Captions of supplementary figures

Figure S1. a) Frequency of false positives for three types of model comparison procedures. The count of each topology was drawn from a trinomial distribution with equal rate, which simulates samples from a single species, and model comparisons using AIC and the Bayesian modelling were conducted. AICs were calculated using the likelihood of eq.(1) for the single-species case and eq.(2) for the three-species case. Posterior probabilities were calculated with eq.(6) and eq.(7). The numbers of trials where the three-species case had larger AICs or smaller posterior probabilities than the single-species case were recorded as false positives. Abbreviations of models are: AIC_MF: AIC with the most frequent topology as dominant topology. AIC_Random: AIC with randomly chosen triplet as dominant topology. Bayes: Bayesian model comparison with six models.

Figure S2. Relationships between the required time for delimitation and the number of loci.

Figure S3. Average number of non-monophyletic species for each setting in the 10-species simulations.

Figure S4. A rooted triple consensus tree inferred from 18 nuclear loci of *Sistrurus* snakes. Polytomies were randomly resolved. Bars in dark grey indicate species delimited by the tr2. Bars in light grey show an alternative delimitation that appeared in the repeated delimitations. The numbers on nodes represent the average difference of log posterior probability scores between the null and alternative models defined by the nodes. A positive value indicates that the delimitation B (separate species) is preferred over the delimitation
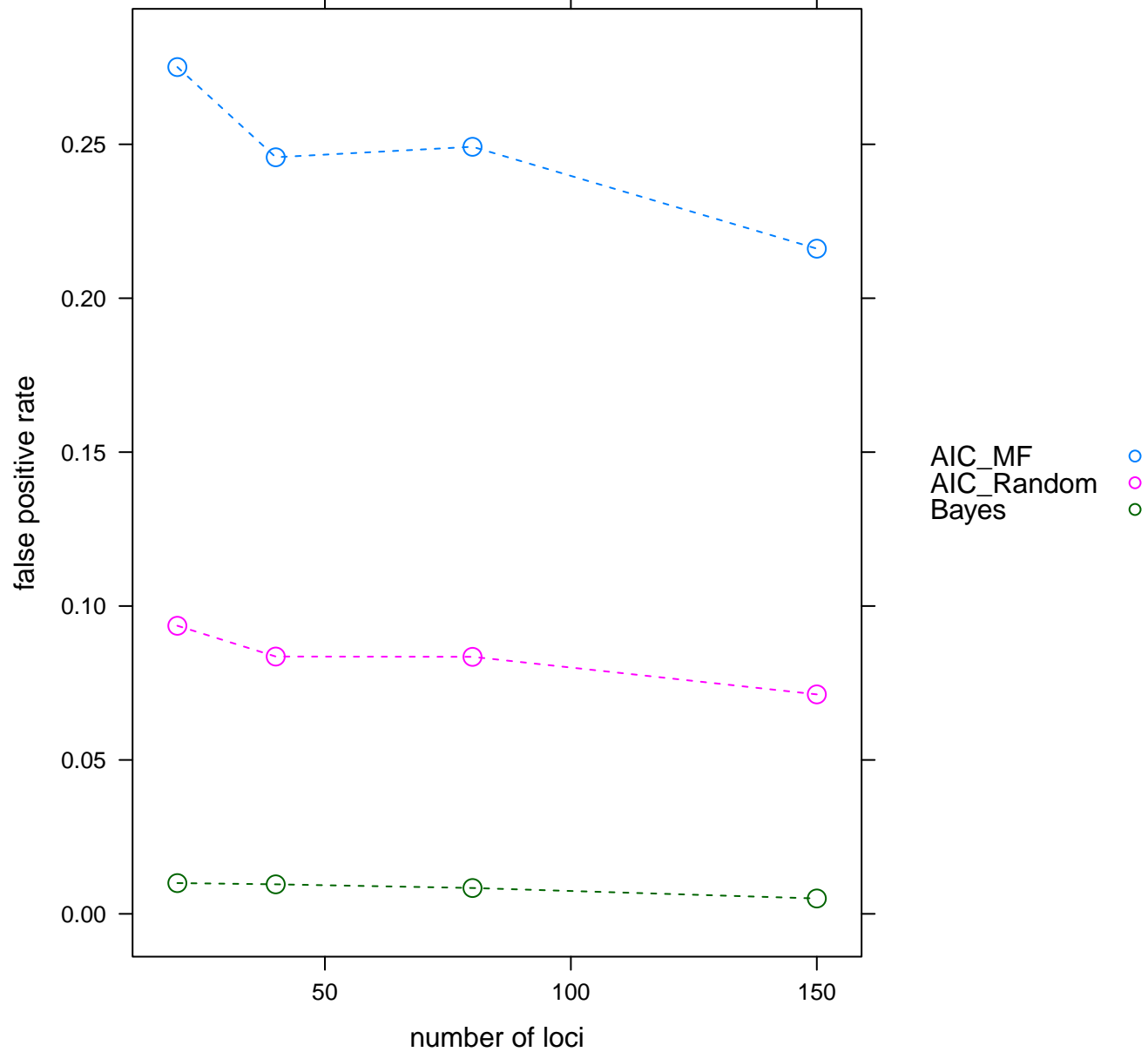
W (same species). Nodes indicated by asterisks are the most recent common ancestor nodes for species groups.

