The Ohio State University, 788 Biomedical Research Tower, 460 West 12th Avenue, Columbus, OH 43210, United States

^eDepartment of Statistics, The Ohio State University, 404 Cockins Hall, 1958 Neil Avenue, Columbus, OH 43210-1247, United States

^fDepartment of Bioinformatics and Genomics, University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC 28223-0001, United States

Abstract

Background—Associations between genotype and phenotype provide insight into the evolution of pathogenesis, drug resistance, and the spread of pathogens between hosts. However, common ancestry can lead to apparent associations between biologically unrelated features. The novel method Cladograms with Path to Event (ClaPTE) detects associations between character-pairs (either a pair of mutations or a mutation paired with a phenotype) while adjusting for common ancestry, using phylogenetic trees.

© 2015 Elsevier Ltd. All rights reserved.

*Corresponding author: Tel.: +1 614 688 1313.

Availability of supporting data he scripts needed to run ClaPTE are available *via* GitHub.

HIV sequences used are deposited in Genbank, with accession numbers JF722674 to JF723201 and JQ778319 to JQ778843.

Influenza sequences used were obtained from Genbank, a list of accession numbers is provided in the Supplement.

Protein clusters were obtained from the NCBI's protein cluster ftp site; accession codes are given in the text.

Conflict of interest statement: None declared.

Appendix A. Supporting information: Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.compbiomed.2014.12.013>.

Methods—ClaPTE tests for character-pairs changing close together on the phylogenetic tree, consistent with an associated character-pair. ClaPTE is compared to three existing methods (independent contrasts, mixed model, and likelihood ratio) to detect character-pair associations adjusted for common ancestry. Comparisons utilize simulations on gene trees for: HIV Env, HIV promoter, and bacterial DnaJ and GuaB; and case studies for Oseltamavir resistance in Influenza, and for DnaJ and GuaB. Simulated data include both true-positive/associated character-pairs, and true-negative/not-associated character-pairs, used to assess type I (frequency of p -values in true-negatives) and type II (sensitivity to true-positives) error control.

Results and conclusions—ClaPTE has competitive sensitivity and better type I error control than existing methods. In the Influenza/Oseltamavir case study, ClaPTE reports no new permissive mutations but detects associations between adjacent (in primary sequence) amino acid positions which other methods miss. In the DnaJ and GuaB case study, ClaPTE reports more frequent associations between positions both from the same protein family than between positions from different families, in contrast to other methods. In both case studies, the results from ClaPTE are biologically plausible.

Keywords

Phylogenetics; Correlated evolution; Genotype-phenotype association; Genetic simulation; Drug resistance; Influenza evolution; HIV evolution; Protein evolution

1. Background

Identifying associations between genotypic and phenotypic features of organisms is a basic method for discovering the consequences of genetic changes. However, common ancestry among the organisms can complicate the interpretation of these associations between genotypes and/or phenotypes. In particular, common ancestry violates the assumption of independence among observations, assumed by classical statistical approaches. Although methods exist to detect these associations while accounting for shared ancestry, large molecular data sets introduce a further challenge because multiple hypothesis correction is required. This study is motivated by observed limitations of existing methods (*i.e.* EMMAX [1], Bayes Traits [2] and Felsenstein's independent contrasts [3]) for proper type I error control in detecting genotype-phenotype associations specifically in simulations modeled on a data set of HIV molecular sequences and phenotypic data [4]. Because these methods did not achieve proper type I error control in our simulations, we did not include these results in Kumar et al. [4]. In these unpublished simulation results, each of these three methods produced an excess of small p -values where no association existed. This excess of small p -values implied that these methods would incorrectly report associations when applied to real data, especially when multiple hypotheses are tested; therefore, we did not publish the results of these methods. However, this study motivated us to develop new methods that we hoped would display proper type I error control to test for associations between genotypes and/or phenotypes.

In most experimental designs, genotype-matched controls allow statistical approaches to detect associations between genotype and phenotype while common ancestry is either matched or controlled. For example, libraries of mutant organisms [5,6] can be screened for

phenotypic deficiencies to identify the function of individually deleted genes, against an otherwise genetically identical control. Likewise, targeted mutation of individual genes is a mainstay of experimental molecular biology [7]. However, for studies outside of model systems, researchers rely on uncontrolled (*e.g.* genetic surveillance data) or partially-controlled experiments (*e.g.* case-control studies) where the results include natural observations. These natural observations will typically exhibit significant covariance [3,8], meaning that if two organisms share any one feature then those two organisms are biased towards sharing many different features. For example, lions and tigers both have canine teeth *and also* have binocular vision because they share a common ancestor with both traits. Common ancestry as in these examples does not rule out the possibility of an underlying association between two traits, but at the same time this dependence between observations can give the false appearance of association or even causation between two unrelated features. The purpose of this study is to develop a method for association analysis, termed Cladograms with Path to Event (ClAPTE), designed to avoid identifying associations arising solely from common descent, thus achieving control over type I error even in samples containing related observations.

Detecting associated changes in character data (*e.g.* mutations and/or phenotypic or environmental change) from naturally occurring organisms is a classic problem in phylogenetics [9]. In the 1970s, Felsenstein described these problems as character-character associations [10]. ClAPTE builds on the approach of Felsenstein [3] by incorporating patristic distances, the number of accumulated mutations between nodes, between reconstructed changes on the phylogenetic tree. Powerful methods have been presented in the evolutionary biology literature [11–26] that are optimized for the refinement of particular types of biological or evolutionary models, on trees or pedigrees. A parallel literature addresses a similar problem in the field of molecular evolution for individual proteins [27–32], although many of these methods do not account for common ancestry. With the advent of large-scale genotyping and whole-genome sequencing in humans, a third group of methods [1,33–39] has been developed to correct for heritable covariance (arising from population structure, family structure and cryptic relatedness) in humans. The simulation experiments reported here bear on the subset of problems where the characters under study are categorical (as opposed to continuous) and where the observations are non-independent, as is typically the case in natural observations.

In this study we use simulation experiments to assess ClAPTE in comparison with alternative approaches in detecting character-character associations, while adjusting for common ancestry. The structure of these simulation experiments were inspired by outcomes for a study on association between viral sequence data with mother-to-child HIV transmission [4], but do not contain the actual data of that report. These simulation experiments compare methodologies according to four possible criteria: (1) sensitivity at low false positive rates, (2) freedom from bias arising from the rate of change in the characters, (3) computational efficiency, and (4) proper type I error control where no association exists. This last criterion is most relevant to conventional hypothesis-rejection approaches given a large number of multiple hypotheses.

This report includes a case study on the use of ClaPTE to search for additional permissive mutations of a well-characterized resistance mutation (H274Y in the influenza reference neuraminidase (NA)) to the anti-viral drug oseltamavir (Tamiflu®) [40], which is of major concern in immuno-compromised populations [41]. When one mutation reduces the fitness cost of a second mutation, the first mutation is said to be permissive of the second mutation. H274Y NA influenza are deficient in export of NA to the virion surface [42,43]; therefore other mutations that rescue this export defect may be permissive of H274Y [44]. In the case of the H1N1 viral lineage used here, one such permissive mutation (R194G NA [45]) is already present [46]. While additional permissive mutations have been reported based on *in vitro* results, it is an open question whether further mutations might accelerate the appearance of oseltamavir resistance *in vivo*.

A second case study utilizes amino acid sequences for two biologically-unrelated protein families: NCBI protein cluster [47] PRK10767 containing *E. coli* protein DnaJ (a folding chaperone), and protein cluster PRK05567 containing *E. coli* protein GuaB (which converts inosine 5' monophosphate to xanthosine 5' monophosphate). These two proteins were chosen in part because we found no biological link between the two had ever been reported, although such a link cannot ever be ruled out entirely. Therefore, in this case study, associations should be detected between pairs of amino acid positions in the same protein chain, and not for pairs of amino acid positions where one is drawn from DnaJ and the second is drawn from GuaB.

2. Methods

2.1. Cladograms with Path to Event (ClaPTE)

ClaPTE is a test for an association between a pair of features of biological data. These two features are termed characters, the independent-character **X** and the dependent-character **Y**. **X** can be any biological feature, including a phenotype, an aspect of the environment, or a nucleotide or amino acid mutation. In order for the ClaPTE test to be valid, **Y** must be a nucleotide or amino acid mutation, because ClaPTE uses a weighted cladogram (a phylogenetic tree) to account for the effect of common ancestry on **Y**. If **Y** is for example a phenotype or a biological niche, the ClaPTE test would be valid only assuming a null hypothesis that changes in **Y** arrive at a rate proportional to the branch lengths, which would depend on an accurate molecular clock, difficult to achieve in practice [48].

