

RESEARCH ARTICLE

Accelerating Bayesian inference of dependency between mixed-type biological traits

Zhenyu Zhang¹, Akihiko Nishimura², Nidia S. Trovão³, Joshua L. Cherry^{3,4}, Andrew J. Holbrook¹, Xiang Ji⁵, Philippe Lemey⁶, Marc A. Suchard^{1,7,8*}

1 Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, California, United States of America, **2** Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, United States of America, **3** Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, Maryland, United States of America, **4** National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, United States of America, **5** Department of Mathematics, Tulane University, New Orleans, Louisiana, United States of America, **6** Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium, **7** Department of Biomathematics, University of California Los Angeles, Los Angeles, California, United States of America, **8** Department of Human Genetics, University of California Los Angeles, Los Angeles, California, United States of America

* msuchard@ucla.edu



OPEN ACCESS

Citation: Zhang Z, Nishimura A, Trovão NS, Cherry JL, Holbrook AJ, Ji X, et al. (2023) Accelerating Bayesian inference of dependency between mixed-type biological traits. *PLoS Comput Biol* 19(8): e1011419. <https://doi.org/10.1371/journal.pcbi.1011419>

Editor: Thierry Chekouo, University of Minnesota Twin Cities; University of Minnesota Twin Cities, UNITED STATES

Received: July 20, 2022

Accepted: August 9, 2023

Published: August 28, 2023

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: We implement our algorithms within BEAST software and provide the data sets and instructions at https://github.com/suchard-group/hzz_data_supplementary.

Funding: ZZ, PL and MAS are partially supported by National Institutes of Health grant R01 AI153044. MAS and PL acknowledge support from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement no.

Abstract

Inferring dependencies between mixed-type biological traits while accounting for evolutionary relationships between specimens is of great scientific interest yet remains infeasible when trait and specimen counts grow large. The state-of-the-art approach uses a phylogenetic multivariate probit model to accommodate binary and continuous traits via a latent variable framework, and utilizes an efficient bouncy particle sampler (BPS) to tackle the computational bottleneck—integrating many latent variables from a high-dimensional truncated normal distribution. This approach breaks down as the number of specimens grows and fails to reliably characterize conditional dependencies between traits. Here, we propose an inference pipeline for phylogenetic probit models that greatly outperforms BPS. The novelty lies in 1) a combination of the recent Zigzag Hamiltonian Monte Carlo (Zigzag-HMC) with linear-time gradient evaluations and 2) a joint sampling scheme for highly correlated latent variables and correlation matrix elements. In an application exploring HIV-1 evolution from 535 viruses, the inference requires joint sampling from an 11,235-dimensional truncated normal and a 24-dimensional covariance matrix. Our method yields a 5-fold speedup compared to BPS and makes it possible to learn partial correlations between candidate viral mutations and virulence. Computational speedup now enables us to tackle even larger problems: we study the evolution of influenza H1N1 glycosylations on around 900 viruses. For broader applicability, we extend the phylogenetic probit model to incorporate categorical traits, and demonstrate its use to study *Aquilegia* flower and pollinator co-evolution.

725422 - ReservoirDOCS) and from the Wellcome Trust through project 206298/Z/17/Z (The Artic Network). JLC is supported by the intramural research program of the National Library of Medicine, National Institutes of Health. AH is supported by NIH grant K25 AI153816. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

We aim to learn the relationships between different biological features, or traits, observed in related specimens that have evolved together over time. This is of great scientific interest because by identifying how different traits influence each other, we gain insights into the mechanisms underlying important biological processes. Learning the relationships between traits across a large number of specimens is computationally challenging, particularly when traits have mixed-type values (continuous and discrete). The previous best approach utilizing a method called bouncy particle sampler (BPS) struggles with increasing specimen and trait counts, resulting in unreliable estimates of trait dependencies. We develop a more efficient approach that largely outperforms BPS, reducing the runtime of our large-scale applications from weeks to days. We apply our method to study the evolution of HIV and influenza viruses, as well as flower and pollinator co-evolution. Our work provides an efficient yet general way to understand the connections between mixed-type traits, offering valuable insights into the evolution of complex biological systems.

Introduction

An essential goal in evolutionary biology is to understand the across-trait covariation observed within biological samples, or *taxa*, ranging from plants and animals to microorganisms and pathogens such as human immunodeficiency virus (HIV) and influenza. This task is difficult because taxa are implicitly correlated through their shared evolutionary history often described with a reconstructed phylogenetic tree. Here, tree tips correspond to the taxa themselves, and internal nodes are their unobserved ancestors. Inferring across-trait covariation requires a highly structured model that can explicitly describe the tree structure and adjust for across-taxa covariation. Phylogenetic models do exactly this but are computationally challenging because one must integrate out unobserved ancestor traits while accounting for uncertainties arising from tree estimation. The computational burden increases when taxon and trait counts grow large and becomes worse when traits include continuous and discrete quantities. Cybis et al. develop the first phylogenetic method that can assess across-trait covariation while controlling for a large, unknown evolutionary tree with hundreds of tips [1]. To jointly model mixed-type traits, this approach assumes discrete traits arise from continuously valued latent variables that follow a Brownian diffusion along the tree [2]. Assuming latent processes is a common strategy for modeling mixed-type data and it finds uses across various fields [3–7]. Subsequent work by [8] solves an essential identifiability issue in [1] by adding specific constraints on the diffusion covariance. The resulting model in particular generalizes the multivariate probit model [9]. The most important contribution of [8], however, is an efficient inference scheme that achieves order-of-magnitudes efficiency gains over [1]. In this work, we significantly advance performance compared to [8] to solve even larger problems.

Here is an intuition on why our new inference scheme, to be formally introduced in Methods section, outperforms the one by [8]. For N taxa each with P continuous or binary traits, Bayesian inference for the phylogenetic probit model involves repeatedly sampling latent variables \mathbf{X} from their conditional posterior, an $(N \times P)$ -dimensional truncated normal distribution. The $(N \times P)$ size of the truncated normal distribution results from having one latent variable for each taxon and each trait. For this task, [8] develop a bouncy particle sampler (BPS) [10] combined with an efficient dynamic programming approach that speeds up the most expensive step in the BPS implementation. Their approach, however, fails to address another source of computational inefficiency in posterior inference under the phylogenetic

probit model—a high degree of correlation between \mathbf{X} and \mathbf{C} . [8] use a separate Hamiltonian Monte Carlo sampler [11, HMC] to sample \mathbf{C} and update the two sets of parameters alternately within a random-scan Gibbs scheme [12]. The phylogenetic probit model assumes \mathbf{X} to follow a multivariate Gaussian distribution whose covariance matrix incorporates \mathbf{C} . By the model's very design, therefore, the values in \mathbf{C} influence the strength and direction of the correlations between elements of \mathbf{X} . This correlation between the two parameters slows down convergence and mixing of the Gibbs scheme as each update of \mathbf{X} or \mathbf{C} is strongly influenced by the current value of the other parameter. To address this issue, our present solution utilizes a state-of-the-art Markov chain Monte Carlo (MCMC) method called Zigzag-HMC [13]. Unlike BPS, this method allows a joint update of \mathbf{X} and \mathbf{C} through differential operator splitting [13, 14] that generalizes the previously proposed split HMC framework based on Hamiltonian splitting [11, 15]. Zigzag-HMC can further take advantage of the same $\mathcal{O}(N)$ gradient evaluation strategy developed by [8].

Our sampling scheme greatly improves the mixing of elements in \mathbf{C} and thus provides a reliable estimate of the across-trait partial correlation matrix \mathbf{R} , the inverse of the correlation matrix normalized to have unit diagonals. The partial correlation between two traits quantifies their conditional dependence that accounts for, and hence removes confounding by, the effects of other traits in the model. Use of partial correlations thus allow us to gain insight into potential causal pathways and help guide further research into underlying biological mechanisms.

We apply our methodology to three real-world examples. First, we re-evaluate the HIV evolution application in [8] and identify HIV-1 *gag* immune-escape mutations linked with virulence through strong conditional dependence relationships. Our findings closely match with the experimental literature and indicate a general pattern in the immune escape mechanism of HIV. Second, we examine the influenza H1N1 glycosylation pattern across different hosts and detect strong conditional dependencies between glycosylation sites closely related to host switching. Finally, we investigate how floral traits of *Aquilegia* flower attract different pollinators, for which we generalize the phylogenetic probit model to accommodate a categorical pollinator trait.

Methods

Mixed-type trait evolution

We describe biological trait evolution with the phylogenetic multivariate probit model following [8] and extend it to unordered categorical traits as in [1]. While we do not consider ordered categorical traits in this work and leave it to future work to support such traits, the mapping of latent variables in this case can also be found in [1]. We either know the phylogenetic tree \mathcal{F} *a priori* or infer it from a molecular sequence alignment \mathbf{S} [16]. In our two large-scale HIV and influenza applications (Results section) with available sequence data, we use a continuous-time Markov chain evolutionary model [17] to construct $p(\mathbf{S}|\mathcal{F})$ and so infer \mathcal{F} simultaneously. We refer interested readers to [16] for more details on tree sampling. When investigating the efficiency gain of our method over [8], we utilize a fixed tree for a more direct comparison and also to reduce the overall run-time. For our third application on flower and pollinator co-evolution, we adopt the same fixed tree as in [18].

Consider N taxa on a tree $\mathcal{F} = (\mathbb{V}, \mathbf{b})$ that is a directed, bifurcating acyclic graph. The node set \mathbb{V} of size $2N - 1$ contains N tip nodes, $N - 2$ internal nodes and one root node. The branch lengths $\mathbf{b} = (b_1, \dots, b_{2N-2})$ denote the child-parent distance in real time. We observe P mixed-type traits for each taxon. The trait data $\mathbf{Y} = \{y_{ij}\} = (\mathbf{Y}^{\text{cont}}, \mathbf{Y}^{\text{disc}})$ partition as \mathbf{Y}^{cont} , an $N \times P_{\text{cont}}$ matrix of continuous traits and \mathbf{Y}^{disc} , an $N \times P_{\text{disc}}$ matrix of discrete ones. We associate with each trait a latent variable $x_{ij} \in \mathbb{R}$, if the j -th trait is continuous or binary, and a $(m_j - 1)$ -

dimensional latent vector $\mathbf{X}_{ij} = \{x_{ij,k}\} \in \mathbb{R}^{m_j-1}$, if the trait is categorical, where m_j denotes the number of categorical classes. Continuous traits y_{ij} can be seen as latent variables that are directly observed so $x_{ij} = y_{ij}$. To relate latent variables to observed discrete traits, we assume a threshold model for binary traits and a choice model for traits with more than two classes. For a binary trait y_{ij} ,

$$y_{ij} = g_b(x_{ij}) = \begin{cases} 0 & \text{if } x_{ij} \leq 0, \\ 1, & \text{if } x_{ij} > 0. \end{cases} \tag{1}$$

For a categorical trait y_{ij} , the possible classes are $\{c_1, \dots, c_{m_j}\}$ with the reference class being c_1 . We have

$$y_{ij} = g_c(x_{ij,1}, \dots, x_{ij,m_j-1}) = \begin{cases} c_1 & \text{if } x_{ij,\max} \leq 0, \\ c_k & \text{if } m_j > 1 \text{ and } x_{ij,\max} = x_{ij,k-1} > 0, \end{cases} \tag{2}$$

where $x_{ij,\max} = \max(x_{ij,1}, \dots, x_{ij,m_j-1})$. This data augmentation strategy is a common choice to model categorical data [19].

After concatenating all the latent variables, for each node $i = 1, \dots, 2N - 1$ in \mathcal{F} we have P_{lat} -dimensional latent variable $\mathbf{X}_i \in \mathbb{R}^{P_{\text{lat}}}$ with $P_{\text{lat}} = P_{\text{cont}} + \sum_{j=P_{\text{cont}}+1}^{P_{\text{cont}}+P_{\text{disc}}} (m_j - 1)$. As a side note, for continuous y_{ij} the corresponding x_{ij} is observed, and so \mathbf{X}_i is actually a partially latent vector. Since in our applications only a small fraction of y_{ij} is continuous, we omit “partial” to ease the notation.

