Supplementary Information

SI 1. MATRIX INVERSION COMPUTATIONS

To evaluate the observed data likelihood, we must compute branch-deflated precisions $\mathbf{P}_i^\star = \left(\mathbf{P}_i^- + t_i \boldsymbol{\delta}_i \boldsymbol{\Sigma} \boldsymbol{\delta}_i\right)^-$ for $i = 1, \ldots, 2N-2$. We demonstrate below that this matrix exists and is well-defined under the definition of our pseudo-inverse. Using the permutation matrix $\mathbf{C}_i$ from Section 2.1.1, we decompose the diffusion variance $\boldsymbol{\Sigma}$ and node precision $\mathbf{P}_i$ such that

$$\boldsymbol{\Sigma} = \mathbf{C}_i \begin{pmatrix} \boldsymbol{\Sigma}_i^{\text{obs}} & \boldsymbol{\Sigma}_i^{\text{ol}} & \boldsymbol{\Sigma}_i^{\text{om}} \\ - & \boldsymbol{\Sigma}_i^{\text{lat}} & \boldsymbol{\Sigma}_i^{\text{lm}} \\ - & - & \boldsymbol{\Sigma}_i^{\text{mis}} \end{pmatrix} \mathbf{C}_i^t \text{ and}$$
$$\mathbf{P}_i = \mathbf{C}_i \text{diag}\left[\infty\mathbf{I}, \tilde{\mathbf{P}}_i, 0\mathbf{I}\right] \mathbf{C}_i^t,$$

for $i = 1, \ldots, 2N-2$. We use this decomposition to identify that:

$$\begin{aligned}
\mathbf{P}_i^\star &= \left(\mathbf{P}_i^- + t_i \boldsymbol{\delta}_i \boldsymbol{\Sigma} \boldsymbol{\delta}_i\right)^- \\
&= \mathbf{C}_i \left(\left(\text{diag}\left[\infty\mathbf{I}, \tilde{\mathbf{P}}_i, 0\mathbf{I}\right] \right)^- + \text{diag}\left[t_i \begin{pmatrix} \boldsymbol{\Sigma}_i^{\text{obs}} & \boldsymbol{\Sigma}_i^{\text{ol}} \\ - & \boldsymbol{\Sigma}_i^{\text{lat}} \end{pmatrix}, 0\mathbf{I} \right] \right)^- \mathbf{C}_i^t \\
&= \mathbf{C}_i \left(\text{diag}\left[0\mathbf{I}, \tilde{\mathbf{P}}_i^{-1}, \infty\mathbf{I}\right] + \text{diag}\left[t_i \begin{pmatrix} \boldsymbol{\Sigma}_i^{\text{obs}} & \boldsymbol{\Sigma}_i^{\text{ol}} \\ - & \boldsymbol{\Sigma}_i^{\text{lat}} \end{pmatrix}, 0\mathbf{I} \right] \right)^- \mathbf{C}_i^t \\
&= \mathbf{C}_i \left(\text{diag}\left[\mathbf{T}, \infty\mathbf{I}\right] \right)^- \mathbf{C}_i^t \\
&= \mathbf{C}_i \, \text{diag}\left[\mathbf{T}^{-1}, 0\mathbf{I}\right] \mathbf{C}_i^t,
\end{aligned} \tag{1}$$

where

$$\mathbf{T} = \text{diag}\left[0\mathbf{I}, \tilde{\mathbf{P}}_i^{-1}\right] + t_i \begin{pmatrix} \boldsymbol{\Sigma}_i^{\text{obs}} & \boldsymbol{\Sigma}_i^{\text{ol}} \\ - & \boldsymbol{\Sigma}_i^{\text{lat}} \end{pmatrix} = \begin{pmatrix} t_i \boldsymbol{\Sigma}_i^{\text{obs}} & t_i \boldsymbol{\Sigma}_i^{\text{ol}} \\ - & \tilde{\mathbf{P}}_i^{-1} + t_i \boldsymbol{\Sigma}_i^{\text{lat}} \end{pmatrix}. \tag{2}$$

The matrix $\mathbf{T}$ is the sum of a positive-definite matrix and positive-semidefinite matrix and is therefore invertible.