In the simulations and use cases presented here, **X** and **Y** can take only two states each, **X** can be in state **c** (ancestral) or **C** (derived), and **Y** can be in state **r** (ancestral) or **R** (derived). Fig. 1A gives an example of two characters reconstructed on a phylogenetic tree: each tip on the tree corresponds to a sequence/observation, and the tree is color coded to represent the different states (For interpretation of the references to color in the figures, the reader is referred to the web version of this article.)

In the ClaPTE method, patristic distances between changes in **X** and changes in **Y** are assumed to be exponentially distributed (as signal times in a Poisson process [49]). The number of times a change in **X** is followed by a change in **Y** (a signal from the Poisson process) is compared against the expected number of such sequentially-ordered pairs of

changes under the null hypothesis of no association. The scaled difference between observed and expected sequential changes should be asymptotically χ^2 distributed, giving a p-value. This approach is derived from the Cochran–Mantel–Haenszel [50] (CMH) statistic for accelerated failure, as applied to a single strata of data. In ClaPTE, the CMH allows for a comparison between those lineages where **X** remains **c**, to those lineages where **X** adopts the derived state **C**. Under the alternative (non-null) hypothesis that **X=C** favors **Y** to become **R**, the lineages where **X=C** will contain more changes of **r** becoming **R** than would be expected, and the null hypothesis can be rejected. A formal exposition of the ClaPTE method, including relevant equations, is located in the Supplement in sections Formal specification of trees and two state character data, Relationship between phylogenetic trees and waiting times, and Formal specification of Cladograms with Path to Event (ClaPTE).

2.2. ClaPTE implementation

The ClaPTE implementation takes as input (aligned) character data including reconstructed values at internal nodes (typically a fasta-formatted nucleotide or amino acid sequence file, or a diagnosis file generated by POY), a phylogenetic tree with nodes matching the entries in the fasta file, and a list of ordered pairs of characters (by default, position numbers in the alignment file) for which an association should be tested, with optional regular expressions to specify ancestral and derived states. The ClaPTE implementation interprets the first character in each pair as independent (**X**) and tests for dependency in the second character (**Y**). In some cases, this ordering will either be obvious or will be a component of the hypothesis that an investigator wishes to test; for example, if testing the hypothesis that a change in drug regimen favors some mutation, the drug-regimen character would be **X** and the mutations would be **Y**s. If there is no obvious order, the default behavior of the ClaPTE implementation is to assign the shorter character (the one with fewer reconstructed changes) to be **X**, with ties broken arbitrarily. If the edge lengths on the tree describe the null-hypothesis-expectation for changes in only one of the characters, that character should be the dependent character **Y**, since ClaPTE does not require changes in **X** to be dependent on the branch lengths.

The ClaPTE implementation then assigns each change in **X** to be either censored (if no change in **Y** follows it), or failed (if a change in **Y** does follow it), compares the number of **X**-failed to the expected number of **X**-failed, and uses the pchisq function built into the statistical programming language R to report a test statistic. For an explicit specification of how ClaPTE operates, see ClaPTE pseudocode in the Supplement.

2.3. Phylogenetic trees used in simulations

Tree topologies and branch lengths chosen for use in this report are derived from two molecular data sets and from two extremes of the possible range of theoretical topologies. Fig. 2 provides a graphical summary of all of the trees used, which can be viewed in full in the Supplement Figs. S1–S6. In order to study the effect of errors in tree reconstruction on ClaPTE and on the other methods described below, these trees were chosen in pairs, which are mutually somewhat similar but which differ in ways that reflect expected failure modes in application of these methods to real data. In the Supplement Full descriptions of trees used in simulations contains the detailed characteristics of each phylogenetic tree used.

2.4. Simulation experiments

The mutation rates used in these forward simulations were guided by results from a previously reported study [4] on the relative transmissibility of different HIV variants from HIV positive pregnant women to their offspring during pregnancy. The independent character **X** is meant to represent movement into a transmission-specific placental compartment, which was observed in approximately 5 pregnancies; therefore, for these simulations **X** will evolve at a rate of 5. The dependent character **Y** is meant to represent an amino acid mutation which supports proliferation in this placental compartment, which we hypothesize would in turn support mother to child transmission; therefore, in those simulations where **X** and **Y** are uncoupled, **Y** evolves at a rate of 25 which was typical for amino acid positions potentially related to tropism [4]. For these forward simulations, all of these trees were normalized to a length of 1 (thus, the rates of 5 and 25 represent the total number of expected changes on the entire tree). Because the lengths of edges are used only to generate simulated transition probabilities, the total length of the tree and the evolutionary rate of any character are identifiable only as a product (*i.e.* a rate of 1 on a tree of length 1 would be exactly the same as a rate of 2 on the same tree scaled to a length of 1/2.) Nodes in these simulations are allowed to adopt exactly four states: (**X**=**c**, **Y**=**r**); (**X**=**C**, **Y**=**r**); (**X**=**c**, **Y**=**R**); (**X**=**C**, **Y**=**R**), and nodes change state from **c** to **C** with rate θ_{Ff} , from **C** to **c** with rate θ_{Fb} , from **r** to **R** with rate θ_{Gf} , and from **R** to **r** with rate θ_{Gb} .

Two types of simulations were performed: in uncoupled simulations (**X**,**Y**), the values of θ_{Gf} and θ_{Gb} are independent of **X**; in coupled simulations, θ_{Gf} and θ_{Gb} take values of 250 or 10 depending on the state of **X** (chosen arbitrarily.)

For each tree, 250,000 uncoupled simulations of **X**,**Y** are performed and 10,000 simulations of **X**,**Y'** are performed (see Table 1). For equations and specifics, see section Forward simulation on phylogenetic trees in the Supplement.

These 250,000 uncoupled character pairs **X**,**Y** are then weighted so that the distribution of relative frequencies of the states **c**, **r**, **C** and **R** are equivalent those in the 10,000 coupled character pairs **X**, **Y'**. This weighting is done to avoid introducing a bias simply towards detecting similarity in frequencies (*i.e.* a method that called every pair coupled when the frequency of **C** was the same as the frequency of **R** might appear to do well, if this correction is not applied.) For details on the weighting of pairs, see Weighting of character pairs in the Supplement.

2.5. Pre-processing associated with ClaPTE

POY [51,52] version 4 is used, along with the tree, to produce a list of character changes (POY command diagnose under unweighted parsimony) associated with each simulated character. This list of changes, along with the tree, is the input used by ClaPTE. Note that neither ClaPTE nor any of the other methods has access to Eq. (S2A–D), or to the parameters used in those equations, or to the original states at internal nodes of the tree. ClaPTE carries an implicit assumption, that these reconstructed changes are correct, however any loss in performance arising from this implicit assumption is included in the simulation results presented here.

2.6. Other methods

BAYESTRAITS [53] is distributed as source code by the authors [54] and was compiled locally. Each phylogenetic tree was converted into nexus format, and each character pair was converted into binary characters ($\mathbf{X}=\mathbf{c} \rightarrow \mathbf{X}=\mathbf{0}$; $\mathbf{X}=\mathbf{C} \rightarrow \mathbf{X}=\mathbf{1}$; $\mathbf{Y}=\mathbf{r} \rightarrow \mathbf{Y}=\mathbf{0}$; $\mathbf{Y}=\mathbf{R} \rightarrow \mathbf{Y}=\mathbf{1}$); the tree and the character values are given as input to BAYESTRAITS for both a model with independent rates, and for a model with dependent rates. BAYESTRAITS reports a difference between the likelihood of the two models, and that difference in likelihoods is treated as a draw from a chi-square distribution with four degrees of freedom, as described in the BAYESTRAITS documentation [53].

EMMAX [1] and EIGENSTRAT [33] are distributed as source code by the respective authors and were each compiled locally. 25,000 Independent characters (*i.e.* character $\mathbf{X}$, all $\theta_F=5$ in Eq. (S2A)) from each simulation were converted into PLINK format [55] and the five principal components of the resulting smartpca run (a component of EIGENSTRAT) were given to EMMAX as covariates. Then, the interleaf distances between each taxon on the corresponding tree were converted into an identity by descent matrix using Eq. (S2A) with $\theta_F=5$ and the interleaf distance used as $\mathbf{L}_{ab}$ (see Supplement). Finally, for each pair of characters studied, $\mathbf{X}$ was converted into a SNP ($\mathbf{X}=\mathbf{c} \rightarrow \mathbf{AA}$ homozygote; $\mathbf{X}=\mathbf{C} \rightarrow \mathbf{TT}$ homozygote) in PLINK format [55], while the dependent character was converted into a phenotype ($\mathbf{Y}=\mathbf{r} \rightarrow \mathbf{0}$; $\mathbf{Y}=\mathbf{R} \rightarrow \mathbf{1}$). EMMAX reports p -values directly.

