Supplementary information for: Simultaneously estimating evolutionary history and repeated traits phylogenetic signal: applications to viral and host phenotypic evolution.

Bram Vrancken[1], Philippe Lemey[1], Andrew Rambaut[2,3], Trevor Bedford[4], Ben Longdon[5], Huldrych Gunthard[6], and Marc A. Suchard[7,8]

[1]Department of Microbiology and Immunology, KU Leuven, Leuven, Belgium.
[2]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK.
[3]Fogarty International Center, National Institutes of Health, Bethesda, MD, USA.
[4]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.
[5]Department of Genetics, University of Cambridge, Cambridge, UK.
[6]Division of Infectious Diseases and Hospital Epidemiology, University Hospital of Zürich, University of Zürich, Zürich, Switzerland.
[7]Departments of Biomathematics and Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA 90095-1766, USA
[8]Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles, CA 90095-1766, USA

September 3, 2014

Additional Figure

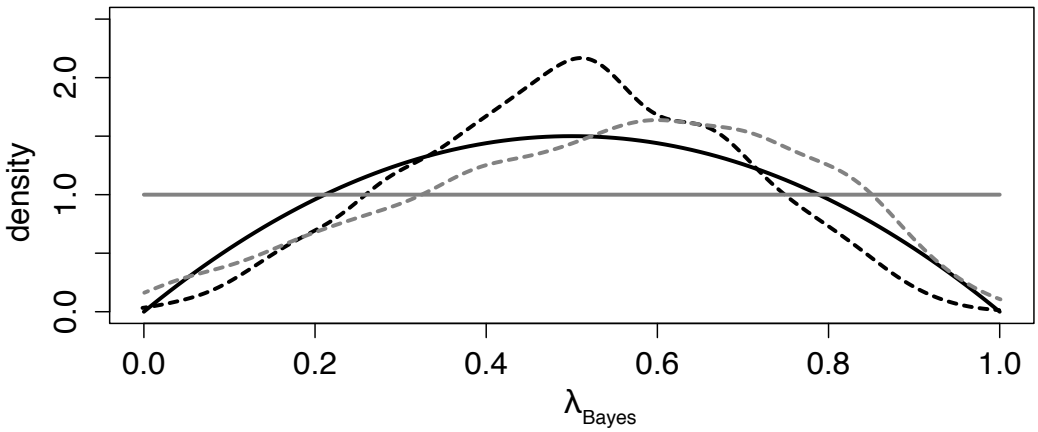

Figure 1: **Marginal prior and posterior densities of $\lambda_{\mathbf{B}}$ for the MSM strict data set.** The black lines indicate the prior (solid) and posterior (dashed) densities under a $\beta(2,2)$ prior, whereas the grey lines indicate the prior (solid) and posterior (dashed) density under a $\beta(1,1)$ prior.

Additional simulations

We complement the performance assessment of λ_{B} based on the approach proposed by Alizon *et al.* (2010) with an additional simulation study based on the approach taken by Shirreff *et al.* (2013). In the Methods section of the main manuscript, we describe how this approach differs from the previous simulation procedure. The simulations are performed along the maximum clade credibility tree inferred from the MSM strict analysis. The additional analyses essentially replicate the findings reported in the main manuscript: the λ_{B} estimator shows less bias and has lower variance than the λ_{ML} and Blomberg's K estimators of phylogenetic signal (Table 1). Again, the markedly larger variance of λ_{ML} is particularly noticeable at intermediate heritability values (Figure 2). Similarly, when comparing the effect of several prior settings on λ_{B} in reducing the sigmoidal pattern of phylogenetic signal estimates (Figure 3), we find the β(2,2) prior again shows higher coverage and lower variance than all other tested priors. However, the bias is slightly higher ($\sim$2%) than under the β(1,1) prior.

Table 1: Comparison of bias, mean squared error (MSE) and coverage for different phylogenetic signal estimators across a range of heritability values.

Estimator	bias			MSE	coverage
	total	$h^2 < 0.5$	$h^2 > 0.5$		
Blomberg's K	-0.400	-0.454	-0.346	0.190	NA
Pagel's λ_{ML}, MLE	0.050	-0.147	0.310	0.042	0.919
λ_{B} , $\beta(1.0, 1.0)$	0.000	-0.009	-0.004	0.020	0.869
λ_{B} , $\beta(2.0, 2.0)$	0.020	-0.068	0.018	0.017	0.954