SI 2. HERITABILITY STATISTIC

We compute the expectation of the empirical variance $\mathbb{E}[\mathbf{S}^2(\mathbf{Y})]$ under the MBD model with residual variance as follows:

$$\begin{aligned}
\mathbb{E}\left[\mathbf{S}^2(\mathbf{Y})\right] &= \mathbb{E}\left[\frac{1}{N}\left(\mathbf{Y}-\bar{\mathbf{Y}}\right)^t\left(\mathbf{Y}-\bar{\mathbf{Y}}\right)\right] \\
&= \frac{1}{N}\mathbb{E}\left[\mathbf{Y}^t\mathbf{Y} - \frac{2}{N}\mathbf{Y}^t\mathbf{J}_N\mathbf{Y} + \frac{1}{N^2}\mathbf{Y}^t\mathbf{J}_N\mathbf{J}_N\mathbf{Y}\right] \\
&= \frac{1}{N}\mathbb{E}\left[\mathbf{Y}^t\mathbf{Y} - \frac{2}{N}\mathbf{Y}^t\mathbf{J}_N\mathbf{Y} + \frac{1}{N}\mathbf{Y}^t\mathbf{J}_N\mathbf{Y}\right] \\
&= \frac{1}{N}\mathbb{E}\left[\mathbf{Y}^t\mathbf{Y} - \frac{1}{N}\mathbf{Y}^t\mathbf{J}_N\mathbf{Y}\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\mathbf{Y}_i\mathbf{Y}_i^t\right] - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left[\mathbf{Y}_i\mathbf{Y}_j^t\right].
\end{aligned} \tag{3}$$

The multivariate normal distribution of $\mathrm{vec}\,[\mathbf{Y}]$ implies $\mathrm{Cov}\,(Y_{ik}, Y_{jl}) = \Sigma_{kl}\Upsilon_{ij} + \Gamma_{kl}^{-1}1_{\{i=j\}}$ where $1_{\{i=j\}}$ is an indicator function. Using this information in SI Equation 3,

$$\begin{aligned}
\mathbb{E}\left[\mathbf{S}^2(\mathbf{Y})\right] &= \frac{1}{N}\sum_{i=1}^{N}\left(\Upsilon_{ii}\boldsymbol{\Sigma} + \boldsymbol{\Gamma}^{-1} + \mathbb{E}[\mathbf{Y}_i]\,\mathbb{E}[\mathbf{Y}_i]^t\right) \\
&\quad - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\Upsilon_{ij}\boldsymbol{\Sigma} + \boldsymbol{\Gamma}^{-1}1_{\{i=j\}} + \mathbb{E}[\mathbf{Y}_i]\,\mathbb{E}[\mathbf{Y}_j]^t\right) \\
&= \frac{1}{N}\mathrm{tr}\,[\boldsymbol{\Upsilon}]\,\boldsymbol{\Sigma} + \boldsymbol{\Gamma}^{-1} - \left(\frac{1}{N^2}\mathbf{1}_N^t\boldsymbol{\Upsilon}\mathbf{1}_N\right)\boldsymbol{\Sigma} - \frac{1}{N}\boldsymbol{\Gamma}^{-1} \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\mathbf{Y}_i]\,\mathbb{E}[\mathbf{Y}_i]^t - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}[\mathbf{Y}_i]\,\mathbb{E}[\mathbf{Y}_j]^t.
\end{aligned} \tag{4}$$

Note that $\mathbb{E}[\mathbf{Y}_i] = \mathbf{Y}_{2N-1}$ for $i = 1 \ldots N$, which implies

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\mathbf{Y}_i]\,\mathbb{E}[\mathbf{Y}_i]^t - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}[\mathbf{Y}_i]\,\mathbb{E}[\mathbf{Y}_j]^t = 0. \tag{5}$$

SI Table 1: Likelihood calculation speed comparison between BEAST and PCMBaseCpp. Each data set was run 10 times for 1,000 likelihood evaluations each. We report the median likelihood evaluations per second and speed-up over the 10 runs.