Felsenstein's independent contrasts (FIC) is distributed as part of the APE [14] and GEIGER [56] packages for the statistical programming language R. Each phylogenetic tree was converted into nexus format, and each character pair was converted into continuous valued characters of either 0 or 1 ($\mathbf{X}=\mathbf{c} \rightarrow \mathbf{X}=\mathbf{0}$; $\mathbf{X}=\mathbf{C} \rightarrow \mathbf{X}=\mathbf{1}$; $\mathbf{Y}=\mathbf{r} \rightarrow \mathbf{Y}=\mathbf{0}$; $\mathbf{Y}=\mathbf{R} \rightarrow \mathbf{Y}=\mathbf{1}$), in order that contrasts could be calculated. These trees and character values were given as input into FIC. FIC directly reports a regression p -value.

Fisher's exact test (implemented in Perl) is used in one comparison, incorporated into the Supplement. For that test, the contingency table is populated with the count of each of the four states, $\mathbf{rc}$, $\mathbf{rC}$, $\mathbf{Rc}$ and $\mathbf{RC}$, observed at the tips.

2.7. Cross-tree comparisons

ClaPTE and each of the other methods were also run using simulation results derived from one tree, but using a second, slightly different tree to adjust for common ancestry. The purpose of this exercise is to estimate the error that would be introduced by either undetected recombination or by a failure to properly reconstruct the underlying tree. Mechanically, this simply means that the different executable programs are run with simulation output from one tree, but with the tree file (or corresponding identity matrix and principle component analysis) from another tree. For the *in silico* trees, the results from a tree with relatively long interior branches (a “branch heavy” tree) are scored against a tree with relatively short interior branches (a “leaf heavy” tree). This corresponds to a case where the level of sequence similarity/cryptic relatedness between most pairs of individuals (except for immediate siblings or cousins) has been underestimated. In a topological sense, the leaves in the two trees correspond exactly so the labels are simply conserved. For the GuaB

and DnaJ trees, two leaves correspond if they originate from the same source organism, a comparison between the two reflects expected loss of power for comparisons dependent on different genes. For the GP120 and promoter trees, two leaves correspond if they are derived from sequences of the same amplicon, with the assumption that these sequences were isolated from two regions of the same HIV genome. Unlike the *in silico* trees, these two trees differ in topology, primarily due to the high recombination rate of HIV [57] but also potentially due to mistakes in the construction of one or another tree [58]. Fig. 2B illustrates these “cross tree” comparisons, showing the same mutations as in Fig. 2A, but reconstructed on the promoter tree instead of the GP120 tree. In the figures in the results, these cross-tree comparisons are identified by an arrow shown pointing from the tree on which the simulations are run, to the tree on which the results are scored for significance.

2.8. Computational complexity estimates

The running time estimates are from single processes, running in more than $1000 \times$ excess of needed memory, as a single thread on a 2.3 GHz AMD Opteron(TM) Processor 6276. Time is user time expended, reported by the GNU/LINUX time command; to convert to CPU cycles, multiply by 2.3 billion. The time estimates are averaged over 500 character pairs, derived from trees used for simulations. For data including a large number of character pairs, ClaPTE parallelizes naturally; that is, each pair of characters runs as an independent process. If needed, ClaPTE can also be parallelized internally by running the test independently on different partitions of the tree and aggregating the results; however, this approach is not used here.

2.9. Influenza data for Oseltamavir resistance study

The data set used included neuraminidase (NA) sequences from 4302 pandemic H1N1 strains available from the NCBI as of March 13, 2013 (a full list of identifiers is provided in the Supplement). NA sequences were excluded where codon 275 was not observed. The sequences were aligned using MAFFT v6.937b [59]. Suspect sequences that either broke the reading frame or that were missing the sialidase domain [60] were removed, leaving 4264 remaining sequences. Finally, a phylogenetic tree for the alignment was searched using RAxML [61] under the GTRCAT model of substitution for 100 replicates.

All amino-acid positions in the resulting alignment were converted into two states (ancestral or non-ancestral). Each pair of positions in the variable region of the sialidase [60] domain (corresponding to amino acid positions 149 to 284 in the neuraminidase reference sequence, see Supplement) with three or more reconstructed changes on the tree (a total of 64 positions or 2,016 position pairs) were processed using BayesTraits, FIC, and ClaPTE as described above.

2.10. Protein cluster case study

Protein clusters PRK05567 (containing *E. coli* protein GuaB) and PRK10767 (containing *E. coli* protein DnaJ) were downloaded from the NCBI protein clusters page. Strain names associated with each protein were extracted from NCBI records, and sequences were removed unless the source strain was represented exactly-once-each in PRK05567 and in PRK10767. Then, amino acid positions were trimmed from the alignment unless they met

the following criteria: no alignment to deletion, consensus amino acid present at between 50% and 75% frequency. The results of this procedure can be seen in the Supplement in sections GuaB (PRK05567) reference sequence, DnaJ (PRK10767) reference sequence and List of Organisms represented exactly once in both PRK05567 and PRK10767.

3. Results and discussion

The overall purpose of the methods considered in this report is to distinguish coupled pairs of characters from otherwise-similar but uncoupled pairs. Therefore, the results in this section deal with comparisons between pairs of characters (*i.e.* pairs of simulated nucleotide positions) that are either coupled ($\mathbf{X}, \mathbf{Y}'$ in Section 2) or uncoupled ($\mathbf{X}, \mathbf{Y}$ in Section 2).

3.1. Summary of methodological comparisons

Overall, ClaPTE controls type I error better than competing methods, while making at-most modest sacrifices in sensitivity. ClaPTE is more computationally efficient than competing approaches. Table 2 compares the 4 methods in this report on these criteria.

3.2. Summary statistics on cross-tree comparisons

In the results below, numerous comparisons are shown for simulations performed on one tree, but tested for associations using the covariance structure of another tree. The topology of the tree is invariant between pairs of the symmetrical trees (Fig. 1E–H), so the pairs of trees differ by 0 prune and regraft operations with a topological similarity of 100%.

The two HIV trees were chosen because of their biological interest and to reflect an extreme case of recombination introducing errors into covariance estimation. The `sprdiff` function in TNT [62], run with 1000 replicates of up to 1000 operations, reports that GP120 tree and the Env tree differed by at most 175 sub-tree prune and regraft operations, with a topological similarity of 42%, reflecting the high frequency of recombination in HIV genomes.

The GuaB and DnaJ trees were chosen to represent a moderate level of recombination, relevant when using a reference gene (for example, ribosomal sequences) to construct a tree used to evaluate associations for some other position in the genome. The `sprdiff` function in TNT [62], run with 1000 replicates of up to 1000 operations, found that GuaB tree and the DnaJ tree differed by at most 33 sub-tree prune and regraft operations, with a topological similarity of 74%, indicating a moderate number of horizontal gene transfer events.

3.3. Coupled characters are more likely than uncoupled characters to vary at a similar rate

Table 3 shows the difference between the Shannon Entropy values between pairs of characters from the different simulations, on different trees (note: this value is not the same as mutual information). Across all trees, the difference between the two entropy values is on average much smaller when the characters are generated by a coupled simulation. However, this change in relative entropy is inconsistent among different tree topologies used in the simulations.

These similarities in rates could confound the assessments reported here. In order to address this, the difference between the entropy at each position is used as a normalizing factor (a

“weight”) when comparisons on each tree are scored (see Section 2; the weight given to each observation is the ratio of the null column to the corresponding coupled column). However, because the bins are not infinitely small, this normalization is not guaranteed to entirely account for indirect effects arising from these differences in rate.

3.4. ClaPTE reports p -values that are suitable to null-hypothesis rejection

ClaPTE is suitable for multiple-null hypothesis rejection because p -values below some threshold α , in a data set where the null hypothesis is true, are reported at a frequency of no-more-than- α . Adherence to this bound is often called Type I error control. For these applications, the null hypothesis is that two characters are evolving independently. For the simulations reported here, methods that do not account for the tree fail to control type I error: in the symmetrical unbalanced tree, which is the most independent/best case for such methods, p -values below 10^{-10} are reported more than 1% of the time by Fisher's exact test on the states of the tips, see Fig. S7. For comparison across different methods, Figs. 3–5 show quantile/quantile plots. In each quantile/quantile plot, the frequency of different p -values on the uncoupled characters $\mathbf{X}, \mathbf{Y}$ is reported for the four methods considered in this report. Ideally, a method would fall on the dotted line exactly; a method falling below the dotted line is conservative but still controlled. For five of the comparisons (Figs. 3A and B, 4A and B, 5A and B), the true tree is known and used. ClaPTE is the only method that consistently reports low p -values bounded near the desired frequencies when the true tree is known.