The latent variables follow a multivariate Brownian diffusion process along \mathcal{F} such that \mathbf{X}_i distributes as a multivariate normal (MVN)

$$\mathbf{X}_i \sim \mathcal{N}(\mathbf{X}_{\text{pa}(i)}, b_i \boldsymbol{\Omega}), i = 1, \dots, 2N - 2, \tag{3}$$

where $\mathbf{X}_{\text{pa}(i)}$ is the parent node value and the $P_{\text{lat}} \times P_{\text{lat}}$ covariance matrix $\boldsymbol{\Omega}$ describes the across-trait association. The intuition behind $b_i \boldsymbol{\Omega}$ is that the further away a child node is from its parent node (larger b_i), the bigger difference between their node values. Assuming a conjugate root prior $\mathbf{X}_{2N-1} \sim \mathcal{N}(\boldsymbol{\mu}_0, \omega^{-1} \boldsymbol{\Omega})$ with prior mean $\boldsymbol{\mu}_0$ and prior variance $\omega^{-1} \boldsymbol{\Omega}$, we can analytically integrate out latent variables on all internal nodes. Marginally, then, the $N \times P_{\text{lat}}$ tip latent variables \mathbf{X} have the matrix normal distribution

$$\mathbf{X} \sim \text{MTN}_{N \times P_{\text{lat}}}(\mathbf{M}, \boldsymbol{\Upsilon}, \boldsymbol{\Omega}), \tag{4}$$

where $\mathbf{M} = (\boldsymbol{\mu}_0, \dots, \boldsymbol{\mu}_0)^T$ is an $N \times P_{\text{lat}}$ mean matrix and the across-taxa covariance matrix $\boldsymbol{\Upsilon}$ equals $\mathbf{V}(\mathcal{F}) + \omega^{-1} \mathbf{J}$ [20]. The diffusion matrix $\mathbf{V}(\mathcal{F})$ is a function of branch lengths such that its diagonal elements represent the sum of branch lengths from a tip to the root, while the off-diagonal elements are the branch length from the root to the most recent common ancestor of two tips. The augmented likelihood of \mathbf{X} and \mathbf{Y} factorizes as

$$p(\mathbf{Y}, \mathbf{X} | \boldsymbol{\Upsilon}, \boldsymbol{\Omega}, \boldsymbol{\mu}_0, \omega) = p(\mathbf{Y} | \mathbf{X}) p(\mathbf{X} | \boldsymbol{\Upsilon}, \boldsymbol{\Omega}, \boldsymbol{\mu}_0, \omega), \tag{5}$$

where $p(\mathbf{Y} | \mathbf{X}) = 1$ if \mathbf{X} are consistent with \mathbf{Y} according to Eqs (1) and (2) and 0 otherwise. Following [8], we decompose $\boldsymbol{\Omega} = \mathbf{D} \mathbf{C} \mathbf{D}$ where \mathbf{C} is a $P_{\text{lat}} \times P_{\text{lat}}$ correlation matrix. The diagonal entries of \mathbf{C} are all equal to 1, while the off-diagonal entries lie in the range of $[-1, 1]$ and represent the correlations between pairs of latent variables and, hence, their corresponding traits. The $P_{\text{lat}} \times P_{\text{lat}}$ diagonal matrix $\mathbf{D} = \{\sigma_{ii}\}$ for $i = 1, \dots, P_{\text{lat}}$ contains the marginal standard deviation of each latent variable. Importantly, since discrete traits only inform the sign or ordering

of their underlying latent variables, certain elements of \mathbf{D} must be set as a fixed value to ensure that the model is parameter-identifiable [8]. Without loss of generality, we fix $\sigma_{ii} = 1$ for σ_{ii} corresponding to discrete traits. For continuous traits, the square of the corresponding element (σ_{ii}^2) multiplied by a branch length is the marginal variance for the Brownian diffusion process along that branch (Eq 3). In other words, this product reports the amount of trait variation that accumulates along a branch. [8] demonstrate the necessity of this DCD decomposition, which also allows a non-informative prior [21, LKJ] on \mathbf{C} . For goodness-of-fit of the phylogenetic probit model we refer interested readers to [8] where the explicit tree modeling leads to a significantly better fit.

A novel inference scheme

We sample from the joint posterior to learn the across-trait correlation \mathbf{C}

$$p(\mathbf{C}, \mathbf{D}, \mathbf{X}, \mathcal{F} | \mathbf{Y}, \mathbf{S}) \propto p(\mathbf{Y} | \mathbf{X}) \times p(\mathbf{X} | \mathbf{C}, \mathbf{D}, \mathcal{F}) \times p(\mathbf{C}, \mathbf{D}) \times p(\mathbf{S} | \mathcal{F}) \times p(\mathcal{F}), \tag{6}$$

where we drop the dependence on hyper-parameters $(\boldsymbol{\Upsilon}, \boldsymbol{\mu}_0, \omega)$ to ease notation. We fix $\boldsymbol{\mu}_0$ to be a P_{lat} -dimensional zero vector and ω to be 1. We then specify the priors $p(\mathbf{C}, \mathbf{D})$ and $p(\mathcal{F})$ as in [8] where $p(\mathcal{F})$ is a typical coalescent tree prior on \mathcal{F} [22] and $p(\mathbf{C}, \mathbf{D}) = p(\mathbf{C})p(\mathbf{D})$. We set independent log normal priors on \mathbf{D} diagonals that correspond to continuous traits. We assume an LKJ prior on the Cholesky factor of \mathbf{C} to ensure that \mathbf{C} and $\boldsymbol{\Omega}$ are positive definite and invertible. [8] use a random-scan Gibbs [12] scheme to alternately update \mathbf{X} , $\{\mathbf{C}, \mathbf{D}\}$ and \mathcal{F} from their full conditionals [16]. They sample \mathbf{X} from an NP_{lat} -dimensional truncated normal distribution with BPS and deploy the standard HMC based on Gaussian momentum [23] to update $\{\mathbf{C}, \mathbf{D}\}$. Instead, we simulate the joint Hamiltonian dynamics on $\{\mathbf{X}, \mathbf{C}, \mathbf{D}\}$ by combining novel Hamiltonian zigzag dynamics on \mathbf{X} [24] and traditional Hamiltonian dynamics on $\{\mathbf{C}, \mathbf{D}\}$. This strategy enables an efficient joint update of the two highly-correlated sets of parameters. The improved efficiency allow us to focus on the across-trait partial correlation matrix $\mathbf{R} = \{r_{ij}\}$. After collecting the MCMC samples of $\boldsymbol{\Omega}$, we obtain \mathbf{R} by the standard transformation [25]:

$$\boldsymbol{\Omega}^{-1} = \mathbf{P} = \{p_{ij}\}, \quad r_{ij} = -\frac{p_{ij}}{\sqrt{p_{ii}p_{jj}}}. \tag{7}$$

Since \mathbf{R} measures the linear relationship between pairs of variables after controlling for effects of all other variables in the model, \mathbf{R} usually lies in a more-constrained space than \mathbf{C} and is more difficult for the sampler to effectively explore its posterior distribution. We demonstrate the improved efficiency of our method in inferring \mathbf{R} in Results section. In the subsequent sections, we first describe how Zigzag-HMC samples \mathbf{X} from a truncated normal and then detail the joint update of $\{\mathbf{X}, \mathbf{C}, \mathbf{D}\}$.

Zigzag-HMC for truncated multivariate normals. We outline the main ideas behind HMC [11] before describing Zigzag-HMC as a version of HMC based on *Hamiltonian zigzag dynamics* [13, 24]. In order to sample a d -dimensional parameter $\mathbf{x} = (x_1, \dots, x_d)$ from the target distribution $\pi(\mathbf{x})$, HMC introduces an auxiliary *momentum* variable $\mathbf{p} = (p_1, \dots, p_d) \in \mathbb{R}^d$ and samples from the product density $\pi(\mathbf{x}, \mathbf{p}) = \pi(\mathbf{x})\pi(\mathbf{p})$ by numerically discretizing the Hamiltonian dynamics

$$\frac{d\mathbf{x}}{dt} = \nabla K(\mathbf{p}), \quad \frac{d\mathbf{p}}{dt} = -\nabla U(\mathbf{x}), \tag{8}$$

where $U(\mathbf{x}) = -\log \pi(\mathbf{x})$ and $K(\mathbf{p}) = -\log \pi(\mathbf{p})$ are the potential and kinetic energy. In each

HMC iteration, we first draw \mathbf{p} from its marginal distribution $\pi(\mathbf{p}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, a standard Gaussian and then approximate (8) from time $t = 0$ to $t = \tau$ by $L = \lfloor \tau/\epsilon \rfloor$ steps of the *leapfrog* update with step size ϵ [26]:

$$\mathbf{p} \leftarrow \mathbf{p} + \frac{\epsilon}{2} \nabla_{\mathbf{x}} \log \pi(\mathbf{x}), \quad \mathbf{x} \leftarrow \mathbf{x} + \epsilon \mathbf{p}, \quad \mathbf{p} \leftarrow \mathbf{p} + \frac{\epsilon}{2} \nabla_{\mathbf{x}} \log \pi(\mathbf{x}). \tag{9}$$

The end state is a valid *Metropolis* proposal that one accepts or rejects according to the standard acceptance probability formula [27, 28].

Zigzag-HMC differs from standard HMC insofar as it posits a Laplace momentum $\pi(\mathbf{p}) \propto \prod_i \exp(-|p_i|)$, $i = 1, \dots, d$. The Hamiltonian differential equations now become

$$\frac{d\mathbf{x}}{dt} = \text{sign}(\mathbf{p}), \quad \frac{d\mathbf{p}}{dt} = -\nabla U(\mathbf{x}), \tag{10}$$

and the velocity $\mathbf{v} := d\mathbf{x}/dt \in \{\pm 1\}^d$ depends only on the sign of \mathbf{p} and thus remains constant until one of p_i 's undergoes a sign change (an "event"). To understand how the Hamiltonian zigzag dynamics (10) evolve over time, one must investigate when such events happen. Before moving to the truncated MVN, we first review the event time calculation for a general $\pi(\mathbf{x})$ following [24]. Let $\tau^{(k)}$ be the k th event time and $(\mathbf{x}(\tau^{(0)}), \mathbf{v}(\tau^{(0)}), \mathbf{p}(\tau^{(0)}))$ is the initial state at time $\tau^{(0)}$. Between $\tau^{(k)}$ and $\tau^{(k+1)}$, \mathbf{x} follows a piecewise linear path and the dynamics evolve as

$$\mathbf{x}(\tau^{(k)} + t) = \mathbf{x}(\tau^{(k)}) + t\mathbf{v}(\tau^{(k)}), \quad \mathbf{v}(\tau^{(k)} + t) = \mathbf{v}(\tau^{(k)}), \quad t \in [0, \tau^{(k+1)} - \tau^{(k)}], \tag{11}$$

and

$$p_i(\tau^{(k)} + t) = p_i(\tau^{(k)}) - \int_0^t \partial_i U[\mathbf{x}(\tau^{(k)}) + s\mathbf{v}(\tau^{(k)})] ds \quad \text{for } i = 1, \dots, d. \tag{12}$$

Therefore we can derive the $(k + 1)$ th event time

$$\tau^{(k+1)} = \tau^{(k)} + \min_i t_i, \quad t_i = \min_{t>0} \left\{ p_i(\tau^{(k)}) - \int_0^t \partial_i U[\mathbf{x}(\tau^{(k)}) + s\mathbf{v}(\tau^{(k)})] ds \right\}, \tag{13}$$

and the dimension causing this event is $i^* = \text{argmin}_i t_i$. At the moment of $\tau^{(k+1)}$, the i^* th velocity component flips its sign

$$v_{i^*}(\tau^{(k+1)}) = -v_{i^*}(\tau^{(k)}), \quad v_j(\tau^{(k+1)}) = v_j(\tau^{(k)}) \text{ for } j \neq i^*. \tag{14}$$

Then the dynamics continue for the next interval $[\tau^{(k+1)}, \tau^{(k+2)})$.