Data set	N	P	Likelihood evaluations/sec		Speed-up
			BEAST	PCMBaseCpp	
Prokaryotes	705	7	240	40	6.0×
HIV	1536	3	490	67	7.2×
Mammals	3649	8	60	12	5.1×

As such, our expression for the expected empirical variance reduces to the following:

$$\mathbb{E}\left[\mathbf{S}^2(\mathbf{Y})\right] = \frac{N-1}{N}\boldsymbol{\Gamma}^{-1} + \left(\frac{1}{N}\mathrm{tr}\left[\boldsymbol{\Upsilon}\right] - \frac{1}{N^2}\mathbf{1}_N^t\boldsymbol{\Upsilon}\mathbf{1}_N\right)\boldsymbol{\Sigma}. \tag{6}$$

SI 3. COMPARISON WITH PCMBaseCpp

As our algorithm for efficiently computing the likelihood with incomplete trait measurements relies on a similar strategy as that presented by Mitov et al. (2020), we compare the likelihood computation speed of our BEAST (Suchard et al., 2018) implementation against and the PCMBaseCpp implementation. We record the time it takes to evaluate the likelihood 1,000 times using the data and trees from all three examples we discuss in the text, and repeat this ten times for each example. We report the median likelihoods per second in SI Table 1. We also perform the same comparisons with simulated trees and data sets, and report these results in SI Table 2.

Note that while we do show consistently faster likelihood evaluations than PCMBase, we do not believe that our implementation is necessarily "better" than that of Mitov et al. (2020). The primary difficulty in comparing the speed of the two software packages is that we implement our software in different languages (BEAST in Java and PCMBase in R and C++), and the specific Java and C++ compilers used could influence their speed. It is difficult to determine the exact sources of the differences in speed without testing both

SI Table 2: Likelihood calculation speed comparison between BEAST and PCMBaseCpp on simulated data. For each N, P combination, data was simulated 10 times under random conditions and run for 1,000 likelihood evaluations each. We report the median likelihood evaluations per second and speed-up over the 10 runs.

N	P	Likelihood evaluations/sec		Speed-up
		BEAST	PCMBaseCpp	
100	2	3300	1300	2.6×
100	10	690	180	3.8×
100	20	220	26	8.3×
1,000	2	780	170	4.5×
1,000	10	100	13	7.9×
1,000	20	25	2.8	8.8×
10,000	2	82	16	5.1×
10,000	10	11	1.7	6.4×
10,000	20	2.5	0.29	8.7×