For six of the comparisons (Figs. 3C and D, 4C and D, 5C and D), a cross-tree comparison is reported. These cross-tree comparisons are relevant for applications in which either measurement errors are made, or in which different portions of the genome have arisen from different phylogenies (horizontal gene transfer, recombination, *etc.*). While no method could perform perfectly when a covariate is subject to systematic error, these comparisons evaluate each method under the expected failure modes, when either the degree of relatedness among the subjects is under-estimated or the phylogenetic signal is lost (Fig. 3C and D), when the fine structure of the relatedness is incorrect but the overall shape is correct (Fig. 4C and D), or when a reference phylogeny might be used from other genes (Fig. 5C and D). ClaPTE does relatively well in minimizing the impact of the error (Figs. 3A and B, 4A and B, 5A and B contrasted with Figs. 3C and D, 4C and D, 5C and D).

We propose that the robust type-I error control of ClaPTE is a reflection of stronger underlying assumptions. Especially when testing multiple hypotheses or otherwise at a low significance cutoff, assumptions which are rarely violated can nonetheless compromise type-I error control; a test based on assumptions which are valid 99.9% of the time might have no issues at a $p < 0.05$ significance cutoff, but significant issues at a $p < 0.001$ cutoff.

3.5. ClaPTE achieves robust null hypothesis rejection without sacrificing sensitivity

In addition to rejecting a null hypothesis, a method should be sensitive to the alternative hypothesis. In this case the alternative hypothesis is that $\mathbf{Y}$ evolves at a different rate depending on the state of $\mathbf{X}$. That is, the type II error rate (false *acceptance* of the null) should be as low as possible for the desired significance threshold. Figs. 6–8 show R.O.C.

curves for each of the four methods across different trees. In the large data sets that are the focus of this report, performance at lower false-positive rates is of chief interest; therefore, the R.O.C. curves in Figs. 6–8 are displayed with the false positive rate in a log scale, emphasizing the important distinctions among the methods. Although EMMAX is less conservative than ClaPTE by the criterion shown in Figs. 3–5, this does not translate into improved sensitivity (Purple lines, Figs 6 and 7, all panels). Although the sensitivity varies among the different trials shown here, FIC (Green lines, Figs. 6–8, all panels) is the most sensitive method at low false-positive rates, albeit by a small margin. BayesTraits (Red lines, Figs. 6 and 8) is superior to other methods at high false-positive rates, and is competitive with other methods when the true tree is known (Figs. 6A and B, 8A and B); however, when errors in the tree branch lengths are introduced (Figs. 6C and D, 8C and D) BayesTraits suffers an immense loss of sensitivity. Furthermore, BayesTraits does not scale well to larger phylogenetic trees (Table 4 and previous sections; this is why BayesTraits results are not reported in Fig. 5). Finally, ClaPTE (Blue lines, Figs. 6–8, all panels) is at least competitive in sensitivity with all other methods in all cases but two (Figs. 6B and D, 8A and B), in exchange for achieving well-bounded p -values, which the other sensitive methods (BayesTraits and FIC, depending) do not.

3.6. Overview of reproducible differences among trees

ClaPTE demonstrates scalability, stringency and sensitivity in simulations, properties critical for detecting character-character associations. In particular, ClaPTE maintains good type I error control across a wide range of circumstances. However, with regards to sensitivity, certain unexpected results bear further explanation. A few method/tree combinations reported above show increased sensitivity when the wrong tree topology is used for scoring (FIC in Fig. 6D and B; EMMAX in Fig. 7C and A). In these cases, long edges/partitions occur in the tree used for simulation, and these long edges correspond to edges of a significantly different length in the alternative tree used for scoring. Because the branch is long, numerous changes occur on the same edge, magnifying the impact of numerically-small differences in the significance score, resulting in large fluctuations in the R.O.C. curve especially at very low false positive rates as seen here. By coincidence these fluctuations can seem to be an improvement. However, such improvements do not occur reproducibly for any of these methods across different ranges of trees, tree errors and simulation parameters.

For simulations carried out on trees derived from amplified HIV genomic sequences, the internal distribution of branch lengths appears to have an effect. On both trees, 34% of the length is on branches ending in leaves. Furthermore, the distribution of branch lengths is heavy-tailed (power-law like) in both trees. However, in the GP120 tree the 26 longest branches account for 33% of the length of the tree, whereas in the promoter tree, only the 19 longest branches account for the same 33% share of the length.

Thus, the distance from the root to each leaf is typically about 20% longer in the promoter tree than in the GP120 tree; for ClaPTE, this leads to a longer average number-of-accumulated-mutations on test, which is why ClaPTE is more sensitive on the promoter tree.

These branch lengths correspond to the time-on-test for ClaPTE. The distance from the root to each leaf is typically about 20% longer in the promoter tree than in the GP120 tree,

corresponding to more average time-on-test. A greater time-on-test is analogous to a process which is observed for longer lengths of time, and the generic expectation is that this would confer greater sensitivity. Overall, the observed trend is consistent with this generic expectation: ClaPTE performs better against competing methods when the average distance from each leaf to the root is long (Figs. 3A, 4B, 6A, 7B), but is less optimal when the average distance from each leaf to the root is short (Figs. 3B, 4A, 6B, 7A). When the tree is exceptionally concentrated on a few long internal branches (as the GuaB/DnaJ trees, Figs. 5 and 8), ClaPTE becomes hyper-conservative. This may be because such simulations are underpowered – with relatively few changes found only on the long internal branches. Under these circumstances BayesTraits performs well, beating the sensitivity of Felsenstein's independent contrasts even in cross-tree comparisons at low false positive rates.

3.7. ClaPTE is not biased towards similar rates

Fig. 9 shows modified R.O.C. curves that reflect the tendency of different methods to bias their results according to the difference in entropy seen in Table 3. In contrast to a conventional R.O.C. curve, these curves have a false positive rate, one for high entropy differences and one for low entropy differences, plotted on each axis. A method biased towards one or the other sort of false positives will deviate from the line across the diagonal. Although these biases are small, they are sufficient to confound results in large-scale genomic data sets. ClaPTE (blue lines, Fig. 9, all panels) shows a lower tendency to prefer highly similar rates, but in some cases (Fig. 9A) it may introduce a reverse bias towards dissimilar rates. Other methods prefer similar rates (the vertical axis) generally but there is not much consistency across the trees.

ClaPTE does not strongly depend on the relative rate of the two evolving characters when a correlation is tested. This is a critical feature in interpreting high-throughput sequencing data, especially as Bayesian approaches (explicit [63] or implicit [64–66]) are used to enrich the interpretation and reproducibility of the results. Consider incorporation of genotype associations in a Bayesian context to protein structure applications. If amino acid positions on the surface of a protein evolve at a similar rate, then a method biased to detect associations between positions with a similar characteristic rate will identify pairs of such amino acid positions as “correlated”. When such correlations are “validated” against the available objective data, experimentally characterized structures, the objective data will appear to verify the predicted correlations. In bacteria, genes with related functions (even if they do not directly interact) may have similar fitness consequences [67] and thus may be gained or lost at similar rates. Similar effects influence the frequency of human alleles [68]. In contrast to this, the reverse-bias seen in Fig. 9A may arise because time to event methods more readily detect acceleration of frequent events. While ClaPTE shows promise in controlling this source of bias, when constructing a predictor on a large volume of real data, the entropy-binning method used here might be needed as well.

3.8. ClaPTE is computationally efficient

Table 4 shows the computational complexity of the different approaches used here; these are measurements of the computational complexity of each comparison operation between two

characters. ClaPTE is at least as fast as FIC and ten times faster than either EMMAX or BayesTraits. However, there are additional computational costs associated with tree reconstruction for ClaPTE, BayesTraits and FIC. The computational cost of tree reconstruction is likely to be considerably higher than the computational cost associated with performing a PCA analysis using EIGENSTRAT, but will vary tremendously depending on the methods chosen and the properties of the data used to construct the tree.

3.9. Case study in pandemic H1N1

While the simulation results above show that ClaPTE may produce better controlled or more sensitive results than existing methods, this case study shows the level of qualitative differences between the results of ClaPTE and of existing approaches. This Oseltamavir use case represents a joint analysis of phylogenetic and epidemiological data [69], a major motivator for the ClaPTE method. ClaPTE is suitable to test for dependency of drug resistance in these data, because drug resistance is conferred by a single amino acid change, so it is suitable to treat it as a dependent-character Y . Under the null hypothesis of no association, this single drug-resistance mutation would arise on the phylogeny with likelihood proportional to the branch lengths.

Suppose the introduction of a drug shifts the fitness landscape experienced by a virus such that a mutation conferring drug resistance becomes favorable. In that case, in drug-exposed lineages the average number of other mutations (the patristic distance, above) that accumulate before the drug resistance mutation becomes fixed is expected to be smaller, compared to drug-naïve viral lineages. This is related to problems on which a time to event method might be applied. ClaPTE is analogous to a test for an association between the use of synthetic engine oil (analogous to a change in the first character) and the rate of engine failure (analogous to a change in the second character): if the average mileage (analogous to a patristic distance) between oilchange and engine failure is lower in cars given synthetic oil, then synthetic oil is associated with engine failure. However, ClaPTE differs from time to event methods in several respects. First, character changes are not directly observed but must be reconstructed on the phylogenetic tree (except in exotic cases such as simulated evolution experiments), while time to event methods may assume that failure times can be precisely assigned to intervals of time. Second, the phylogenetic tree itself includes parallel tests in sister lineages which are not found in the traditional time to event case, although the edges on the tree are similar to intervals on a timeline.