We now consider simulating the Hamiltonian zigzag dynamics for a truncated MVN arising from the phylogenetic probit model.

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ subject to } \mathbf{x} \in \{\text{map}(\mathbf{x}) = \mathbf{y}\}, \tag{15}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and covariance matrix for the MVN and $\text{map}(\cdot)$ is the mapping from the vectorized latent variables \mathbf{x} to \mathbf{y} as in Eqs (1) and (2). In other words, \mathbf{y} is the NP -dimensional vectorized discrete data such that $\mathbf{x} \in \mathbb{R}^d$ for $d = NP_{\text{lat}}$. Since vectorizing the random variables under a matrix normal distribution (3) results in a MVN distribution, we have $\boldsymbol{\Sigma} = \boldsymbol{\Omega} \otimes \boldsymbol{\Upsilon}$ where \otimes denotes the Kronecker product. The mean vector $\boldsymbol{\mu}$ is N copies of the pre-specified root prior mean vector $\boldsymbol{\mu}_0$ concatenated together.

In the setting of Eq (15), we have $\nabla U(\mathbf{x}) = \boldsymbol{\Sigma}^{-1} \mathbf{x}$ whenever $\mathbf{x} \in \{\text{map}(\mathbf{x}) = \mathbf{y}\}$. Importantly, this structure allows us to simulate the Hamiltonian zigzag dynamics exactly and efficiently [24]. We handle the constraint $\text{map}(\mathbf{x}) = \mathbf{y}$ with a technique from [11] where the constraint

boundaries embody “hard walls” that the Hamiltonian zigzag dynamics “bounce” against upon impact. To distinguish different types of events, we define *gradient events* arising from solutions of Eq (13), *binary events* arising from hitting binary data boundaries and *categorical events* arising from hitting categorical data boundaries.

We first consider how to find the gradient event time. Starting from a state $(\mathbf{x}, \mathbf{v}, \mathbf{p})$, by plugging in $\nabla U(\mathbf{x}) = \Sigma^{-1}\mathbf{x}$ to Eq (13), we can calculate the gradient event time t_g by first solving d quadratic equations

$$\mathbf{p} = t\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) + \frac{t^2}{2}\Sigma^{-1}\mathbf{v}, \tag{16}$$

and then taking the minimum among all positive roots of Eq (16). When $\pi(\mathbf{x})$ is a truncated MVN arising from the phylogenetic probit model, we exploit the efficient gradient evaluation strategy in [8] to obtain $\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})$ and $\Sigma^{-1}\mathbf{v}$ without the notorious $\mathcal{O}(d^3)$ cost to invert Σ . In our application, $\boldsymbol{\mu}$ is a vector of all zeros since we set the root prior mean $\boldsymbol{\mu}_0$ to be all zero. If there is prior knowledge about $\boldsymbol{\mu}_0$, we can use another fixed value without increasing the computational cost.

Next, we focus on the binary and categorical events. We partition \mathbf{x} into two sets: $S_{\text{bin}} = \{x_i: x_i \text{ is for binary data}\}$ and $S_{\text{cat}} = \{x_i: x_i \text{ is for categorical data}\}$. Starting from a state $(\mathbf{x}, \mathbf{v}, \mathbf{p})$, a binary event happens at time t_b when the trajectory first reaches a binary boundary at dimension i_b

$$t_b = |x_{i_b}/v_{i_b}|, \quad i_b = \operatorname{argmin}_{i \in I_{\text{bin}}} |x_i/v_i| \quad \text{for } I_{\text{bin}} = \{i: x_i v_i < 0 \text{ and } x_i \in S_{\text{bin}}\}. \tag{17}$$

Here, we only need to check the dimensions satisfying $x_i v_i < 0$, i.e., those for which the trajectory is heading towards the boundary. At time t_b , the trajectory bounces against the binary boundary, and so the i_b th velocity and momentum element both undergo an instantaneous flip $v_{i_b} \leftarrow -v_{i_b}, p_{i_b} \leftarrow -p_{i_b}$, while other dimensions stay unchanged.

Finally, we turn to categorical events. Suppose that a categorical trait $y_j = c_k$ belongs to one of m possible classes, and x_1, x_2, \dots, x_{m-1} the underlying latent variables. Eq (2) specifies the boundary constraints. If $k = 1$, the $m - 1$ latent variables must be all negative, which poses the same constraint as if they were for $n - 1$ binary traits, therefore we can solve the event time using Eq (17). If $k > 1$, we must check when and which two dimensions first violate the order constraint $x_{k-1} = \max(x_1, \dots, x_{m-1}) > 0$. With the dynamics starting from $(\mathbf{x}, \mathbf{v}, \mathbf{p})$, the categorical event time t_c^j is given by

$$t_c^j = |(x_{k-1} - x_{i_c})/(v_{k-1} - v_{i_c})|, \quad i_c = \operatorname{argmin}_{i \in I_{\text{cat}}} |(x_{k-1} - x_i)/(v_{k-1} - v_i)|, \tag{18}$$

for $I_{\text{cat}} = \{i: v_{k-1} < v_i \text{ and } x_i \in S_{\text{cat}}\},$

when x_{i_c} reaches x_{k-1} and violates the constraint. To identify i_c we only need to check dimensions with $v_{k-1} < v_i$ where the distance $x_{k-1} - x_i$ is decreasing. At t_c^j , the two dimensions involved ($k - 1$ and i_c) bounce against each other such that $v_{k-1} \leftarrow -v_{k-1}, v_{i_c} \leftarrow -v_{i_c}, p_{k-1} \leftarrow -p_{k-1}, p_{i_c} \leftarrow -p_{i_c}$. Note t_c^j is for a single y_j and we need to consider all categorical data to find the actual categorical event time $t_c = \min_j t_c^j$.

We now present the dynamics simulation with all three event types included, starting from a state $(\mathbf{x}, \mathbf{v}, \mathbf{p})$ with $\mathbf{x} \in \{\text{map}(\mathbf{x}) = \mathbf{y}\}$:

1. Solve t_g, t_b, t_c using Eqs (16), (17) and (18) respectively.
2. Determine the actual (first) event time $t = \min\{t_g, t_b, t_c\}$ and update \mathbf{x} and \mathbf{p} as in Eqs (11) and (12) for a duration of t .

3. Make instantaneous velocity and momentum sign flips according to the rules of the actual event type, then go back to Step 1.

Based on the above discussion, Algorithm 1 describes one iteration of Zigzag-HMC on truncated MVNs where we simulate the Hamiltonian zigzag dynamic for a pre-specified duration t_{total} . For a truncated MVN arising from the phylogenetic probit model, the most computationally expensive step is the gradient evaluation in Line 3, where a matrix-vector multiplication by the precision matrix $\Phi = \Sigma^{-1}$ is involved. A matrix inversion to evaluate Φ directly is expensive since $\Phi = \Omega^{-1} \otimes \Upsilon^{-1}$ and computing Υ^{-1} has a cost of $\mathcal{O}(N^3)$. We adopt the dynamic programming strategy of [8] to reduce the cost of Line 3 from either $\mathcal{O}(N^2 P_{\text{lat}} + NP_{\text{lat}}^2)$ when \mathcal{F} is fixed, or $\mathcal{O}(N^3 + P_{\text{lat}}^3)$ when \mathcal{F} is random, to $\mathcal{O}(NP_{\text{lat}}^2)$. We refer interested readers to [8] for details on the dynamic programming strategy. In brief, this strategy avoids explicitly inverting Υ by recursively traversing the tree [20] to obtain N conditional densities that directly translate to the desired gradient φ_x .

Algorithm 1 Zigzag-HMC for multivariate truncated normal distributions

```

1: function HzzTMVN( $\mathbf{x}$ ,  $\mathbf{p}$ ,  $t_{\text{total}}$ )
2:    $\mathbf{v} \leftarrow \text{sign}(\mathbf{p})$ 
3:    $\varphi_x \leftarrow \Phi(\mathbf{x} - \boldsymbol{\mu})$ 
4:    $t_{\text{remain}} \leftarrow t_{\text{total}}$ 
5:   while  $t_{\text{remain}} > 0$  do
6:      $\mathbf{a} \leftarrow \varphi_v/2$ ,  $\mathbf{b} \leftarrow \varphi_x$ ,  $\mathbf{c} \leftarrow -\mathbf{p}$ 
7:      $t_g \leftarrow \min_i \{\text{minPositiveRoot}(a_i, b_i, c_i)\} \triangleright$  "minPositiveRoot"
       defined below
8:      $t_b \leftarrow \min_i x_i/v_i$ , for  $i$  with  $x_i v_i < 0$  and  $x_i \in S_{\text{bin}}$ 
        $\triangleright$  find binary boundary event time
9:     for  $j = 1, \dots, n_c$  do
10:       $t_c^j \leftarrow \min_i |(x_{k-1} - x_{i_c}) / (v_{k-1} - v_i)|$  for  $i$  with  $v_{k-1} < v_i$  and  $x_i \in S_{\text{cat}}$ 
11:    end for
12:     $t_c \leftarrow \min_j t_c^j$ 
        $\triangleright$  the actual event happens at time  $t$ 
13:     $t \leftarrow \min\{t_g, t_b, t_c, t_{\text{remain}}\}$ 
14:     $\mathbf{x} \leftarrow \mathbf{x} + t\mathbf{v}$ ,  $\mathbf{p} \leftarrow \mathbf{p} - t\varphi_x - t^2\varphi_v/2$ ,  $\varphi_x \leftarrow \varphi_x + t\varphi_v$ 
15:    if a gradient event happens at  $i_g$  then
16:       $v_{i_g} \leftarrow -v_{i_g}$ 
17:    else if a binary boundary event happens at  $i_b$  then
18:       $v_{i_b} \leftarrow -v_{i_b}$ ,  $p_{i_b} \leftarrow -p_{i_b}$ 
19:    else if a categorical boundary event happens at  $i_{c1}$ ,  $i_{c2}$  then
20:       $v_{i_{c1}} \leftarrow -v_{i_{c1}}$ ,  $v_{i_{c2}} \leftarrow -v_{i_{c2}}$ ,  $p_{i_{c1}} \leftarrow -p_{i_{c1}}$ ,  $p_{i_{c2}} \leftarrow -p_{i_{c2}}$ 
21:    end if
22:     $\varphi_v \leftarrow \varphi_v + 2v_i\Phi e_i$ 
23:     $t_{\text{remain}} \leftarrow t_{\text{remain}} - t$ 
24:  end while
25: return  $\mathbf{x}$ ,  $\mathbf{p}$ 
26: end function

```

* minPositiveRoot(a_i, b_i, c_i) returns the minimal positive root of the equation $a_i x^2 + b_i x + c = 0$, or else returns $+\infty$ if no positive root exists.

Jointly updating latent variables and across-trait covariance. The $N \times P_{\text{lat}}$ latent variables and $P_{\text{lat}} \times P_{\text{lat}}$ across-trait covariance are highly correlated with each other, so individual Gibbs updates can be inefficient. The posterior conditional of \mathbf{X} is truncated normal and thus allows for the efficient Hamiltonian zigzag simulation. The conditional distribution for

covariance components \mathbf{C} and \mathbf{D} has no such special structure, so we map them to an unconstrained space and deploy Hamiltonian dynamics based on Gaussian momentum. We use a standard mapping of \mathbf{C} elements to real numbers [29] that first transforms \mathbf{C} to canonical partial correlations (CPC) that fall in $[-1, 1]$ and then apply the Fisher transformation to map CPC to the real line. We then construct the joint update of latent variables and covariance via differential operator splitting [13, 14] to approximate the joint dynamics of Laplace-Gauss mixed momenta.