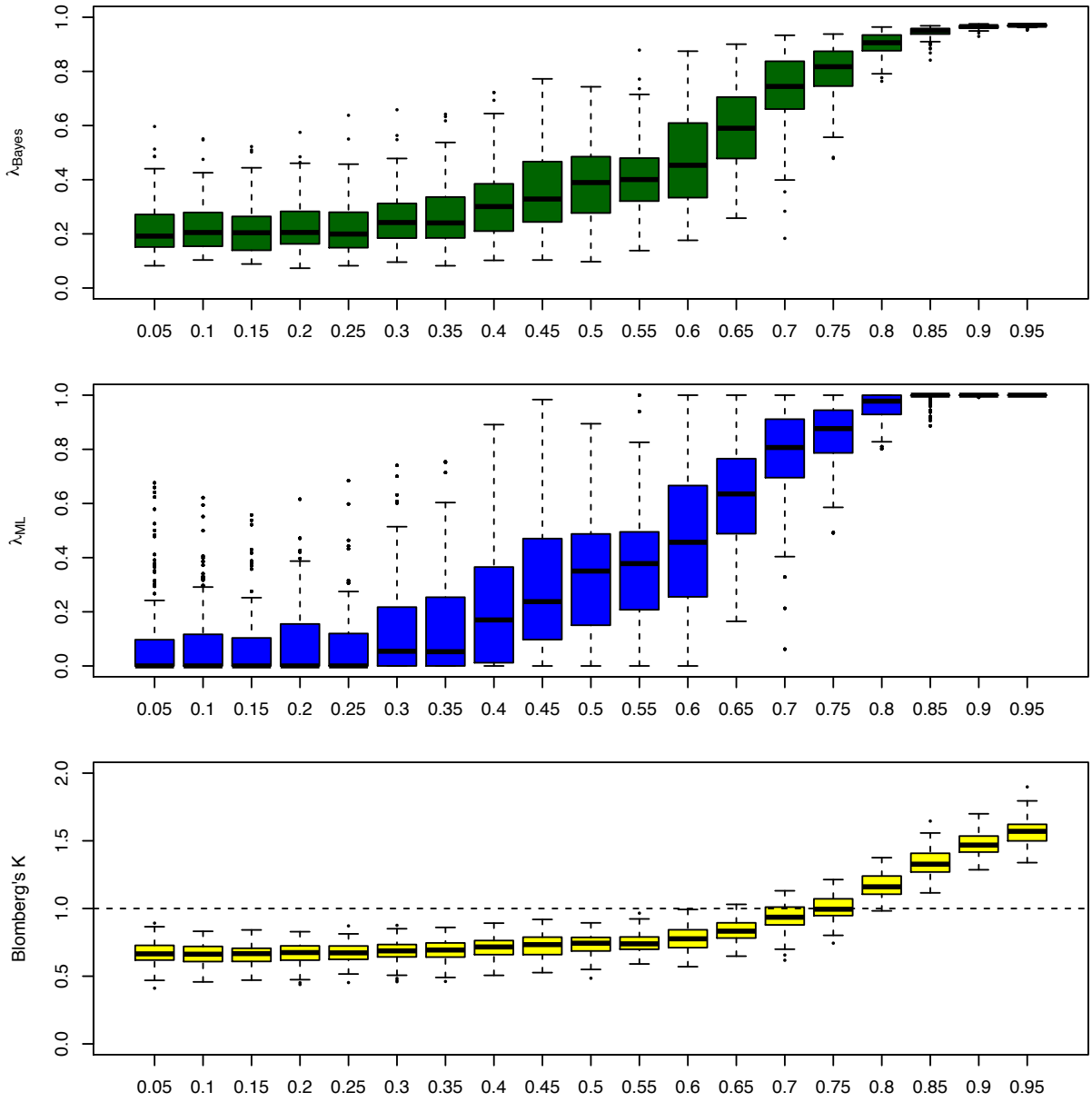

Figure 2: **Estimator performance of Blombergs K, Pagels $\lambda_{\mathbf{ML}}$ and $\lambda_{\mathbf{B}}$ on simulated data.** The evolution of an infection trait with known heritability (h^2= 0.05 to 0.95, stepsize 0.05) was modeled through 100 simulations of the trait on a fixed tree following the method of Shirreff *et al.* (2013). The box plots show the median values, the three quartiles and the outliers for Blombergs K (yellow), Pagels λ_{ML} (blue) and λ_{B} (green) estimates of phylogenetic signal. The dashed line in the plot for Blomberg's K indicates estimated phylogenetic signal equal to 1.

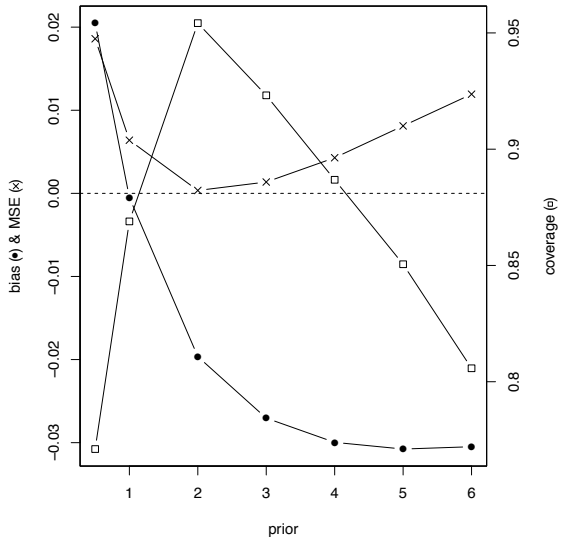

Figure 3: **Performance of the Bayesian phylogenetic signal estimation under various priors on $\lambda_{\mathbf{B}}$.** Various $\beta(\alpha,\beta)$ priors are explored for $\alpha = \beta$. We plot bias (filled black circles) and MSE (crosses) according to the primary axis and coverage (open squares) according to the secondary axis. The dotted horizontal line represents zero bias and MSE.

References

Alizon, S., von Wyl, V., Stadler, T., Kouyos, R.D., Yerly, S., Hirschel, B., Böni, J., Shah, C., Klimkait, T., Furrer, H., Rauch, A., Vernazza, P.L., Bernasconi, E., Battegay, M., Bürgisser, P., Telenti, A., Günthard, H.F., Bonhoeffer, S. & Swiss HIV Cohort Study (2010) Phylogenetic approach reveals that virus genotype largely determines hiv set-point viral load. *PLoS Pathogens*, **6**, e1001123.

Shirreff, G., Alizon, S., Cori, A., Günthard, H.F., Laeyendecker, O., van Sighem, A., Bezemer, D. & Fraser, C. (2013) How effectively can hiv phylogenies be used to measure heritability? *Evolution, Medicine and Public Health*, **2013**, 209–24.