In the sialidase [60] domain of the pandemic H1N1 sequenced used for this study, neither BayesTraits nor ClaPTE report any new permissive mutations leading to the Oseltamavir resistance mutation at NA-274. The sialidase domain was chosen because previous permissive mutations were found in this region, and to make the problem more computationally tractable. Among all pairs of positions in the sialidase domain, BayesTraits reports no associations that are significant after multiple hypothesis correction: the lowest p -value BayesTraits reported is 0.00063, between positions NA-205 with NA-206, corresponding to a multiple-hypothesis corrected e -value of greater than 1. Among all pairs of positions in the sialidase domain, ClaPTE reports: a p -value of 0.00001 for NA-205 with NA-206, corresponding to an e -value of 0.02 and being nominally significant; and a p -value

of 0.00005 for NA-221 with NA-222 (BayesTraits reports a p -value of 0.04 for this position), which is nominally significant at a Benjamini-Hochberg false discovery rate of 0.05. FIC reports a p -value below precision (less than 2×10^{-16} in this install of R) for 53 position pairs including NA-205w/NA-206 and NA-221w/NA-222. EMMAX was not used in this case study, because the covariance matrix derived from these influenza sequences was rejected by the EMMAX executable as containing too many non-zero eigenvalues (*i.e.* EMMAX believes that the data set contained too many “closely related individuals” for it to process, and refuses to run.) As NA-205 and NA-206 are adjacent amino acid positions, a dependency between the two is entirely plausible, but we do not find any particular interest in these positions in the literature.

These results may seem counterintuitive given that ClaPTE is generically more conservative than BayesTraits. However, ClaPTE is more sensitive than BayesTraits in some regimes. Although there is no gold standard by which the correctness of these results can be rigorously assessed, associations between adjacent positions in the amino acid sequence are highly plausible, and the ability of ClaPTE to detect such associations in this alignment is a positive indicator that ClaPTE would detect associations with NA-274 if any such existed. FIC does appear quite sensitive, but the very low p -values it reports are consistent with a general failure of type I error control.

3.10. Results in protein clusters

Associations between character pairs within protein sequence alignments can be driven by numerous factors outside the scope of this manuscript [70–72]. Thus, a gold standard to judge position specific correlation predictions is difficult to achieve in protein sequence alignments. However, a reasonable negative control is available: pairs of characters drawn from different protein sequence alignments, where the two proteins do not interact in any biological sense, should be conditionally independent when ancestry is taken into account. Pairs of characters from the same protein sequence alignment may be substantially non-independent, and sensitivity to this non-independence is desirable, although this may not be useful as a positive control. Furthermore, the design of this study (excluding sequences from organisms that do not carry exactly one gene from each of the two clusters) is not optimized to assess sensitivity. Nonetheless it is important to avoid a method that achieves type-I error control (works properly on the false negative example) simply by sacrificing all sensitivity. Therefore, we report for three methods (ClaPTE, BayesTraits and Felsenstein's independent contrasts) the following statistics on alignments derived from two NCBI protein clusters, associated with the genes DnaJ and GuaB (see Section 2) which are not likely to interact: associated pairs ($p < 0.002$) in DnaJ, conditioned on the DnaJ and GuaB trees; associated pairs ($p < 0.002$) in GuaB, likewise conditioned on either tree; and associated pairs ($p < 0.002$) from different proteins, likewise conditioned on either tree. Counts and frequencies are given in Table 5; ClaPTE produces highly plausible results, where a modest number of additional associations are reported for pairs originating from the same protein family, but few associations (roughly consistent with type I error control at the $p < 0.002$ threshold) between positions in one alignment and positions in the other. Bayes Traits also gives plausible results, although the coupling within each alignment is slightly smaller than the coupling between alignments, and BayesTraits shows great sensitivity to which tree is used

(as would be expected). Felsenstein's independent contrasts produces, again, an implausibly large number of small p -values, and in this case does not seem to show sensitivity to the intra-sequence versus inter-sequence position pairs. For alignments of this size, such associations are likely to arise from functional coupling (*via* interaction with binding partners and the like) [73,74]; much larger alignments with tens of thousands of structurally diverse representatives would be required for applications such as protein folding prediction [75].

3.11. ClaPTE vs. BayesTraits

In the simulations reported here, BayesTraits shows two significant shortcomings in the context of testing for an association between characters. First, the p -value approximations from BayesTraits are not properly bounded under the null hypothesis, which is a major practical issue in the use of the approach. BayesTraits may show this property due to a lack of independence among the parameters of the likelihood model, effectively reducing the degrees of freedom, or to a breakdown of the chi-square approximation for likelihood differences much greater than 0. Second, BayesTraits loses power/sensitivity when the false-positive rate is forced to be very low. The non-independence of the parameters is the likely cause of this effect as well, by distorting the likelihood differences associated with rare events. This is unexpected, since likelihood-ratio approaches such as BayesTraits are generally expected to perform well in decision-making and related areas. On average, BayesTraits does make the best decisions when the two alternatives are balanced, that is, at a false positive rate on the order of 0.5 (Fig. 6, towards the upper-right-hand corner of each panel.) However, the theoretical guarantees regarding likelihood do not hold when the ground truth is more complex, or when the false-positive rate must be very low, as when a large multiple hypothesis correction is applied. Even so, BayesTraits may be more efficient than ClaPTE for small numbers of independent transformations (Fig. 7).

3.12. Caveats and future directions

First, because ClaPTE is intended for categorical outcomes other methods may be more suitable when the outcome is continuous (e.g. body mass, metabolic rate). These results in particular do not address the use of FIC for continuous characters. Although, a continuous character which shifts between distinct optima gives results resembling those for a categorical trait.

As shown in Table S1 (in the Supplement), ClaPTE suffers a considerable loss of power when one of the characters studied is ambiguously reconstructed. This could arise in cases where multiple character histories give similar parsimony scores or likelihood values (depending on method used), or in cases where the actual measurements are subject to uncertainty at individual tips. In particular, as is probably the case in the comparison shown on Table S1, uncertainty about the identity of the transmitting variant causes changes in the transmission status character to be falsely-reconstructed as deeply ancestral. Thus, all of the apparent intervals on the tree were lengthened, and the test for accelerated failure loses considerable power. For ClaPTE to be useful, the phenotype under study should be directly and uniquely available for each observation. In the example from Kumar et al. [4] and shown in Table S1, ClaPTE would be expected to be more powerful if HIV sequences were

available from each of the infected infants, enabling a much more precise determination of exactly where on the phylogeny the transmission event occurred.

Better handling of such ambiguities is one area where more advanced time to event methods may be useful [76]. These results show that given a multiple hypothesis correction and uncertainty in underlying phylogeny, such methods may be useful to adapt to the phylogenetic case. In protein sequences, alternative amino acid mutations could be treated as competing hazards; likewise, HIV escape variants are mutually exclusive competing hazards due to competition for host resources [77], as would be zoonosis out of an endangered population where the competing hazard is extinction of the host [78]. However, for these applications an additional technical challenge arises due to the need for a molecular clock [48]. Treatment regime changes [79] could lead to a declining hazard (that is, a fitness advantage that existed only for a period of time) resulting in an improper survival function. Finally, ClaPTE is most robust when the phylogenetic history of the outcome can be characterized with a minimum of recombination or gene transfer [80]. The presence of recombination corresponds to simulations scored using a different tree than the tree driving the simulations. A systematic study of the performance of ClaPTE across a much wider range of such recombinant cases, and across a wider range of parameters, is under way. These future results will be applied to questions of power analysis and study design, allowing investigators to incorporate both the level of diversity among the observations and the uncertainty in the estimates of inter-relatedness.

3.13. Summary

ClaPTE is suitable to detect associations between any pair of discrete/categorical characters. Character is a general term to refer to any feature of a biological observation. Suitable characters for ClaPTE include: environmental niche; host organism; host tissue compartment; predator or prey relationships; host cell-type affinities; drug exposures; the presence, absence or malfunction of an individual gene, protein, or network interaction; and substitutions, insertions or deletions at individual nucleotide or amino acid positions. ClaPTE is primarily designed to be computationally-efficient on categorical data, and robust to possible covariance in the characters [81]. ClaPTE is efficient in simulation experiments for a range of biologically-relevant values, and these results show a broad trend towards increasing efficiency (relative to other approaches) as the simulation experiments become more realistic, and less idealized (as shown in Figs. 3–9). Although the simulation results reported here cannot generalize to all cases or ranges of parameters, the simulation experiments are geared to detect a trend towards higher efficiency in more difficult problems [82,83], when the multiple hypothesis adjustment is large, the uncertainty in the covariance structure is high, or the covariance matrix among the observations is heterogeneous. Although significant additional work is required on this question, the “cross tree” comparisons provide a preliminary assessment on the performance of different methods in recombinant genomes, and ClaPTE suffers less degradation in performance compared to either BayesTraits or to FIC.