We denote the two concatenated sets of parameters \mathbf{X} and $\{\mathbf{C}, \mathbf{D}\}$ as $\mathbf{x} = (\mathbf{x}_G, \mathbf{x}_L)$ with momenta $\mathbf{p} = (\mathbf{p}_G, \mathbf{p}_L)$, where indices G and L refer to Gaussian or Laplace momenta. The joint sampler updates $(\mathbf{x}_G, \mathbf{p}_G)$ first, then $(\mathbf{x}_L, \mathbf{p}_L)$, followed by another update of $(\mathbf{x}_G, \mathbf{p}_G)$. This symmetric splitting ensures that the simulated dynamics is reversible and hence constitutes a valid *Metropolis* proposal mechanism [13]. The LG-STEP function in Algorithm 2 describes the process of simulating the joint dynamics for time duration 2ϵ via the analytical Hamiltonian zig-zag dynamics for $(\mathbf{x}_L, \mathbf{p}_L)$ and the approximate leapfrog dynamics (9) for $(\mathbf{x}_G, \mathbf{p}_G)$. Because \mathbf{x}_G and \mathbf{x}_L can have very different scales, we incorporate a tuning parameter, the step size ratio r , to allow different step sizes for the two dynamics. To approximate a trajectory of the joint dynamics from $t = 0$ to $t = \tau$, we apply the function LG-STEP $m = \lfloor \tau/2\epsilon \rfloor$ times, and accept or reject the end point following the standard acceptance probability formula [27, 28]. We call this version of HMC based on Laplace-Gauss mixed momenta as *LG-HMC* and describe one iteration of LG-HMC in Algorithm 2 where the inputs include the joint potential function $U(\mathbf{x}_G, \mathbf{x}_L)$. We use LG-HMC to update $\{\mathbf{X}, \mathbf{C}, \mathbf{D}\}$ as a Metropolis-within-Gibbs step of our random-scan Gibbs scheme. The overall sampling efficiency strongly depends on m , the step size ϵ and the step size ratio r , so it is preferable to auto-tune all of them. We provide an empirical method to automatically tune r in [S1 File](#). We provide another option utilizing the no-U-turn algorithm to automatically decide the trajectory length m [23] and call the resulting algorithm *LG No-U-Turn Sampler* (LG-NUTS). We adapt the step size ϵ with primal-dual averaging to achieve an optimal acceptance rate [23].

Algorithm 2 One LG-HMC iteration

```

1: function LG-HMC ( $\mathbf{x}_G, \mathbf{x}_L, \mathbf{p}_G, \mathbf{p}_L, U, m, \epsilon, r$ )
  ▷ Record the initial state
2:    $\mathbf{x}_G^0 \leftarrow \mathbf{x}_G, \mathbf{x}_L^0 \leftarrow \mathbf{x}_L, \mathbf{p}_G^0 \leftarrow \mathbf{p}_G, \mathbf{p}_L^0 \leftarrow \mathbf{p}_L$ 
3:   for  $i = 1, \dots, m$  do
4:      $\mathbf{x}_G, \mathbf{x}_L, \mathbf{p}_G, \mathbf{p}_L \leftarrow$  LG-STEP ( $\mathbf{x}_G, \mathbf{x}_L, \mathbf{p}_G, \mathbf{p}_L, \epsilon, r$ )
5:   end for
  ▷ Calculate the acceptance probability  $a$ , where  $K_G$  and  $K_L$  denote the
  kinetic energy based on Gaussian or Laplace momentum and  $\|\cdot\|_1, \|\cdot\|_2$ 
  are the  $L^1$  and  $L^2$  norm.
6:    $K_G^0 \leftarrow (\|\mathbf{p}_G^0\|_2)^2/2, K_L^0 \leftarrow \|\mathbf{p}_L^0\|_1$ 
7:    $K_G \leftarrow (\|\mathbf{p}_G\|_2)^2/2, K_L \leftarrow \|\mathbf{p}_L\|_1$ 
8:    $a \leftarrow \min\{1, \exp[U(\mathbf{x}_G^0, \mathbf{x}_L^0) - U(\mathbf{x}_G, \mathbf{x}_L) + K_G^0 + K_L^0 - K_G - K_L]\}$ 
  ▷ Accept or reject
9:    $u \leftarrow$  one draw from uniform(0, 1)
10:  if  $u < a$  then
11:    return  $\mathbf{x}_G, \mathbf{x}_L, \mathbf{p}_G, \mathbf{p}_L$ 
12:  else
13:    return  $\mathbf{x}_G^0, \mathbf{x}_L^0, \mathbf{p}_G^0, \mathbf{p}_L^0$ 
14:  end if
15: end function
16: function LG-STEP ( $\mathbf{x}_G, \mathbf{x}_L, \mathbf{p}_G, \mathbf{p}_L, \epsilon, r$ )
17:   $\mathbf{x}_G, \mathbf{p}_G \leftarrow$  LEAPFROG ( $\mathbf{x}_G, \mathbf{p}_G, \epsilon$ )
18:   $\mathbf{x}_L, \mathbf{p}_L \leftarrow$  HZZTMVN ( $\mathbf{x}_G, \mathbf{p}_G, r\epsilon$ )
19:   $\mathbf{x}_G, \mathbf{p}_G \leftarrow$  LEAPFROG ( $\mathbf{x}_G, \mathbf{p}_G, \epsilon$ )

```

```

20:   return  $\mathbf{x}_G, \mathbf{x}_L, \mathbf{p}_G, \mathbf{p}_L$ 
21: end function
22: function LEAPFROG( $\mathbf{x}_G, \mathbf{p}_G, \epsilon$ )
23:    $\mathbf{p}_G \leftarrow \mathbf{p}_G + \frac{\epsilon}{2} \nabla_{\mathbf{x}_G} \log P(\mathbf{x})$ 
24:    $\mathbf{x}_G \leftarrow \mathbf{x}_G + \epsilon \mathbf{p}_G$ 
25:    $\mathbf{p}_G \leftarrow \mathbf{p}_G + \frac{\epsilon}{2} \nabla_{\mathbf{x}_G} \log P(\mathbf{x})$ 
26:   return  $\mathbf{x}_G, \mathbf{x}_L$ 
27: end function

```

Results

To illustrate the broad applicability of our method, we detail three real-world applications and discuss the scientific findings. We first apply our method to the HIV virulence application of [8]. The improved efficiency allows us to estimate the across-trait partial correlation with adequate effective sample size (ESS) and to reveal the conditional dependence among traits of scientific interest. We use the same HIV data set to demonstrate that LG-HMC and LG-NUTS outperform BPS (Section “Efficiency gain from the new inference scheme”), followed by two more LG-NUTS applications on influenza and *Aquilegia* flower evolution. We conclude this section with MCMC convergence criteria and timing results.

HIV immune escape

In the HIV evolution application of [8], a main scientific focus lies on the association between HIV-1 immune escape mutations and virulence, the pathogen’s ability to cause disease. The human leukocyte antigen (HLA) system is predictive of the disease course as it plays an important role in the immune response against HIV-1. Through its rapid evolution, HIV-1 can acquire mutations that aid in escaping HLA-mediated immune response, but the escape mutations may reduce its fitness and virulence [30, 31]. [8] identify HLA escape mutations associated with virulence while controlling for the unknown evolutionary history of the viruses. However, [8] interpret their results based on the across-trait correlation \mathbf{C} which only informs marginal associations that can remain confounded. Now armed with a more efficient inference method, we direct our attention towards the across-trait partial correlation matrix \mathbf{R} .

The data contain $N = 535$ aligned HIV-1 *gag* gene sequences collected from 535 patients between 2003 and 2010 in Botswana and South Africa [31]. Each sequence is associated with 3 continuous and 21 binary traits. The continuous virulence measurements are replicative capacity (RC), viral load (VL) and cluster of differentiation 4 (CD4) cell count. The binary traits include the existence of HLA-associated escape mutations at 20 different amino acid positions in the *gag* protein and another trait for the sampling country (Botswana or South Africa). Fig 1 depicts across-trait correlations and partial correlations with posterior medians > 0.2 (or < -0.2). Compared to correlations (Fig 1A), we observe more partial correlations with greater magnitude (Fig 1B). They indicate conditional dependencies among traits after removing effects from other variables in the model, helping to explore the causal pathway. For example, we only detect a negative conditional dependence between RC and CD4. In other words, holding one of CD4 and RC as constant, the other does not affect VL, suggesting that RC increases VL via reducing CD4. The fact that RC is not found to share a strong conditional dependence with VL may be explained by the strong modulatory role of immune system on VL. Only when viruses with higher RC also lead to more immune damage, as reflected in the CD4 count, higher VL may be observed as a consequence of less suppression of viral replication. As such, our findings are in line with the demonstration that viral RC impacts HIV-1 immunopathogenesis independent of VL [32].

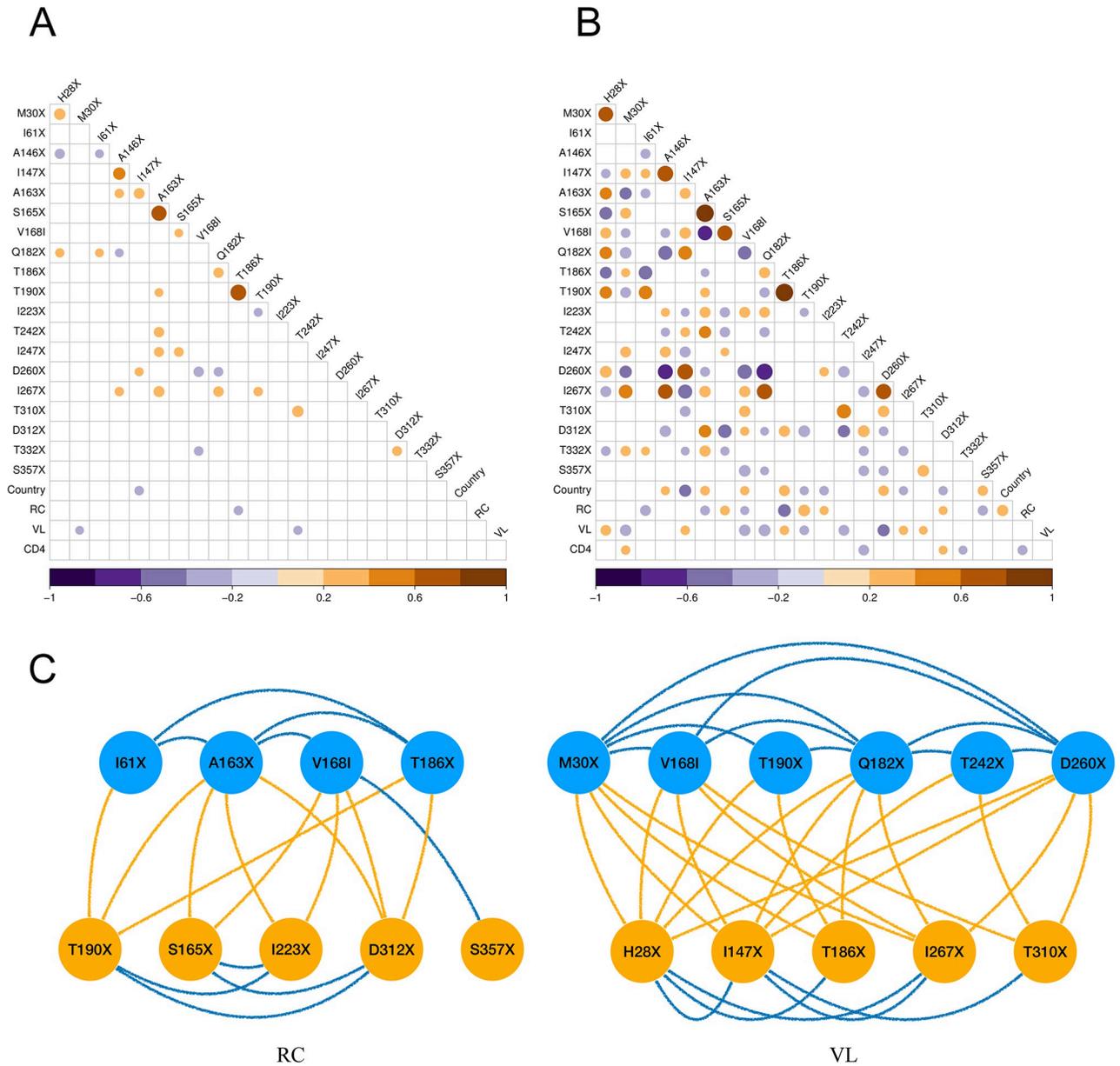


Fig 1. (A) Across-trait correlation and (B) partial correlation with a posterior median > 0.2 or < -0.2 (in color). HIV *gag* mutation names start with the wild type amino acid state, followed by the amino acid site number according to the HXB2 reference genome and end with the amino acid as a result of the mutation ('X' means a deletion). Country = sample region: 1 = South Africa, -1 = Botswana; RC = replicative capacity; VL = viral load; CD4 = CD4 cell count. (C) Conditional dependencies between HIV-1 immune escape mutations that affect RC or VL. Node and edge color indicates whether the dependence is positive (orange) or negative (blue).