Figure S5. Results of delimitation of *Bacillus* with 100 sets of gene trees sampled from MCMC runs using rooted triple consensus as guide trees. Left) The rooted triple consensus built with seven MCC trees. Right) Frequency that each isolate is grouped by the tr2.
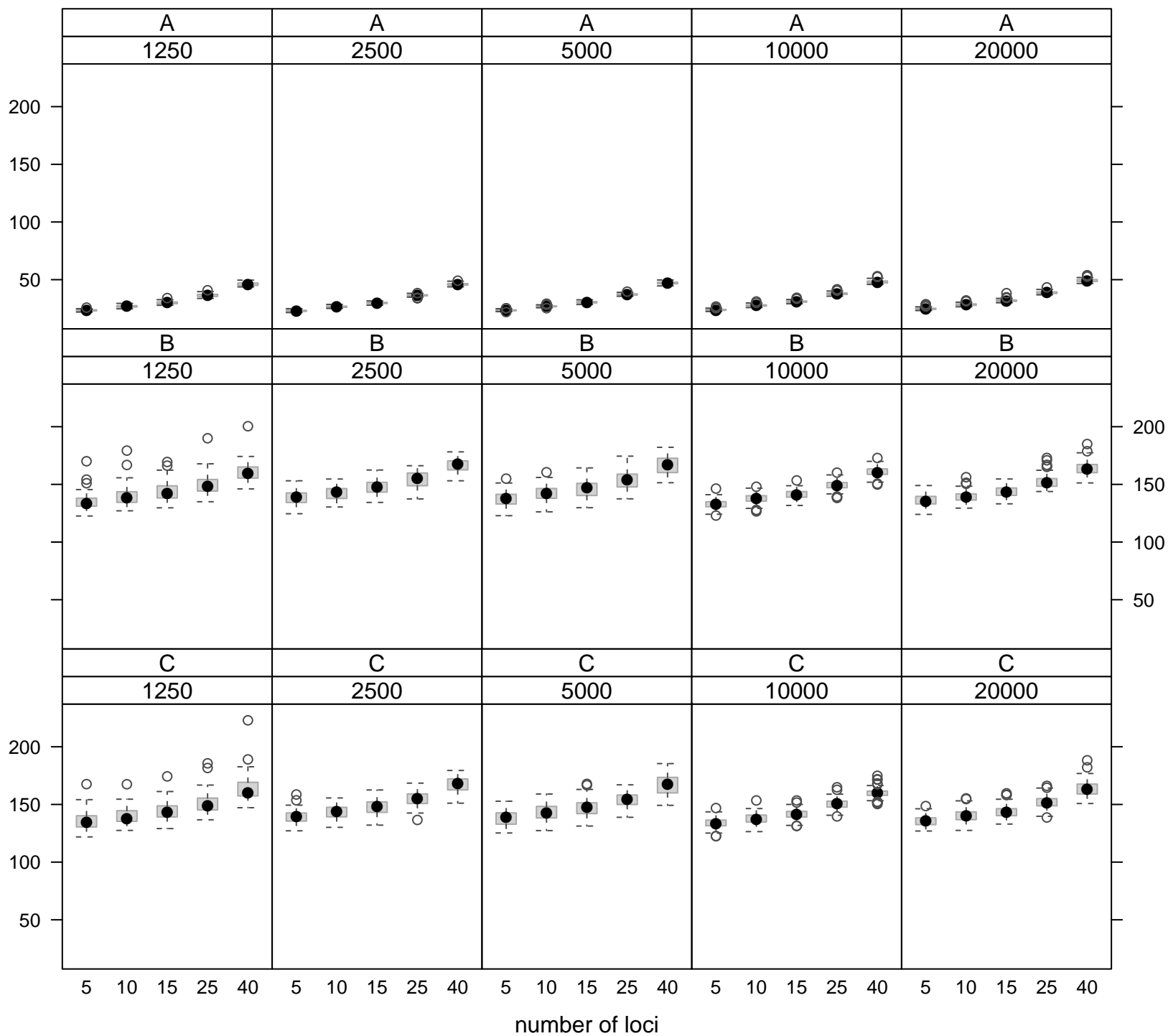
Figure S6. Distribution of linkage disequilibrium measured by R^2. Dotted grey lines indicate the border of loci and each grid cell represents the positions of each locus, from left hand side, glpF, gmk, ilv, pta, purH, pycA and tipD.