The simulations reported in the results show that ClaPTE is generally competitive to other methods on the following criteria: sensitivity (the “true positive” rate, also related to

statistical efficiency), stringency (the relationship between the “false positive” rate and the reported *p*-value), robustness (the degree to which performance suffers when there is uncertainty in the tree reconstruction/measurement of covariance) and computational efficiency. These criteria support the suitability of ClaPTE as a test for thousands or millions of possible associations on very large data sets of genotypes and phenotypes.

More broadly, two approaches to express association are considered here, trees and covariance matrices. Each tree implies a particular covariance matrix (structure). Not all covariance matrices are constructible from trees, but the tree-induced matrices can be quite diverse. Most importantly, a phylogenetic tree implies additional constraints than a two-dimensional covariance matrix. ClaPTE takes advantage of these constraints. To the extent that these constraints are reflected in the data and are not an artifact of tree construction, ClaPTE should outperform regression methods which utilize a covariance matrix instead of a tree. It is true that in some cases an incorrect tree may be reconstructed from the data; however, this does not always favor regression methods which utilize a covariance matrix only, since an incorrectly specified covariance structure can adversely impact performance even for such methods (as shown in Figs. 3, 4, 6 and 7).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors acknowledge that this material is based upon work supported by or in part by: NIGMS grant U01 092655 to the XGEN consortium; NSF agreement no. 0931642 to the Mathematical Biosciences Institute; NIH grant R00 HD05686 to Dr. Jesse Kwiek; and, the U.S. Army Research and Laboratory and Office under grant number W911NF-05-1-0271. This work was supported in part by an allocation of computing time from the Ohio Supercomputer Center. None of these funding agencies had any part in the research presented here, or in the preparation thereof. The authors would like to acknowledge the gracious contributions of Dr. Surender Kumar and Jessy Mates for the experimental work which provided structure to the simulation studies.

References

1. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. *Nat Methods*. 2011; 8(10):833–835. [PubMed: 21892150]
2. Pagel M, Meade A. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am Nat*. 2006; 167(6):808–825. [PubMed: 16685633]
3. Felsenstein J. Phylogenies and the comparative method. *Am Nat*. 1985
4. Kumar SB, Handelman SK, Voronkin I, Mwapasa V, Janies D, Rogerson SJ, Meshnick SR, Kwiek JJ. Different regions of HIV-1 subtype C env are associated with placental localization and in utero mother-to-child transmission. *J Virol*. 2011; 85(14):7142–7152. [PubMed: 21543508]
5. Winterberg KM, Luecke J, Bruegl AS, Reznikoff WS. Phenotypic screening of *Escherichia coli* K-12 Tn5 insertion libraries, using whole-genome oligonucleotide microarrays. *Appl Environ Microbiol*. 2005; 71(1):451–459. [PubMed: 15640221]
6. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*. 2006 2:2006 0008.
7. Strepp R, Scholz S, Kruse S, Speth V, Reski R. Plant nuclear gene knockout reveals a role in plastid division for the homolog of the bacterial cell division protein FtsZ, an ancestral tubulin. *Proc Natl Acad Sci U S A*. 1998; 95(8):4368–4373. [PubMed: 9539743]

8. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science*. 1994; 265(5181):2037–2048. [PubMed: 8091226]
9. Slatkin M. Spatial patterns in the distributions of polygenic characters. *J Theor Biol*. 1978; 70(2): 213–228. [PubMed: 633917]
10. Felsenstein J. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am J Hum Genet*. 1973; 25(5):471. [PubMed: 4741844]
11. Paradis E, Claude J. Analysis of comparative data using generalized estimating equations. *J Theor Biol*. 2002; 218(2):175–185. [PubMed: 12381290]
12. Lorch PD, Eadie JMA. Power of the concentrated changes test for correlated evolution. *Syst Biol*. 1999; 48(1):170–191. [PubMed: 12078639]
13. Habib F, Johnson AD, Bundschuh R, Janies D. Large scale genotype– phenotype correlation analysis based on phylogenetic trees. *Bioinformatics*. 2007; 23(7):785–788. [PubMed: 17267431]
14. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004; 20(2):289–290. [PubMed: 14734327]
15. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007; 23(19):2633–2635. [PubMed: 17586829]
16. Tuller T, Birin H, Gophna U, Kupiec M, Ruppin E. Reconstructing ancestral gene content by coevolution. *Genome Res*. 2010; 20(1):122–132. [PubMed: 19948819]
17. Lavin SR, Karasov WH, Ives AR, Middleton KM, Garland T Jr. Morpho-metrics of the avian small intestine compared with that of nonflying mammals: a phylogenetic approach. *Physiol Biochem Zool*. 2008; 81(5):526–550. [PubMed: 18754728]
18. Giannini NP, Goloboff PA. Delayed-response phylogenetic correlation: an optimization-based method to test covariation of continuous characters. *Evolution*. 2010; 64(7):1885–1898. [PubMed: 20100217]
19. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. GARD: a genetic algorithm for recombination detection. *Bioinformatics*. 2006; 22(24):3096–3098. [PubMed: 17110367]
20. Maddison WP, Slatkin M. Null models for the number of evolutionary steps in a character on a phylogenetic tree. *Evolution*. 1991; 45(5):1184–1197.
21. Slatkin M, Maddison WP. Detecting isolation by distance using phylogenies of genes. *Genetics*. 1990; 126(1):249–260. [PubMed: 2227384]
22. Maddison W, Slatkin M. Parsimony reconstructions of ancestral states do not depend on the relative distances between linearly-ordered character states. *Syst Zool*. 1990; 39(2):175–178.
23. Slatkin M, Maddison WP. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics*. 1989; 123(3):603–613. [PubMed: 2599370]
24. Pond SLK, Frost SDW. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. *Mol Biol Evol*. 2005; 22(3):478–485. [PubMed: 15509724]
25. Jiao W, Vembu S, Deshwar AG, Stein L, Morris Q. Inferring clonal evolution of tumors from single nucleotide somatic mutations. *BMC Bioinf*. 2014; 15:35.
26. Sirakoulis G, Karafyllidis I, Mizas C, Mardiris V, Thanailakis A, Tsalides P. A cellular automaton model for the study of DNA sequence evolution. *Comput Biol Med*. 2003; 33(5):439–453. [PubMed: 12860467]
27. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science*. 1999; 286(5438):295–299. [PubMed: 10514373]
28. Pazos F, Ranea JAG, Juan D, Sternberg MJE. Assessing protein co-evolution in the context of the tree of life assists in the prediction of the interactome. *J Mol Biol*. 2005; 352(4):1002–1015. [PubMed: 16139301]
29. Pazos F, Valencia A. Protein co-evolution, co-adaptation and interactions. *EMBO J*. 2008; 27(20): 2648–2655. [PubMed: 18818697]
30. Pazos F, Helmer-Citterich M, Ausiello G, Valencia A. Correlated mutations contain information about protein-protein interaction 1. *J Mol Biol*. 1997; 271(4):511–523. [PubMed: 9281423]
31. Burger L, van Nimwegen E. Disentangling direct from indirect co-evolution of residues in protein alignments. *Plos Comput Biol*. 2010; 6:1.