<https://doi.org/10.1371/journal.pcbi.1011419.g001>

The partial correlation also helps to decipher epistatic interactions and how the escape mutations and potential compensatory mutations affect HIV-1 virulence. For example, we find a strong positive partial correlation between T186X and T190X. Studies have shown that T186X is highly associated with reduced VL [33, 34] and it requires T190I to partly compensate for this impaired fitness so the virus stays replication competent [35]. The negative conditional dependence between T186X and RC and the positive conditional dependence between

T190I and RC are consistent with this experimental observation. In contrast, with the strong positive association between T186X and T190, the marginal association fails to identify their opposite effects on RC. Another pair of mutations that potentially shows a similar interaction is H28X and M30X, which have a positive and negative partial correlation with VL, respectively. These mutations have indeed been observed to co-occur in *gag* epitopes from longitudinally followed-up patients [36]. Fig 1B keeps all the other compensatory mutation pairs in Fig 1A such as A146X-I147X and A163X-S165X that find confirmation in experimental studies [37, 38].

More generally, when considering the viral trait RC and the infection trait VL, for which their variation are to a considerable extent attributable to viral genetic variation [39], we reveal an intriguing pattern. As in Fig 1C, when two escape mutations impair virulence, and there is a conditional dependence between them, it is always negative. When two mutations have opposing effects on these virulence traits, the conditional dependence between them (if present) is almost always positive, with one exception of the negative effect between V168I and S357X. For example, T186X and I61X both have a negative impact on RC and the negative effect between them suggests that their additive, or even potentially synergistic, impact on RC is inhibited. Moreover, they appear to benefit from a compensatory mutation, T190X, which has been corroborated for the T186X-T190X pair at least as reported above. Also for VL, the conditional dependence between mutations that both have a negative impact on this virulence trait is consistently negative. Several of these individual mutations may benefit from H28X as a compensatory mutation, as indicated by the positive effect between pairs that include this mutation, and as suggested above for H28X—M30X. This illustrates the extent to which escape mutations may have a negative impact on virulence and the need to evolve compensatory mutations to restore it. We note that our analysis is not designed to recover compensatory mutations at great length as we restrict it to a limited set of known escape mutations, while mutations on many other sites may be compensatory. In fact, our analysis suggests that some of the considered mutations may be implicated in immune escape due to their compensatory effect rather than a direct escape benefit.

Efficiency gain from the new inference scheme

We demonstrate that the joint update of latent variables \mathbf{X} and the covariance matrix Ω significantly improve inference efficiency. For this purpose we use the large HIV dataset from Section “HIV immune escape” with $N = 535$, $P_{\text{disc}} = 21$, $P_{\text{cont}} = 3$, where the efficiency gain becomes significant. Our implementations of the algorithms have been validated on smaller truncated MVNs, on which simple rejection sampling can provide the ground truth up to quantifiable Monte Carlo errors. The Zigzag-HMC implementation has also been validated through the standalone implementation in an R package “htdg” [40].

We consider 4 sampling schemes BPS, Zigzag-HMC, LG-HMC, and LG-NUTS. To enable a more direct comparison while saving computational time, we separate tree inference from the inference for Ω and \mathbf{X} and fix \mathcal{F} as the maximum clade credibility tree from the HIV immune escape application. BPS and Zigzag-HMC only update \mathbf{X} and we use the standard NUTS transition kernel (i.e. standard HMC combined with no-U-turn algorithm) for the Ω elements. LG-HMC employs the joint update of \mathbf{X} and Ω described in Section “Jointly updating latent variables and across-trait covariance”. LG-NUTS additionally employs the No-U-Turn algorithm to decide the number of steps and a primal-dual averaging algorithm to calibrate the step size. We set the same t_{total} for BPS and Zigzag-HMC for a fair comparison. To tune LG-HMC, we first supply it with an optimal step size ϵ learned by LG-NUTS, then decide the number of steps $m = 100$ as it gives the best performance among the choices (10, 100,

Table 1. Efficiency comparison among different sampling schemes (BPS, Zigzag-HMC, LG-HMC, LG-NUTS). We calculate effective sample size (ESS) per hour run-time for the elements of \mathbf{C} , \mathbf{R} , \mathbf{X} , log joint density $\log p(\mathbf{X}, \Omega)$, and likelihood $l(\mathbf{X}, \Omega)$. For the three multivariate parameters (\mathbf{C} , \mathbf{R} , \mathbf{X}) with dimensions 276, 276, and 11,235, respectively, we report the minimal ESS across all dimensions. We conduct three independent simulations for each method and report the ESS values in the first three rows. We include the mean and standard deviation in the last row for each method to provide a summary of its overall performance. The bold number indicates the highest value in each of the five columns. For BPS, given the larger number of iterations required to achieve convergence, we record one sample of \mathbf{X} every 1,000 iterations to comply with storage limitations, and report upper bounds of the actual ESS by multiplying the ESS from thinned samples by 1,000.

ESS/hour	C(276d)	R(276d)	X (11,235d)	$\log p(\mathbf{X}, \Omega)$	$l(\mathbf{X}, \Omega)$
BPS	6.05	1.46	< 760*	0.56	0.56
	5.86	2.41	< 670	0.52	0.52
	0.55	0.49	< 100	0.42	0.43
	4.16(3.13)	1.45(0.96)	-	0.5(0.07)	0.5(0.07)
Zigzag-HMC	13.75	2.23	1480	4.42	4.44
	7.79	2.36	1057	5.38	5.38
	14.9	2.53	927	5.16	5.2
	12.15(3.82)	2.37(0.15)	1155(289)	4.99(0.5)	5.01(0.5)
LG-HMC	8.26	7.33	4.92	4.79	4.81
	7.11	8.59	7.76	5.09	5.1
	7.44	6.49	4.46	5.33	5.34
	7.6(0.59)	7.47(1.06)	5.71(1.79)	5.07(0.27)	5.08(0.26)
LG-NUTS	1.31	1.29	1.69	0.7	0.7
	11.93	7.52	6.37	1	1.06
	7.77	6.09	2.64	2.71	2.72
	7.01(5.35)	4.97(3.26)	3.57(2.47)	1.47(1.09)	1.49(1.07)

* The ESS estimates after 1/1000 thinning are 0.76, 0.67, 0.10

<https://doi.org/10.1371/journal.pcbi.1011419.t001>

1000). We conduct 3 independent simulations for each sampling scheme and report the per-run-time ESS for 5 parameters—the across-trait correlation \mathbf{C} , partial correlation \mathbf{R} , latent variable \mathbf{X} , log joint density $\log p(\mathbf{X}, \Omega)$ and log likelihood $l(\mathbf{X}, \Omega)$. \mathbf{C} and \mathbf{R} are of primary scientific interest as they provide insights into correlation structure among the traits. Examining ESS of the highest dimensional parameter \mathbf{X} is also important for diagnostic purposes. ESS's of $\log p(\mathbf{X}, \Omega)$ and $l(\mathbf{X}, \Omega)$ help us additionally evaluate how well the samplers explore the target distribution overall. As reported in Table 1, BPS is outperformed by the three other samplers in terms of efficiency for all five parameters. While a formal theoretical analysis is beyond the scope of this work, we provide an empirical explanation for the different performances of BPS and Zigzag-HMC in S2 File. LG-HMC achieves the highest per run-time ESS for \mathbf{R} , resulting in a $5\times$ speed-up compared to BPS. The result also highlights that inferring \mathbf{R} is more challenging than inferring \mathbf{C} , with the elements of \mathbf{R} generally having lower ESS, but the difficulty can be largely eliminated by jointly updating \mathbf{X} and Ω through LG-HMC and LG-NUTS. Although Zigzag-HMC achieves much higher ESS for \mathbf{X} than LG-HMC, the latter performs best in the most difficult and critical task of updating \mathbf{R} . Compared to LG-HMC, LG-NUTS exhibits lower efficiency and higher variance across the 3 runs, likely due to the No-U-Turn algorithm's tendency to require some extraneous leapfrog steps [41, 42]. We also provide the histograms for the per run-time ESS of \mathbf{R} elements in S1 Fig. Based on our findings, we recommend using LG-HMC with multiple choices of hyper parameters (m , ϵ), with a good starting point being (100, 0.01), or the auto-tuned LG-NUTS.

Glycosylation of Influenza A virus H1N1

Influenza A viruses of the H1N1 subtype currently circulate in birds, humans, and swine [43–45], where they are responsible for substantial morbidity and mortality [46, 47]. The two

surface glycoproteins hemagglutinin (HA) and neuraminidase (NA) interact with a cell surface receptor and so their characteristics largely affect virus fitness and transmissibility. Mutations in the HA and NA, particularly in their immunodominant head domain, sometimes produce glycosylations that shield the antigenic sites against detection by host antibodies and so help the virus evade antibody detection [48–51]. On the other hand, glycosylation may interfere with the receptor binding and also be targeted by the innate host immunity to neutralize viruses. Therefore there must be an equilibrium between competing pressures to evade immune detection and maintain virus fitness [52, 53]. The number of glycosylations that leads to this balance is expected to vary in host species experiencing different strengths of immune selection. Despite decades of tracking IAVs evolution in humans for vaccine strain selection and recent expansions of zoonotic surveillance, the evolvability and selective pressures on the HA and NA have not been rigorously compared across multiple host species. Here, we examine the conditional dependence between host type and multiple glycosylation sites by estimating the posterior distribution of across-trait partial correlation while jointly inferring the IAVs evolutionary history.

We use hemagglutinin (H1) and neuraminidase (N1) sequence data sets for influenza A H1N1 produced by Trovão et al. as described in [54]. We scan all H1 and N1 sequences to identify potential N-linked glycosylation sites, based on the motif Asn-X-Ser/Thr-X, where X is any amino acid other than proline (Pro) [55]. We then set a binary trait for each sequence encoding for the presence or absence of glycosylations at a particular amino acid site. We keep sites with a glycosylation frequency between 20% and 80% for our analysis. This gives six sites in H1 and four sites in N1. We include another binary trait for the host type being mammalian (human or swine) or avian, so the sample sizes are $N = 964$, $P = 7$ (H1) and $N = 896$, $P = 5$ (N1).

The six H1 glycosylation sites consist of three pairs that are physically close (63/94, 129/163, and 278/289, see Fig 2). Sites 63 and 94 are particularly close to each other, though distances will vary slightly with sequence. A negative conditional dependence suggests glycosylation at two close sites may be harmful for the virus (63/94 and 278/289) while a positive effect between two sites suggests a potential benefit (63/129 and 94/278). We detect a negative conditional dependence between mammalian host and glycosylation site 94 and 289. Avian viruses have a stronger tendency to have site 289 glycosylated (Fig 2).

In N1, glycosylations are more strongly correlated than H1 (Fig 3). Two pairs of glycosylation sites have a positive conditional dependency in between (50/68 and 50/389) and two pairs (44/68 and 68/389) have a negative one. We omit a structural interpretation since all sites but 389 are located in the NA stalk, for which no protein structure is available. There is a positive conditional dependence between mammalian host and glycosylations at sites 44 and 68. None of the avian lineages has glycosylation site 44 while most swine and some human lineages have it. Similarly, glycosylation at site 68 is present in most swine and human lineages but only in avian lineages circulating in wild birds, not those in poultry.