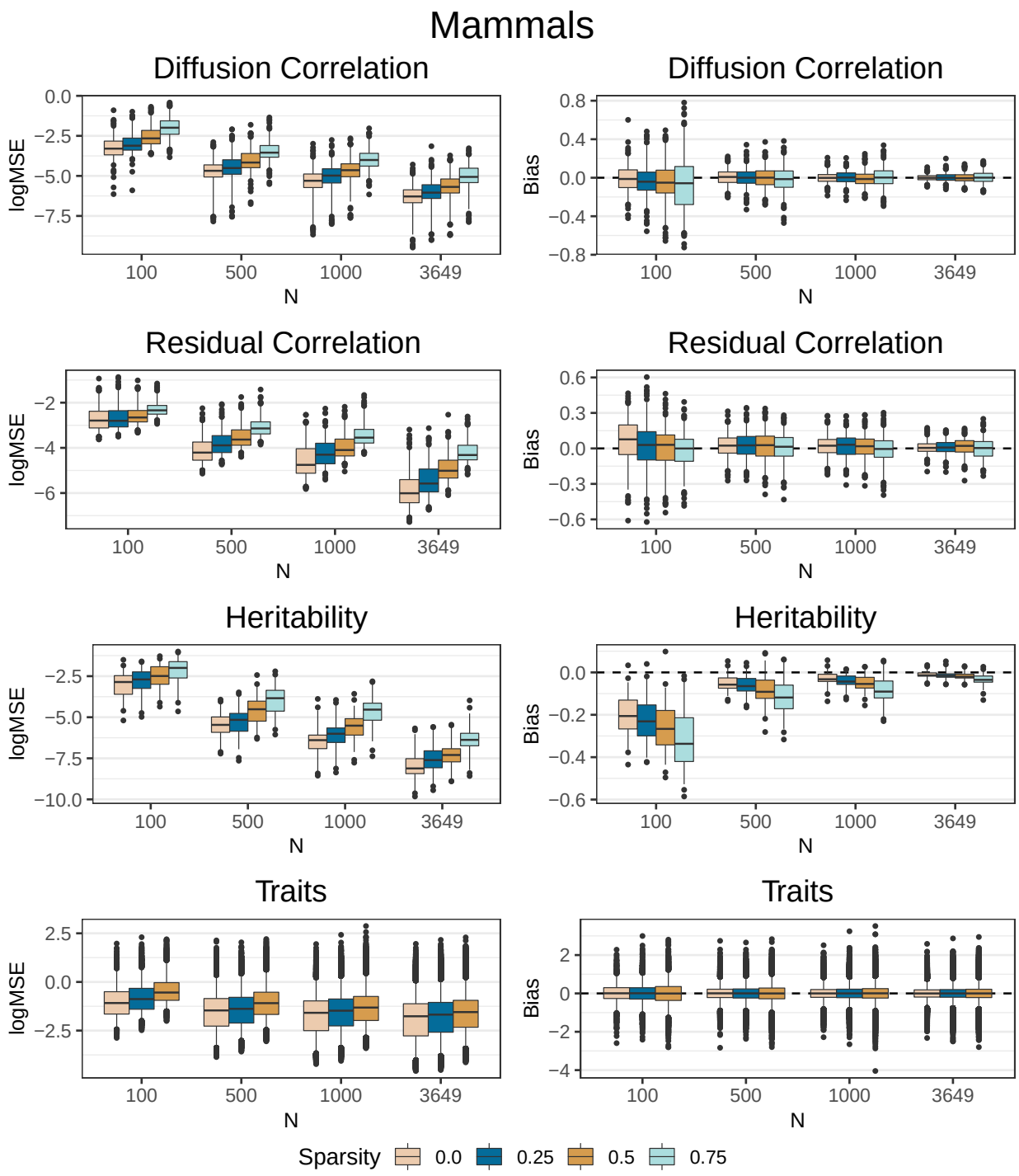
implementations on a wide range of computer architectures and compilers.

Nevertheless, the PCMBase / PCMFit packages and BEAST are fundamentally different in that PCMFit relies on maximum likelihood estimation (MLE) while BEAST performs Bayesian inference. The MLE framework is certainly appropriate when the phylogenetic tree is known with a high degree of certainty, but poses problems when the phylogenetic tree is unknown and must be jointly inferred with the trait evolutionary process. Specifically, MLE will likely produce biased results and has difficulty constructing confidence intervals that take into account the uncertainty of the tree. From the Bayesian perspective, however, we can simply integrate out the tree via Markov Chain Monte Carlo, that results in posterior estimates of the trait evolution parameters that accurately reflect the uncertainty of the tree.