32. Poon AFY, Swenson LC, Dong WWY, Deng W, Pond SLK, Brumme ZL, Mullins JI, Richman DD, Harrigan PR, Frost SDW. Phylogenetic analysis of population-based and deep sequencing data to identify coevolving sites in the nef gene of HIV-1. *Mol Biol Evol.* 2010; 27(4):819–832. [PubMed: 19955476]
33. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38(8): 904–909. [PubMed: 16862161]
34. Halder I, Shriver M, Thomas M, Fernandez JR, Frudakis T. A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Hum Mutat.* 2008; 29(5):648–658. [PubMed: 18286470]
35. Ramani AK, Marcotte EM. Exploiting the co-evolution of interacting proteins to discover interaction specificity. *J Mol Biol.* 2003; 327(1):273–284. [PubMed: 12614624]
36. Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. *Am J Hum Genet.* 2010; 86(1):6–22. [PubMed: 20074509]
37. Enoch MA, Shen PH, Xu K, Hodgkinson C, Goldman D. Using ancestry-informative markers to define populations and detect population stratification. *J Psychopharmacol.* 2006; 20(4 Suppl):19–26. [PubMed: 16785266]
38. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al. Finding the missing heritability of complex diseases. *Nature.* 2009; 461(7265):747–753. [PubMed: 19812666]
39. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 2010; 42(7):565–569. [PubMed: 20562875]
40. Higuchi R, Krummel B, Saiki RK. A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res.* 1988; 16(15):7351–7367. [PubMed: 3045756]
41. Renaud C, Boudreault AA, Kuypers J, Lofy KH, Corey L, Boeckh MJ, Englund JA. H275Y mutant pandemic (H1N1) 2009 virus in immunocompromised patients (quiz 765). *Emerg Infect Dis.* 2011; 17(4):653–660. [PubMed: 21470455]
42. Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, Ostell J, Lipman D. The influenza virus resource at the National Center for Biotechnology Information. *J Virol.* 2008; 82(2):596–601. [PubMed: 17942553]
43. Pinilla LT, Holder BP, Abed Y, Boivin G, Beauchemin CA. The H275Y neuraminidase mutation of the pandemic A/H1N1 influenza virus lengthens the eclipse phase and reduces viral output of infected cells, potentially compromising fitness in ferrets. *J Virol.* 2012; 86(19):10651–10660. [PubMed: 22837199]
44. Gong LI, Suchard MA, Bloom JD. Stability-mediated epistasis constrains the evolution of an influenza protein. *Elife.* 2013; 2:e00631. [PubMed: 23682315]
45. Matrosovich M, Matrosovich T, Carr J, Roberts NA, Klenk HD. Overexpression of the alpha-2,6-sialyltransferase in MDCK cells increases influenza virus sensitivity to neuraminidase inhibitors. *J Virol.* 2003; 77(15):8418–8425. [PubMed: 12857911]
46. Bloom JD, Gong LI, Baltimore D. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science.* 2010; 328(5983):1272–1275. [PubMed: 20522774]
47. Klimke W, Agarwala R, Badretin A, Chetvernin S, Ciufo S, Fedorov B, Kiryutin B, O'Neill K, Resch W, Resenchuk S. The national center for biotechnology information's protein clusters database. *Nucleic Acids Res.* 2009; 37(Database issue):D216–223. [PubMed: 18940865]
48. Sanderson MJ. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics.* 2003; 19(2):301–302. [PubMed: 12538260]
49. Durrett, R. *Essentials of Stochastic Processes.* Springer; 2012.
50. Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data.* Wiley; New York, NY: 1980.
51. Wheeler W. History AMoN: dynamic homology and phylogenetic systematics: a unified approach using POY. *Am Museum Nat Hist N Y.* 2006

52. Varón A, Vinh LS, Wheeler WC. POY version 4: phylogenetic analysis using dynamic homologies. *Cladistics*. 2010; 26(1):72–85.
53. Pagel M, Meade A, Barker D. Bayesian estimation of ancestral character states on phylogenies. *Syst Biol*. 2004; 53(5):673–684. [PubMed: 15545248]
54. Pagel, M.; Meade, A. BayesTraits Univ Reading. 2007. <http://www.evolution.rdg.ac.uk/BayesTraits.html>
55. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81(3):559–575. [PubMed: 17701901]
56. Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. GEIGER: investigating evolutionary radiations. *Bioinformatics*. 2008; 24(1):129–131. [PubMed: 18006550]
57. Onafuwa-Nuga A, Telesnitsky A. The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol Mol Biol Rev*. 2009; 73(3):451–480. [PubMed: 19721086]
58. Zwickl DJ, Hillis DM. Increased taxon sampling greatly reduces phylogenetic error. *Syst Biol*. 2002; 51(4):588–598. [PubMed: 12228001]
59. Katoh K, Frith MC. Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics*. 2012; 28(23):3144–3146. [PubMed: 23023983]
60. Taylor G. Sialidases: structures, biological significance and therapeutic potential. *Curr Opin Struct Biol*. 1996; 6(6):830–837. [PubMed: 8994884]
61. Stamatakis A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006; 22(21):2688–2690. [PubMed: 16928733]
62. Goloboff PA, Pol D. On divide-and-conquer strategies for parsimony analysis of large data sets: Rec-I-DCM3 versus TNT. *Syst Biol*. 2007; 56(3):485–495. [PubMed: 17562472]
63. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. *J Comput Biol*. 2000; 7(3-4):601–620. [PubMed: 11108481]
64. Chowdhury, S.; Nibbe, R.; Chance, M.; Koyutürk, M. Subnetwork State Functions Define Dysregulated Subnetworks in Cancer. Springer; 2010. p. 80-95. In: 2010
65. Eddy JA, Hood L, Price ND, Geman D. Identifying tightly regulated and variably expressed networks by differential rank conservation (DIRAC). *PLoS Comput Biol*. 2010; 6(5):e1000792. [PubMed: 20523739]
66. Jia P, Zheng S, Long J, Zheng W, Zhao Z. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. *Bioinformatics*. 2011; 27(1):95–102. [PubMed: 21045073]
67. Anitha P, Anbarasu A, Ramaiah S. Computational gene network study on antibiotic resistance genes of *Acinetobacter baumannii*. *Comput Biol Med*. 2014; 48:17–27. [PubMed: 24637144]
68. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol*. 2006; 4(3):e72. [PubMed: 16494531]
69. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, Holmes EC. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*. 2004; 303(5656):327–332. [PubMed: 14726583]
70. Horner DS, Pirovano W, Pesole G. Correlated substitution analysis and the prediction of amino acid structural contacts. *Brief Bioinform*. 2008; 9(1):46–56. [PubMed: 18000015]
71. Halperin I, Wolfson H, Nussinov R. Correlated mutations: advances and limitations. A study on fusion proteins and on the Cohesin–Dockerin families. *Proteins: Struct Funct Bioinf*. 2006; 63(4): 832–845.
72. Lavanya P, Ramaiah S, Anbarasu A. Computational analysis of N-Hcdots, three dots, centeredpi interactions and its impact on the structural stability of beta-lactamases. *Comput Biol Med*. 2014; 46:22–28. [PubMed: 24529202]
73. Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, Sander C. Protein 3D structure computed from evolutionary sequence variation. *PLoS One*. 2011; 6(12):e28766. [PubMed: 22163331]

74. Nath A, Chaube R, Subbiah K. An insight into the molecular basis for convergent evolution in fish antifreeze Proteins. *Comput Biol Med.* 2013; 43(7):817–821. [PubMed: 23746722]
75. Kajan L, Hopf TA, Kalas M, Marks DS, Rost B. FreeContact: fast and free software for protein contact prediction from residue co-evolution. *BMC Bioinform.* 2014; 15:85.
76. Snapinn SM. Survival analysis with uncertain endpoints. *Biometrics.* 1998; 54(1):209–218. [PubMed: 9574966]
77. Ball CL, Gilchrist MA, Coombs D. Modeling within-host evolution of HIV: mutation, competition and strain replacement. *Bull Math Biol.* 2007; 69(7):2361–2385. [PubMed: 17554585]
78. Thompson RC, Kutz SJ, Smith A. Parasite zoonoses and wildlife: emerging issues. *Int J Environ Res Public Health.* 2009; 6(2):678–693. [PubMed: 19440409]
79. Kitchen CM, Philpott S, Burger H, Weiser B, Anastos K, Suchard MA. Evolution of human immunodeficiency virus type 1 coreceptor usage during antiretroviral therapy: a Bayesian approach. *J Virol.* 2004; 78(20):11296–11302. [PubMed: 15452249]
80. Jain R, Rivera MC, Lake JA. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A.* 1999; 96(7):3801. [PubMed: 10097118]
81. Hennig W. Phylogenetic systematics. *Annu Rev Entomol.* 1965; 10(1):97–116.
82. Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst Biol.* 1978; 27(4):401–410.
83. Sober E. Parsimony in systematics: philosophical issues. *Annu Rev Ecol Syst.* 1983; 14:335–357.

Abbreviations

ClaPTE	Cladogram with Path to Event
Fig. or Figs	Figure or Figures
Eq. or Eqs	Equation or Equations
PCA	principal components analysis
R.O.C	receiver operating characteristic
FIC	Felsenstein's independent contrasts

Fig. 1.

Illustration of cross-tree comparisons. (A and B) show reconstructions of two characters **X** and **Y** from a simulation run carried out on the GP120 tree, but in panel B reconstructed on the promoter tree instead. The color coding is black for state (**X** in **r**, **Y** in **c**), blue for state (**X** in **R**, **Y** in **c**), red for state (**X** in **r**, **Y** in **C**) and magenta for state (**X** in **R**, **Y** in **C**). The alphanumeric at the leaves identify individual amplicons; the same amplicons are present in each tree, and each tip is in the same state for both trees. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 2.

Illustrations of simulations used. (A) Shows the GP120 tree, and (B) shows the promoter tree, with four digit codes identifying mother/infant pairs. (C) Shows the GuaB tree, (D) shows the DnaJ tree. Four panels show six-node cartoon reductions of the (E) “branch heavy balanced”, (F) “leaf heavy balanced”, (G) “branch heavy unbalanced” and (H) “leaf heavy unbalanced” trees; the actual trees used are much larger containing 132 tips (see Supplement). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 3.