***Aquilegia* flower and pollinator co-evolution**

Reproductive isolation allows two groups of organisms to evolve separately, eventually forming new species. For plants, pollinators play an important role in reproductive isolation [56]. We examine the relationship between floral phenotypes and the three main pollinators for the columbine genus *Aquilegia*: bumblebees, hummingbirds, and hawk moths [18]. Here, the pollinator species represents a categorical trait with three classes and we choose bumblebee with the shortest tongue as the reference class. Fig 4 provides the across-trait correlation and partial correlation. Compared to a similar analysis on the same data set that only looks at correlation

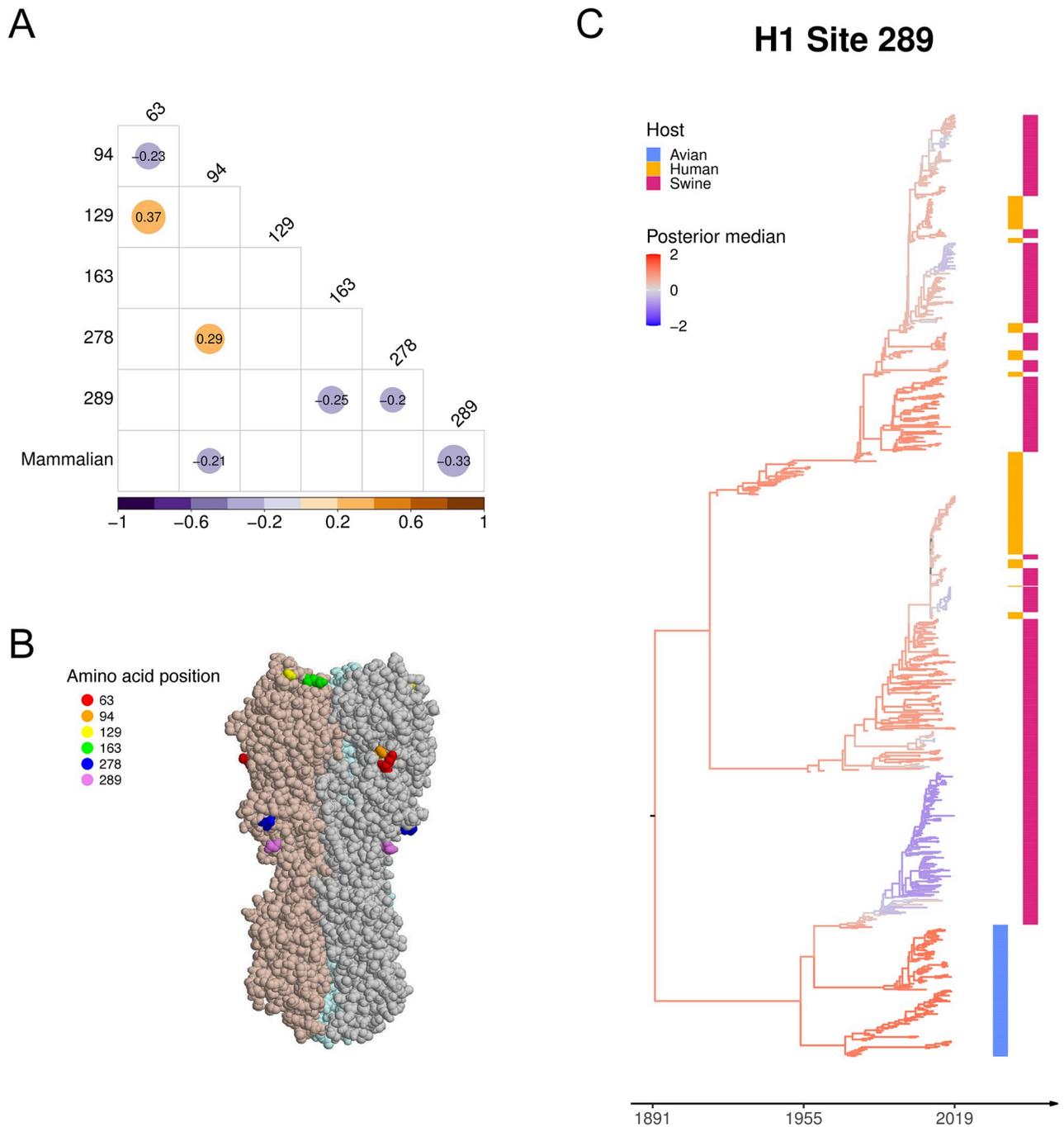


Fig 2. (A) Across-trait partial correlation among H1 glycosylation sites and host type with a posterior median > 0.2 or < -0.2 (in color and number). (B) HA structure of a 2009 H1N1 influenza virus (PDB entry 3LZG) with six glycosylation sites highlighted. Site 278 and 289 are in the stalk domain and all others are in the head domain. (C) The maximum clade credibility (MCC) tree with branches colored by the posterior median of the latent variable underlying H1 glycosylation site 289. The heatmap on the right indicates the host type of each taxon.

<https://doi.org/10.1371/journal.pcbi.1011419.g002>

or marginal association [1], partial correlation controls confounding and indicates the conditional dependencies between pollinators and floral phenotypes that can bring new insights.

For example, we observe a positive marginal association between hawk moth pollinator and spur length but no conditional dependence between them. The marginal association matches

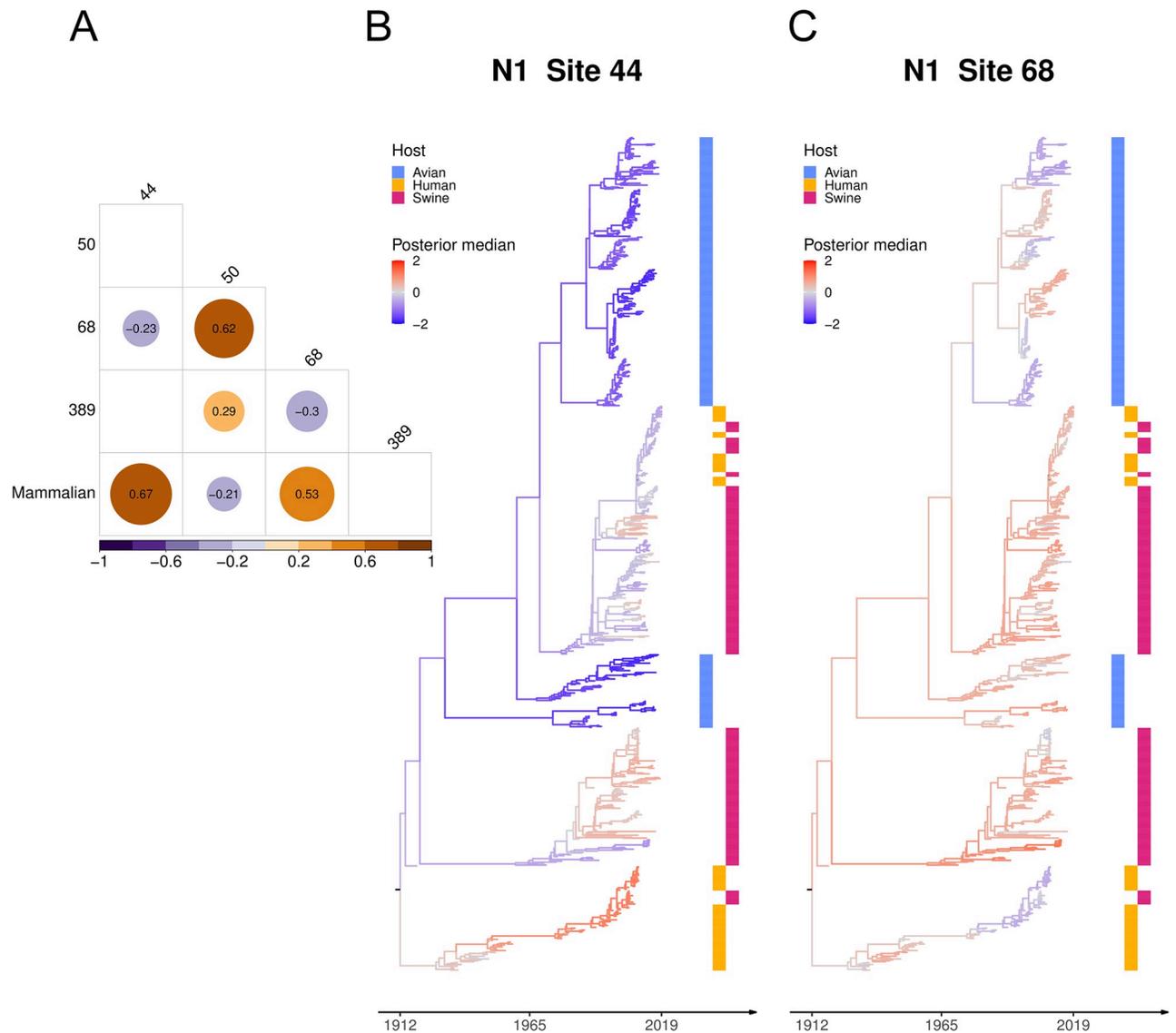


Fig 3. (A) Across-trait partial correlation among N1 glycosylation sites and host type with a posterior median > 0.2 or < -0.2 (in color and number). (B)(C) The maximum clade credibility (MCC) tree with branches colored by the posterior median of the latent variable underlying N1 glycosylation site 44 and 68.

<https://doi.org/10.1371/journal.pcbi.1011419.g003>

with the observation that flowers with long spur length have pollinators with long tongues [18, 57]. The absence of a conditional dependence makes intuitive sense because hawk moth's long tongue is not likely to stop them from visiting a flower with short spurs when the other floral traits are held constant. In fact, researchers observe that shortening the nectar spurs does not affect hawk moth visitation [58]. Similarly, the positive partial correlation between orientation and hawk moth also finds experimental support. The orientation trait is the angle of flower axis relative to gravity, in the range of $(0, 180)$. A small orientation value implies a pendent flower whereas a large value represents a more upright flower [59]. Due to their different morphologies, hawk moths prefer upright flowers while hummingbirds tend to visit pendent ones.

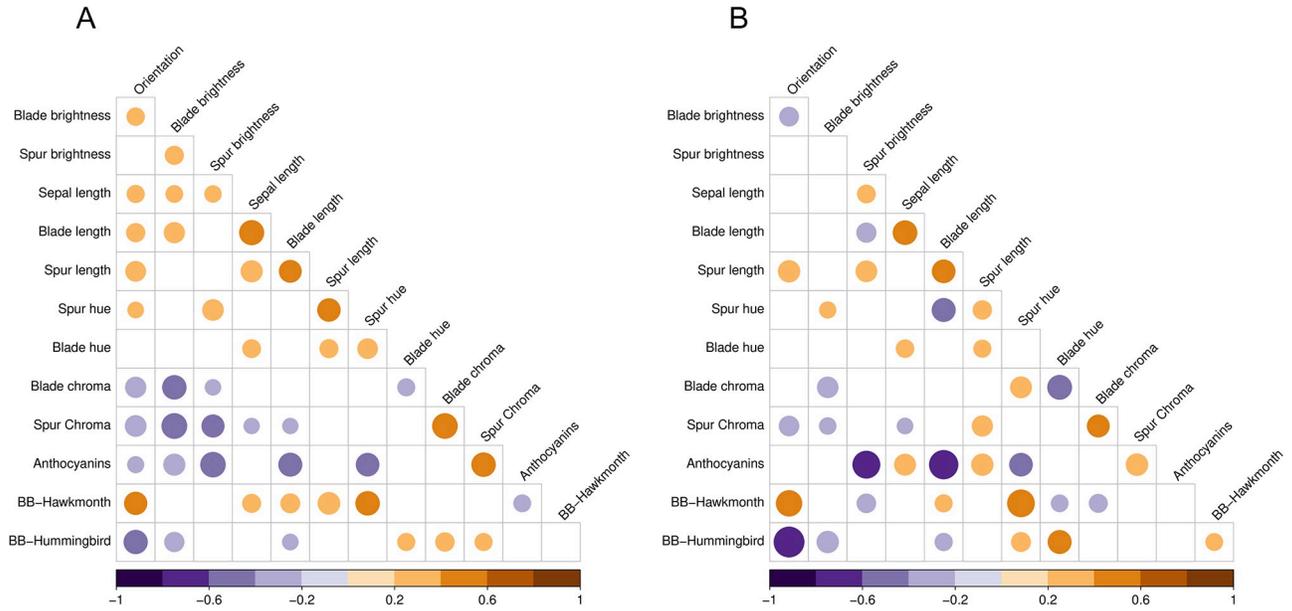


Fig 4. Across-trait correlations (A) and partial correlations (B) with posterior medians > 0.2 or < -0.2 (in color). BB = bumblebee.