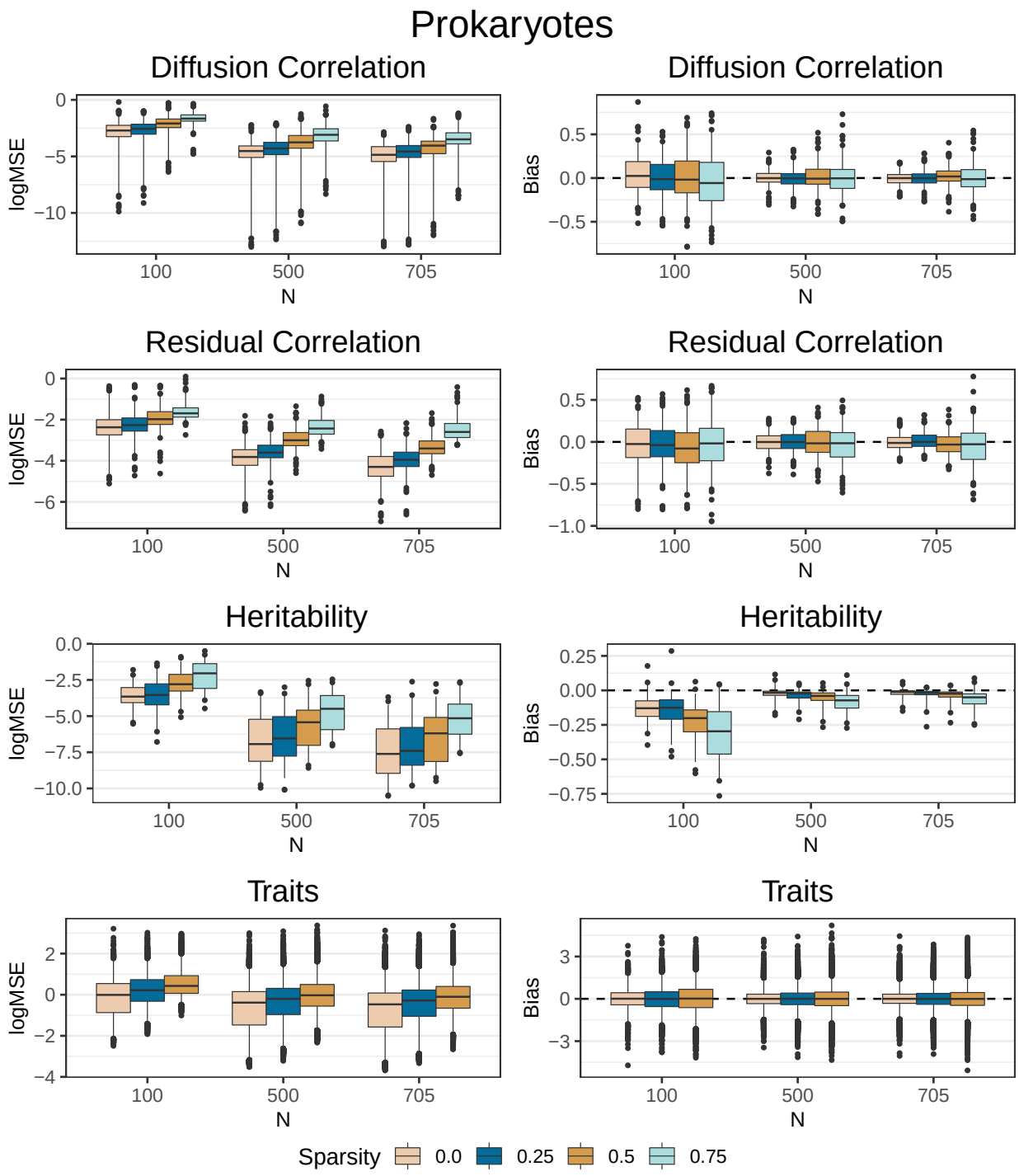
SI 4. SIMULATION STUDY

The setup of our simulation study is described in Section 6. SI Figures 1, 2, and 3 present the results of our simulation study. In general, results indicate that our inference machinery