Quantile/quantile plots to assess type I error (null-hypothesis rejection) rates in different simulations for uncoupled characters on simple trees. Each panel shows the frequency with which different methods report a p -value below a given threshold, with both the p -value (horizontal axis) and the frequency (vertical axis) on log scale. Panels are labeled with cartoons (see Fig. 1) indicating which trees were used. Any rise above the dotted line indicates a method that cannot be used for null hypothesis rejection at the advertised stringency. A method below the dotted line is overly-conservative. The first two panels shows the result for (A) balanced leaf heavy, (B) unbalanced leaf heavy, while the next two panels show results for (C) balanced leaf heavy scored on branch heavy and (D) unbalanced leaf heavy scored on branch heavy. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4.

Quantile–quantile plots to assess type I error (null-hypothesis rejection) rates in different simulations for uncoupled characters on HIV trees. As Fig. 3, except with text rather than cartoon labels, showing results for (A) GP120; (B) promoter phylogenetic trees; (C) GP120 scored on promoter and (D) promoter scored on GP120.

Fig. 5.

Quantile-quantile plots to assess type I error (null-hypothesis rejection) rates in different simulations for uncoupled characters on bacterial protein trees. As Fig. 3, except with text rather than cartoon labels, showing results for (A) GuaB; (B) DnaJ; (C) GuaB scored on DnaJ and (D) DnaJ scored on GuaB.

Fig. 6.

Receiver operating characteristic (R.O.C.) curves to assess type II error (sensitivity) on simple trees. Each panel shows the sensitivity/recall of different methods towards coupled position pairs ($\mathbf{X}, \mathbf{Y}'$, vertical axis shows “true positive rate”) as a function of declining specificity (more pairs $\mathbf{X}, \mathbf{Y}$; horizontal axis shows “false positive rate”). Panels are labeled with cartoons (see Fig. 1). These R.O.C. curves are presented with the horizontal axis in log scale. A perfect predictor would go straight to the upper-left-hand-corner; a random prediction would follow the dotted line. The first two panels shows the result for (A) balanced leaf heavy, (B) unbalanced leaf heavy, while the next two panels show results for (C) balanced leaf heavy scored on branch heavy and (D) unbalanced leaf heavy scored on branch heavy. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 7.

Receiver operating characteristic (R.O.C.) curves to assess type II error (sensitivity) on HIV trees. As Fig. 6., except with text rather than cartoon labels, showing results for (A) GP120; (B) promoter phylogenetic trees; (C) GP120 scored on promoter and (D) promoter scored on GP120.

Fig. 8. Receiver operating characteristic (R.O.C.) curves to assess type II error (sensitivity) on bacterial protein trees. As Fig. 6., except with text rather than cartoon labels, showing results for (A) GuaB; (B) DnaJ; (C) GuaB scored on DnaJ and (D) DnaJ scored on GuaB.

Fig. 9.

Sensitivity among methods to entropy differences for uncoupled characters. Each panel shows the bias of different methods towards uncoupled position pairs with high differences in entropy (vertical axis) or low differences in entropy (horizontal axis). Panels are labeled with cartoons (see Fig. 1) or tree names, and values $S_Y - S_X$ arising in null simulations (see Table 3). As with a normal R.O.C. curve, the path along the diagonal indicates a random prediction (neither axis is on log scale). In this case, a random/unbiased prediction is desirable because there are in no sense any “true positive” pairs to detect. For the sake of brevity, only (A) balanced leaf heavy (B) promoter and (C) balanced leaf heavy scored on steam heavy phylogenetic trees are shown; differences among methods on other trees are less pronounced. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

Simulation parameters.

Simulation type	θ_F	θ_G , when $X=c$	θ_G , when $X=C$
Uncoupled Y	$\theta_{Ff}=\theta_{Fb}=5$	$\theta_{Gf}=\theta_{Gb}=25$	$\theta_{Gf}=\theta_{Gb}=25$
Coupled Y'	$\theta_{Ff}=\theta_{Fb}=5$	$\theta_{Gf}=10, \theta_{Gb}=250$	$\theta_{Gf}=250, \theta_{Gb}=10$

Each cell gives the evolutionary rate(s) used in different simulation experiments.

Table 2

Summary of methods.

Method	Col I: avg. false positive rate at nominal $p < 0.002$	Col II: avg. true positive rate at false positive < 0.002	Col III: avg. seconds per taxon ²	Col IV: Primary Reference (citations as of Mar 10, 2014)
ClaPTE	0.0025/0.0033	0.75/ 0.86	7.7E-8	n/a
BayesTraits	0.0244/n.a.	0.54/n.a.	4.3E-4	Pagel et. al. [2] (328)
FIC	0.0164/0.201	0.87/0.12	4.2E-7	Felsenstein [3] (5179)
EMMAX	0.0026/0.048	0.77/0.50	1.8E-6	Lippert et. al. [1] (75)

Each column gives a different summary figure of merit for each of the four methods described in this report. The values in bold and underlined are the best figures of merit in each category. Col I: average stringency, first across four tests on symmetrical trees (corresponding to all panels of Fig. 3), followed by across four tests on HIV trees (corresponding to all panels of Fig. 4). Col II: average sensitivity across first the symmetrical trees (corresponding to all panels of Fig. 6), followed by across four tests on HIV trees (corresponding to all panels of Fig. 7). Col III: estimated computational complexity, in second per observation² for each method, corresponding to Table 4. Col IV: Gives primary references, with citation estimates as reported by google scholar.

Table 3

Number of simulated character pairs used, binned by differences in entropy.

$S_Y - S_X$	Balanced Tree (Figs. 1C, 2A and C, 4A and C, 6A and C)		Unbalanced tree (Figs. 1E, 2B and D, 4B and D)		GP120 tree (Figs. 1A, 3A and C, 5A and C)		Promoter tree (Figs. 1B, 3B and D, 5B and D, 6B)	
	TP	TN	TP	TN	TP	TN	TP	TN
-0.06 to -0.04	0		0		1	849	3	1,627
-0.04 to -0.02	10	3644	37	9,695	94	2,596	112	4,049
-0.02 to 0	513	5138	1368	30,288	1351	7,301	1476	12,005
exactly 0	175	709	434	3,031	263	452	248	595
0 to 0.02	340	4497	1278	37,630	1255	12,723	1236	14,421
0.02 to 0.04	5	4740	41	21,368	125	16,645	77	14,883
0.04 to 0.06	0		2	14,959	5	20,978	3	18,633

Each column gives a tree used for simulations, as well as the figure panels where those simulations are described. Each cell gives the number of pairs of characters used in simulations for the corresponding entropy bin; TP (true positive) columns are pairs of coupled position pairs $\mathbf{X}, \mathbf{Y}'$, while TN (true negative) columns are uncoupled position pairs $\mathbf{X}, \mathbf{Y}$.

Table 4

Empirical computational complexity, in s.

Method	132 Taxa, balanced (s)	132 Taxa, unbalanced (s)	304 Taxa, biological (s)
ClaPTE	0.001	0.002	0.004
BayesTraits	10	7	28
FIC	0.009	0.009	0.01
EMMAX	0.03	0.03	0.19

Each cell gives the number of seconds (approximately two billion CPU cycles on modern hardware) each method requires to test for an association between two characters, for each of three trees—the balanced 132 taxa tree (see Fig. 2E), the unbalanced 132 taxa tree (see Fig. 2G), and the tree derived from HIV GP120 (see Fig. 1A).

Table 5

Results on GuaB and DnaJ protein clusters.

Position pairs, $p < 0.002$	GuaB $\times$ GuaB		DnaJ $\times$ DnaJ		GuaB $\times$ DnaJ	
	GuaB	DnaJ	GuaB	DnaJ	GuaB	DnaJ
Total	7750	7750	3240	3240	10,125	10,125
BT (fraction)	29 (0.004)	61 (0.008)	4 (0.001)	86 (0.027)	14 (0.001)	144 (0.014)
FIC (fraction)	2050 (0.265)	1690 (0.218)	671 (0.207)	640 (0.198)	2543 (0.251)	2068 (0.204)
ClapTE (fraction)	234 (0.03)	47 (0.006)	63 (0.019)	17 (0.005)	13 (0.001)	36 (0.004)

Each cell shows counts of pairs of positions from NIH PRK05567 (GuaB), NIH PRK10767 (DnaJ), or pairs of positions straddling the two alignments, which different methods report as associated at $p < 0.002$. Only a fraction of position pairs were considered and each amino acid column was converted to two possible states (see Supplement). ClapTE gives roughly correct type-I error control when comparing positions between the two alignments (an association would be unlikely), but detects a signal for pairs of positions from within the same sequence (a complete lack of associations would be implausible).