<https://doi.org/10.1371/journal.pcbi.1011419.g004>

Making the naturally pendent *Aquilegia formosa* flowers upright increases hawk moth visitation [59]. These results suggest that partial correlation may have predictive power for results from carefully designed experiments with controlled variables.

MCMC setup and convergence assessment

We run all simulations on a node equipped with AMD EPYC 7642 server processors which possess 48 cores and 96 threads, with a base clock speed of 2.3 GHz. For every MCMC run, the minimal effective sample size (ESS) across all dimensions of \mathbf{X} and \mathbf{R} after burn-in is above 100. As another diagnostic, for our two large-scale applications on HIV-1 and H1N1 influenza, we run three independent chains and confirm the potential scale reduction statistic \hat{R} for all partial correlation elements falls between [1, 1.03], below the common criterion of 1.1 [60]. To reach a minimal ESS = 100 across all \mathbf{R} elements, the post burn-in run-time and number of MCMC transition kernels applied for the joint inference are 21 hours and 1.3×10^6 (HIV-1), 113 hours and 7.9×10^7 (H1), 76 hours and 1.4×10^8 (N1). These run-times suggest the difficulty of our large-scale inference tasks where besides the main challenge of sampling $\{\mathbf{X}, \mathbf{C}, \mathbf{D}\}$, updating the many tree parameters with Metropolis-Hastings transition kernels also takes a large number of iterations. To reduce the computational burden associated with tree inference, one practical approach is to utilize a set of pre-computed trees and incorporate tree swaps within the MCMC transition kernel.

Discussion

Learning how different biological traits interact with each other from many evolutionarily related taxa is a long-standing problem of scientific interest that sheds light on various aspects of evolution. Towards this goal, we develop a scalable solution that significantly improves inferential efficiency compared to established state-of-the-art approaches [1, 8]. Our novel strategy enables learning across-trait conditional dependencies that are more informative than

the previous marginal association based analyses. This approach provides reliable estimates of across-trait partial correlations for large problems, on which the established BPS-based method struggles. In two large-scale analyses featuring HIV-1 and H1N1 influenza, the improved efficiency allows us to infer conditional dependencies among traits of scientific interest and therefore investigate some of the most important molecular mechanisms underlying the disease. In addition, our approach incorporates automatic tuning, so that the most influential tuning parameters automatically adapt to the specific challenge the target distribution presents. Finally, we extend the phylogenetic probit model to include categorical traits and illustrate its use in examining the co-evolution of *Aquilegia* flower and pollinators.

We leverage the cutting-edge Zigzag-HMC [13] to tackle the exceedingly difficult computational task of sampling from a high-dimensional truncated normal distribution in the context of the phylogenetic probit model. Zigzag-HMC proves to be more efficient than the previously optimal approach that uses the BPS, especially when combined with differential operator splitting to jointly update two sets of parameters \mathbf{X} and Ω that are highly correlated. The improved efficiency allows us to obtain reliable estimates of the conditional dependencies among traits. In our applications, we find that these conditional dependencies better describe trait interactions than do the marginal associations. It is worth mentioning that another closely related sampler, the Markovian zigzag sampler [61], or MZZ, may also be appropriate for this task but provides lower efficiency than Zigzag-HMC [24]. While Zigzag-HMC is a recent and less explored version of HMC, BPS and MZZ are two central methods within the piecewise deterministic Markov process literature that have attracted growing interest in recent years [62, 63]. Intriguingly, the most expensive step of all three samplers is to obtain the log-density gradient, and the same linear-order gradient evaluation method [8] largely speeds it up.

We now consider limitations of this work and the future directions to which they point. First, the phylogenetic probit model does not currently accommodate a directional effect among traits since it only describes pairwise and symmetric correlations. However, the real biological processes are often not symmetric but directional, where it is common that one reaction may trigger another but not the opposite way. A model allowing directed paths is preferable since it better describes the complicated causal network among multiple traits. Graphical models with directed edges [64] are commonly used to learn molecular pathways [65, 66], but challenges remain to integrate these methods with a large and randomly distributed phylogenetic tree. Toward this goal, one may construct a continuous-time Markov chain to describe how discrete traits evolve [67, 68], but with P binary traits the transition rate matrix grows to the astronomical size 2^P . Second, though our method achieves the current best inference efficiency under the phylogenetic probit model, there is still room for improvement. In the influenza glycosylation example, we use a binary trait indicating the host being either avian or mammal (human or swine), instead of setting a categorical trait for host type. In fact, we choose not to use a three-class host type trait because it causes poor mixing for the partial correlation elements. We suspect two potential reasons for this. First, according to our model assumptions for categorical traits (Eq 2), the latent variables underneath the same trait are very negatively correlated, leading to a more correlated and challenging posterior. Second, in our specific data sets, the glycosylation sites tend to be similar in human and swine viruses, further increasing the correlation among posterior dimensions. One potential solution is to de-correlate some latent variables by grouping them into independent factors using phylogenetic factor analysis [69, 70]. Finally, one may consider a logistic or softmax function to map latent variables to the probability of a discrete trait. This avoids the hard truncations in the probit model but also adds another layer of noise. It requires substantial effort to develop an approach that overcomes the above limitations while supporting efficient inference at the scale of applications in this work.

Supporting information

S1 File. Auto-tuning of r . A heuristic to auto-tune the step size ratio r .
(PDF)

S2 File. Zigzag-HMC explores the energy space more efficiently than BPS. An intuition for BPS's slow movement in energy space.
(PDF)

S1 Fig. Histograms of per run-time ESS for r_{ij} .
(PDF)

Acknowledgments

We thank Kristel Van Laethem for useful discussion about HIV replicative capacity, CD4 counts and viral load. This work uses computational and storage services provided by the Hoffman2 Shared Cluster through the UCLA Institute for Digital Research and Education's Research Technology Group. The opinions expressed in this article are those of the authors and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

Author Contributions

Conceptualization: Zhenyu Zhang, Akihiko Nishimura, Marc A. Suchard.

Data curation: Nidia S. Trovão, Joshua L. Cherry, Philippe Lemey.

Formal analysis: Zhenyu Zhang, Nidia S. Trovão, Joshua L. Cherry, Philippe Lemey.

Funding acquisition: Marc A. Suchard.

Investigation: Philippe Lemey.

Methodology: Zhenyu Zhang, Akihiko Nishimura, Marc A. Suchard.

Software: Zhenyu Zhang, Marc A. Suchard.

Supervision: Akihiko Nishimura, Marc A. Suchard.

Writing – original draft: Zhenyu Zhang, Akihiko Nishimura, Nidia S. Trovão, Philippe Lemey, Marc A. Suchard.

Writing – review & editing: Zhenyu Zhang, Akihiko Nishimura, Nidia S. Trovão, Joshua L. Cherry, Andrew J. Holbrook, Xiang Ji, Philippe Lemey, Marc A. Suchard.

References

1. Cybis GB, Sinsheimer JS, Bedford T, Mather AE, Lemey P, Suchard MA. Assessing phenotypic correlation through the multivariate phylogenetic latent liability model. *Annals of Applied Statistics*. 2015; 9(2):969–991. <https://doi.org/10.1214/15-AOAS821>
2. Felsenstein J. Phylogenies and the comparative method. *The American Naturalist*. 1985; 125(1):1–15. <https://doi.org/10.1086/284325>
3. Fedorov V, Wu Y, Zhang R. Optimal dose-finding designs with correlated continuous and discrete responses. *Statistics in medicine*. 2012; 31(3):217–234. <https://doi.org/10.1002/sim.4388> PMID: 22162014
4. Schliep EM, Hoeting JA. Multilevel latent Gaussian process model for mixed discrete and continuous multivariate response data. *Journal of Agricultural, Biological, and Environmental Statistics*. 2013; 18(4):492–513. <https://doi.org/10.1007/s13253-013-0136-z>

5. Irvine KM, Rodhouse T, Keren IN. Extending ordinal regression with a latent zero-augmented beta distribution. *Journal of Agricultural, Biological and Environmental Statistics*. 2016; 21(4):619–640. <https://doi.org/10.1007/s13253-016-0265-2>
6. Pourmohamad T, Lee HK, et al. Multivariate stochastic process models for correlated responses of mixed type. *Bayesian Analysis*. 2016; 11(3):797–820. <https://doi.org/10.1214/15-BA976>
7. Clark JS, Nemergut D, Seyednasrollah B, Turner PJ, Zhang S. Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data. *Ecological Monographs*. 2017; 87(1):34–56. <https://doi.org/10.1002/ecm.1241>
8. Zhang Z, Nishimura A, Bastide P, Ji X, Payne RP, Goulder P, et al. Large-scale inference of correlation among mixed-type biological traits with phylogenetic multivariate probit models. *The Annals of Applied Statistics*. 2021; 15(1):230–251. <https://doi.org/10.1214/20-AOAS1394>
9. Chib S, Greenberg E. Analysis of multivariate probit models. *Biometrika*. 1998; 85(2):347–361. <https://doi.org/10.1093/biomet/85.2.347>
10. Bouchard-Côté A, Vollmer SJ, Doucet A. The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *Journal of the American Statistical Association*. 2018; 113(522):855–867. <https://doi.org/10.1080/01621459.2017.1294075>
11. Neal RM. MCMC using Hamiltonian dynamics. In: Brooks S, Gelman A, Jones GL, Meng XL, editors. *Handbook of Markov Chain Monte Carlo*. vol. 2. CRC Press New York, NY; 2011.
12. Liu JS, Wong WH, Kong A. Covariance structure and convergence rate of the Gibbs sampler with various scans. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1995; 57(1):157–169.
13. Nishimura A, Dunson DB, Lu J. Discontinuous Hamiltonian Monte Carlo for discrete parameters and discontinuous likelihoods. *Biometrika*. 2020; 107(2):365–380. <https://doi.org/10.1093/biomet/asz083>
14. Strang G. On the construction and comparison of difference schemes. *SIAM journal on numerical analysis*. 1968; 5(3):506–517. <https://doi.org/10.1137/0705041>
15. Shahbaba B, Lan S, Johnson WO, Neal RM. Split Hamiltonian Monte Carlo. *Statistics and Computing*. 2014; 24(3):339–349. <https://doi.org/10.1007/s11222-012-9373-1>
16. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*. 2018; 4(1):vey016. <https://doi.org/10.1093/ve/vey016> PMID: 29942656
17. Suchard MA, Weiss RE, Sinsheimer JS. Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular biology and evolution*. 2001; 18(6):1001–1013. <https://doi.org/10.1093/oxfordjournals.molbev.a003872> PMID: 11371589
18. Whittall JB, Hodges SA. Pollinator shifts drive increasingly long nectar spurs in columbine flowers. *Nature*. 2007; 447(7145):706–709. <https://doi.org/10.1038/nature05857> PMID: 17554306
19. Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*. 1993; 88(422):669–679. <https://doi.org/10.1080/01621459.1993.10476321>
20. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, et al. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proceedings of the National Academy of Sciences*. 2012; 109(37):15066–15071. <https://doi.org/10.1073/pnas.1206598109> PMID: 22927414
21. Lewandowski D, Kurowicka D, Joe H. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*. 2009; 100(9):1989–2001. <https://doi.org/10.1016/j.jmva.2009.04.008>
22. Kingman JFC. The coalescent. *Stochastic processes and their applications*. 1982; 13(3):235–248. [https://doi.org/10.1016/0304-4149\(82\)90011-4](https://doi.org/10.1016/0304-4149(82)90011-4)
23. Hoffman MD, Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*. 2014; 15(1):1593–1623.
24. Nishimura A, Zhang Z, Suchard MA. Hamiltonian zigzag sampler got more momentum than its Markovian counterpart: Equivalence of two zigzags under a momentum refreshment limit. *arXiv preprint arXiv:210407694*. 2021;.
25. Whittaker J. *Graphical models in applied multivariate statistics*. Wiley Publishing; 2009.
26. Leimkuhler B, Reich S. *Simulating Hamiltonian dynamics*. 14. Cambridge university press; 2004.
27. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*. 1953; 21(6):1087–1092. <https://doi.org/10.1063/1.1699114>
28. Hastings WK. *Monte Carlo sampling methods using Markov chains and their applications*. 1970;.
29. Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual, Version 2.18.0.*; 2018. Available from: <http://mc-stan.org/>.