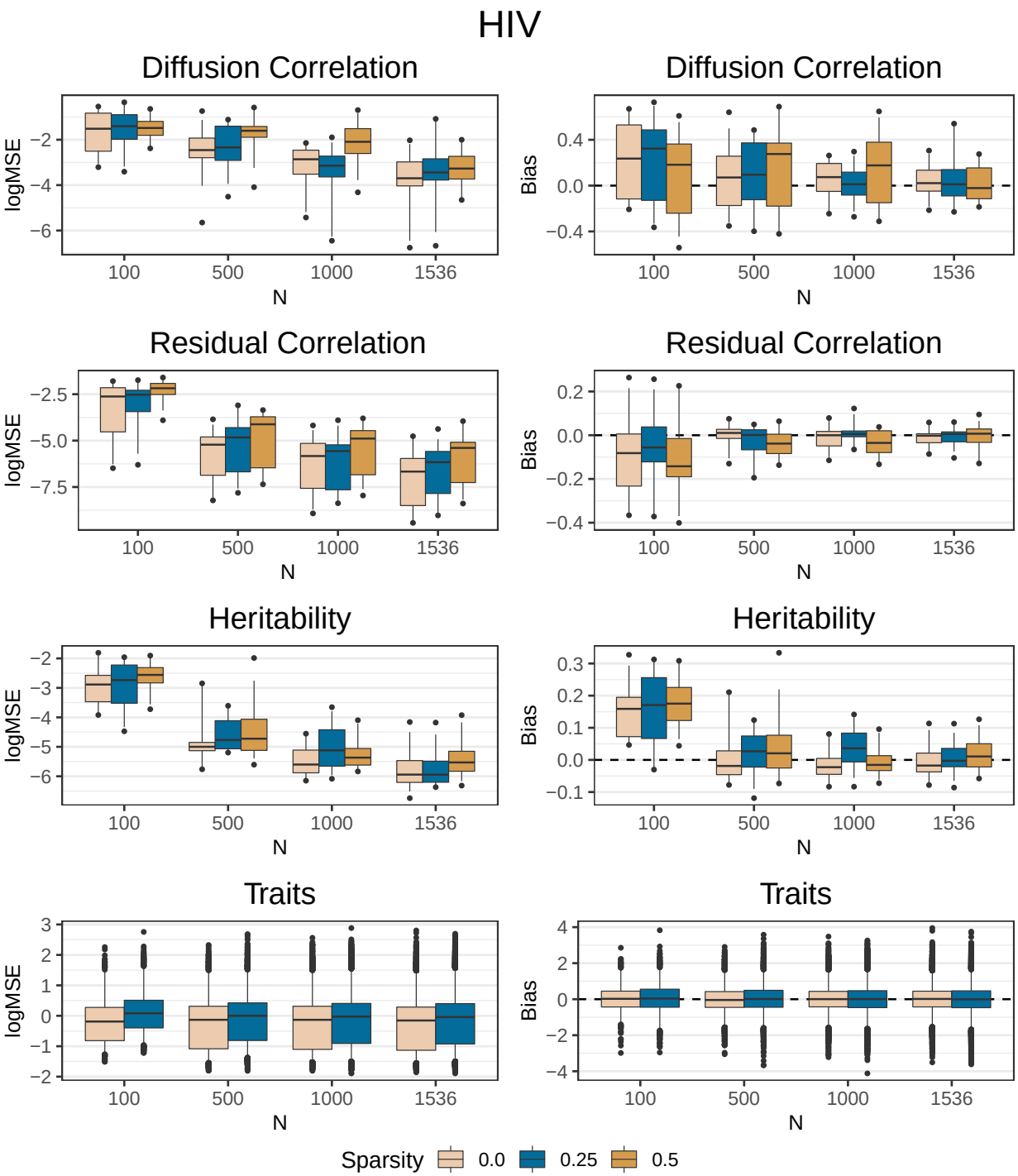
is sufficiently well-powered to accurately and precisely recapture the parameters used to simulate the data. All parameters of interest achieve low posterior mean squared error (MSE) when all available taxa are included. Additionally, there is no apparent bias in our parameter estimation with the notable exception of the diagonal heritabilities. Note that despite the fact that there is some bias in the heritability estimates, they also achieve low logMSE and are indeed close to their "true" values. We believed the induced prior on the diagonal heritabilities may be responsible for this bias, but have not fully explored this phenomenon. Regardless, these results suggest that (conditioning on the model being appropriate) our results accurately reflect biological reality.



SI Figure 1: Mammals simulation study. Posterior log mean squared-error and bias of the parameters of interest over ten simulated replicates. The boxes extend from the 25^{th} to the 75^{th} posterior percentiles with the middle bar representing the median. The lines extend from the 2.5^{th} through the 97.5^{th} percentiles, with outliers depicted as dots. The sparsity depicted by different colors represents different percentages of randomly removed data.



SI Figure 2: Prokaryote simulation study results. See SI Figure 1 for description.



SI Figure 3: HIV-1 simulation study results. See SI Figure 1 for description.

Bibliography

Mitov, V., K. Bartoszek, G. Asimomitis, and T. Stadler (2020). Fast likelihood calculation for multivariate gaussian phylogenetic models with shifts. *Theoretical Population Biology 131*, 66–78.

Suchard, M. A., P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, and A. Rambaut (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution 4*(1), vey016.