30. Nomura S, Hosoya N, Brumme ZL, Brockman MA, Kikuchi T, Koga M, et al. Significant reductions in Gag-protease-mediated HIV-1 replication capacity during the course of the epidemic in Japan. *Journal of Virology*. 2013; 87(3):1465–1476. <https://doi.org/10.1128/JVI.02122-12> PMID: 23152532
31. Payne R, Muenchhoff M, Mann J, Roberts HE, Matthews P, Adland E, et al. Impact of HLA-driven HIV adaptation on virulence in populations of high HIV seroprevalence. *Proceedings of the National Academy of Sciences*. 2014; 111(50):E5393–E5400. <https://doi.org/10.1073/pnas.1413339111> PMID: 25453107
32. Claiborne DT, Prince JL, Scully E, Macharia G, Micci L, Lawson B, et al. Replicative fitness of transmitted HIV-1 drives acute immune activation, proviral load in memory CD4+ T cells, and disease progression. *Proceedings of the National Academy of Sciences*. 2015; 112(12):E1480–E1489. <https://doi.org/10.1073/pnas.1421607112> PMID: 25730868
33. Huang KHG, Goedhals D, Carlson JM, Brockman MA, Mishra S, Brumme ZL, et al. Progression to AIDS in South Africa is associated with both reverting and compensatory viral mutations. *PLoS One*. 2011; 6(4):e19018. <https://doi.org/10.1371/journal.pone.0019018> PMID: 21544209
34. Wright JK, Brumme ZL, Carlson JM, Heckerman D, Kadie CM, Brumme CJ, et al. Gag-protease-mediated replication capacity in HIV-1 subtype C chronic infection: associations with HLA type and clinical parameters. *Journal of Virology*. 2010; 84(20):10820–10831. <https://doi.org/10.1128/JVI.01084-10> PMID: 20702636
35. Wright JK, Naidoo VL, Brumme ZL, Prince JL, Claiborne DT, Goulder PJ, et al. Impact of HLA-B* 81-associated mutations in HIV-1 Gag on viral replication capacity. *Journal of Virology*. 2012; 86(6):3193–3199. <https://doi.org/10.1128/JVI.06682-11> PMID: 22238317
36. Olusola BA, Olaleye DO, Odaibo GN. Non-synonymous Substitutions in HIV-1 GAG Are Frequent in Epitopes Outside the Functionally Conserved Regions and Associated With Subtype Differences. *Front Microbiol*. 2020; 11:615721. <https://doi.org/10.3389/fmicb.2020.615721> PMID: 33505382
37. Crawford H, Prado JG, Leslie A, Hué S, Honeyborne I, Reddy S, et al. Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. *J Virol*. 2007; 81(15):8346–51. <https://doi.org/10.1128/JVI.00465-07> PMID: 17507468
38. Troyer RM, McNevin J, Liu Y, Zhang SC, Krizan RW, Abraha A, et al. Variable fitness impact of HIV-1 escape mutations to cytotoxic T lymphocyte (CTL) response. *PLoS Pathog*. 2009; 5(4):e1000365. <https://doi.org/10.1371/journal.ppat.1000365> PMID: 19343217
39. Blanquart F, Wymant C, Cornelissen M, Gall A, Bakker M, Bezemer D, et al. Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. *PLoS Biol*. 2017; 15(6): e2001855. <https://doi.org/10.1371/journal.pbio.2001855> PMID: 28604782
40. Zhang Z, Chin A, Nishimura A, Suchard MA. hdtg: An R package for high-dimensional truncated normal simulation. *arXiv preprint arXiv:221001097*. 2022;.
41. Wang Z, Mohamed S, Freitas N. Adaptive Hamiltonian and Riemann manifold Monte Carlo. In: *Proceedings of the 30th International Conference on Machine Learning*; 2013. p. 1462–1470.
42. Wu C, Stoehr J, Robert CP. Faster Hamiltonian Monte Carlo by learning leapfrog scale. *arXiv:181004449*. 2018;.
43. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. Evolution and ecology of influenza A viruses. *Microbiological reviews*. 1992; 56(1):152–179. <https://doi.org/10.1128/mr.56.1.152-179.1992> PMID: 1579108
44. Song D, Kang B, Lee C, Jung K, Ha G, Kang D, et al. Transmission of avian influenza virus (H3N2) to dogs. *Emerging infectious diseases*. 2008; 14(5):741. <https://doi.org/10.3201/eid1405.071471> PMID: 18439355
45. Trovão NS, Nelson MI. When Pigs Fly: Pandemic influenza enters the 21st century. *PLoS pathogens*. 2020; 16(3):e1008259. <https://doi.org/10.1371/journal.ppat.1008259> PMID: 32191770
46. Boni MF, Galvani AP, Wickelgren AL, Malani A. Economic epidemiology of avian influenza on small-holder poultry farms. *Theoretical population biology*. 2013; 90:135–144. <https://doi.org/10.1016/j.tpb.2013.10.001> PMID: 24161559
47. Ma W. Swine influenza virus: Current status and challenge. *Virus research*. 2020; 288:198118. <https://doi.org/10.1016/j.virusres.2020.198118> PMID: 32798539
48. Skehel J, Stevens D, Daniels R, Douglas A, Knossow M, Wilson I, et al. A carbohydrate side chain on hemagglutinins of Hong Kong influenza viruses inhibits recognition by a monoclonal antibody. *Proceedings of the National Academy of Sciences*. 1984; 81(6):1779–1783. <https://doi.org/10.1073/pnas.81.6.1779> PMID: 6584912
49. Hebert DN, Zhang JX, Chen W, Foellmer B, Helenius A. The number and location of glycans on influenza hemagglutinin determine folding and association with calnexin and calreticulin. *The Journal of cell biology*. 1997; 139(3):613–623. <https://doi.org/10.1083/jcb.139.3.613> PMID: 9348279

50. Daniels R, Kurowski B, Johnson AE, Hebert DN. N-linked glycans direct the cotranslational folding pathway of influenza hemagglutinin. *Molecular cell*. 2003; 11(1):79–90. [https://doi.org/10.1016/S1097-2765\(02\)00821-3](https://doi.org/10.1016/S1097-2765(02)00821-3) PMID: 12535523
51. Östbye H, Gao J, Martinez MR, Wang H, de Gier JW, Daniels R. N-linked glycan sites on the influenza A virus neuraminidase head domain are required for efficient viral incorporation and replication. *Journal of Virology*. 2020; 94(19):e00874–20. <https://doi.org/10.1128/JVI.00874-20> PMID: 32699088
52. Tate MD, Job ER, Deng YM, Gunalan V, Maurer-Stroh S, Reading PC. Playing hide and seek: how glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection. *Viruses*. 2014; 6(3):1294–1316. <https://doi.org/10.3390/v6031294> PMID: 24638204
53. Lin B, Qing X, Liao J, Zhuo K. Role of protein glycosylation in host-pathogen interaction. *Cells*. 2020; 9(4):1022. <https://doi.org/10.3390/cells9041022> PMID: 32326128
54. Trovão NS, Khan SM, Lemey P, Nelson MI, Cherry JL. Evolution of influenza A virus hemagglutinin H1 and H3 across host species. *bioRxiv*. 2022; p. 2022–04.
55. Mellquist J, Kasturi L, Spitalnik S, Shakin-Eshleman S. The amino acid following an Asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. *Biochemistry*. 1998; 37(19):6833–6837. <https://doi.org/10.1021/bi972217k> PMID: 9578569
56. Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2008; 363(1506):3009–3021. <https://doi.org/10.1098/rstb.2008.0064> PMID: 18579478
57. Rosas-Guerrero V, Aguilar R, Martén-Rodríguez S, Ashworth L, Lopezaiza-Mikel M, Bastida JM, et al. A quantitative review of pollination syndromes: do floral traits predict effective pollinators? *Ecology letters*. 2014; 17(3):388–400. <https://doi.org/10.1111/ele.12224> PMID: 24393294
58. Fulton M, Hodges SA. Floral isolation between *Aquilegia formosa* and *Aquilegia pubescens*. *Proceedings of the Royal Society of London Series B: Biological Sciences*. 1999; 266(1435):2247–2252. <https://doi.org/10.1098/rspb.1999.0915>
59. Hodges SA, Whittall JB, Fulton M, Yang JY. Genetics of floral traits influencing reproductive isolation between *Aquilegia formosa* and *Aquilegia pubescens*. *The American Naturalist*. 2002; 159(S3):S51–S60. <https://doi.org/10.1086/338372> PMID: 18707369
60. Gelman A, Rubin DB, et al. Inference from iterative simulation using multiple sequences. *Statistical science*. 1992; 7(4):457–472. <https://doi.org/10.1214/ss/1177011136>
61. Bierkens J, Fearnhead P, Roberts G, et al. The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *The Annals of Statistics*. 2019; 47(3):1288–1320. <https://doi.org/10.1214/18-AOS1715>
62. Fearnhead P, Bierkens J, Pollock M, Roberts GO. Piecewise deterministic Markov processes for continuous-time Monte Carlo. *Statistical Science*. 2018; 33(3):386–412. <https://doi.org/10.1214/18-STS648>
63. Dunson DB, Johndrow J. The Hastings algorithm at fifty. *Biometrika*. 2020; 107(1):1–23. <https://doi.org/10.1093/biomet/asz066>
64. Lauritzen SL. *Graphical models*. vol. 17. Clarendon Press; 1996.
65. Neapolitan R, Xue D, Jiang X. Modeling the altered expression levels of genes on signaling pathways in tumors as causal Bayesian networks. *Cancer Informatics*. 2014; 13:CIN–S13578. <https://doi.org/10.4137/CIN.S13578> PMID: 24932098
66. Benedetti E, Pučić-Baković M, Keser T, Wahl A, Hassinen A, Yang JY, et al. Network inference from glycoproteomics data reveals new reactions in the IgG glycosylation pathway. *Nature communications*. 2017; 8(1):1–15. <https://doi.org/10.1038/s41467-017-01525-0> PMID: 29133956
67. Pagel M. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London Series B: Biological Sciences*. 1994; 255(1342):37–45. <https://doi.org/10.1098/rspb.1994.0006>
68. O'Meara BC. Evolutionary inferences from phylogenies: a review of methods. *Annual Review of Ecology, Evolution, and Systematics*. 2012; 43:267–285. <https://doi.org/10.1146/annurev-ecolsys-110411-160331>
69. Tolkoﬀ MR, Alfaro ME, Baele G, Lemey P, Suchard MA. Phylogenetic factor analysis. *Systematic biology*. 2018; 67(3):384–399. <https://doi.org/10.1093/sysbio/syx066> PMID: 28950376
70. Hassler GW, Gallone B, Aristide L, Allen WL, Tolkoﬀ MR, Holbrook AJ, et al. Principled, practical, flexible, fast: a new approach to phylogenetic factor analysis. *arXiv preprint arXiv:210701246*. 2021;.