Figure S7. Relationship between the number of exact matches and total sample size (loci X individual samples). Dark grey: trials with sample size=50. Light grey: trials with sample size = 100.
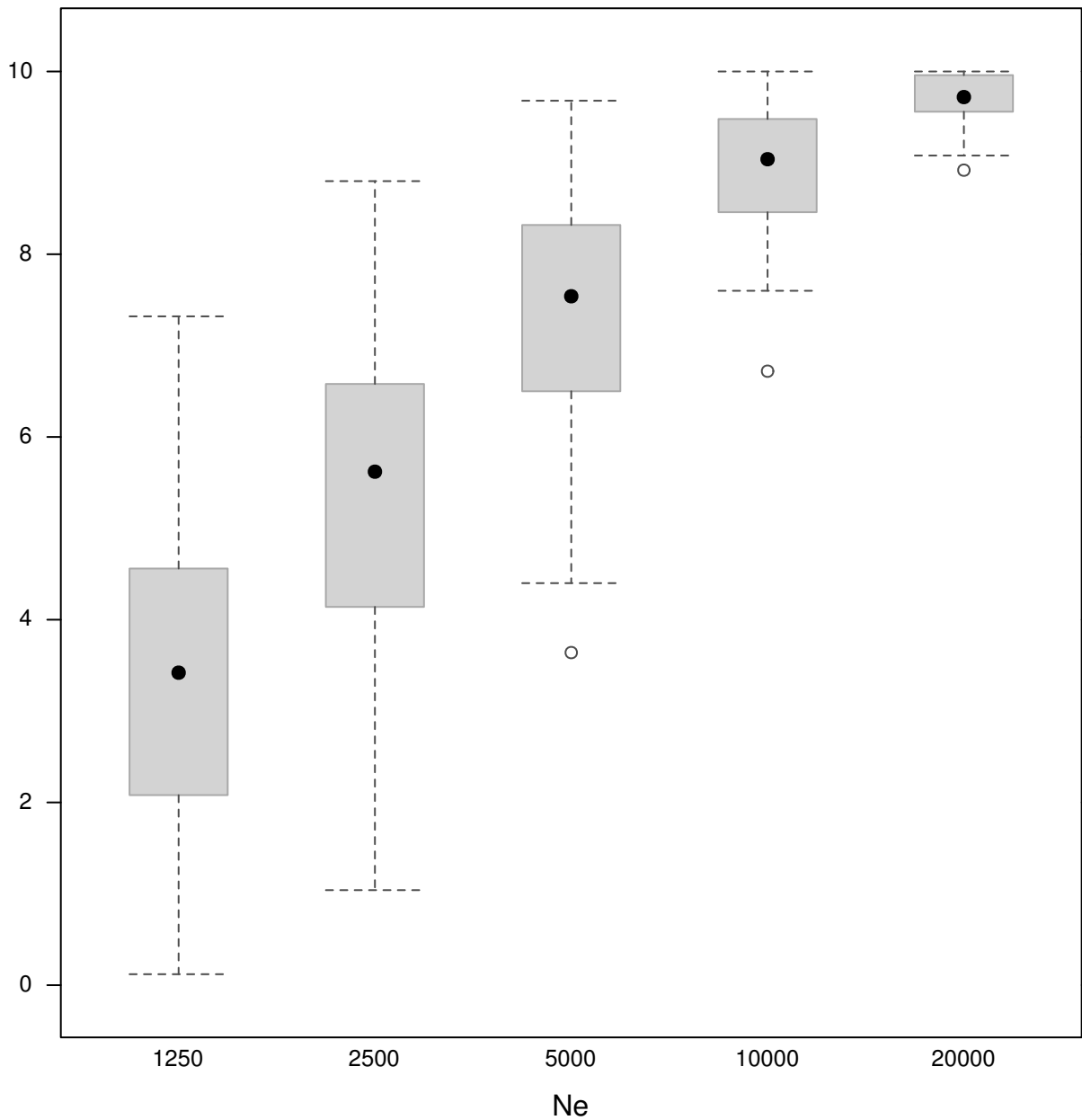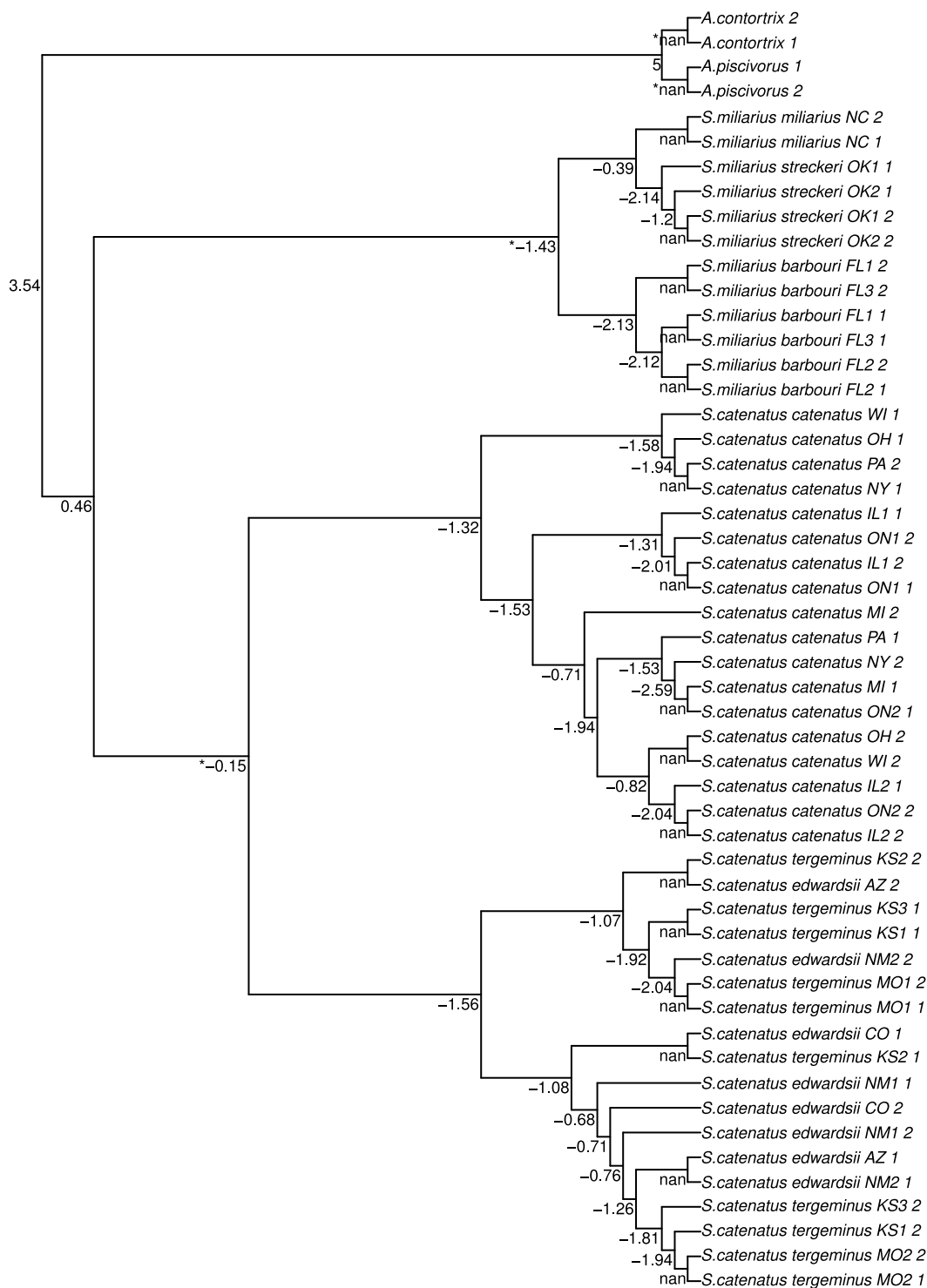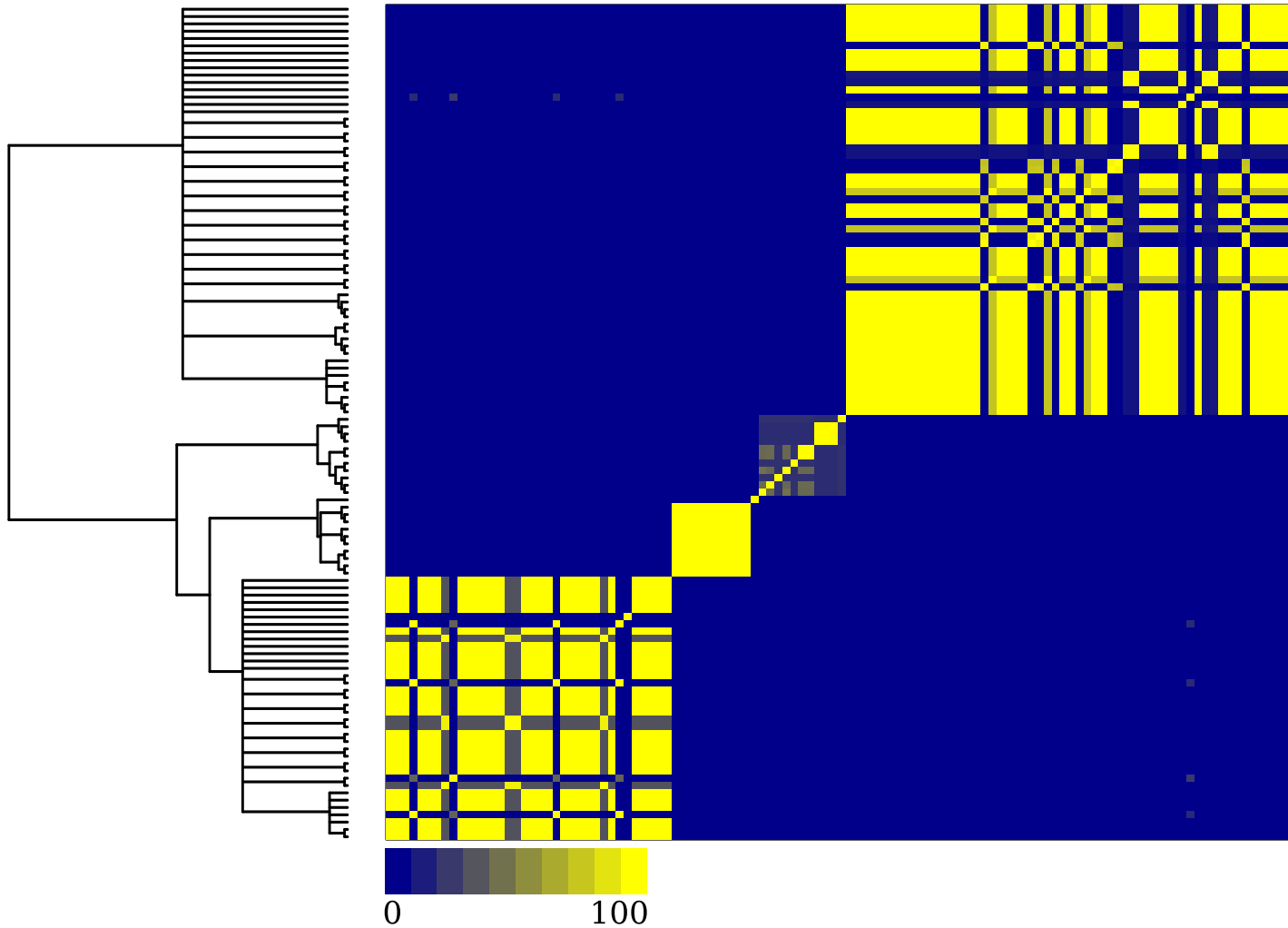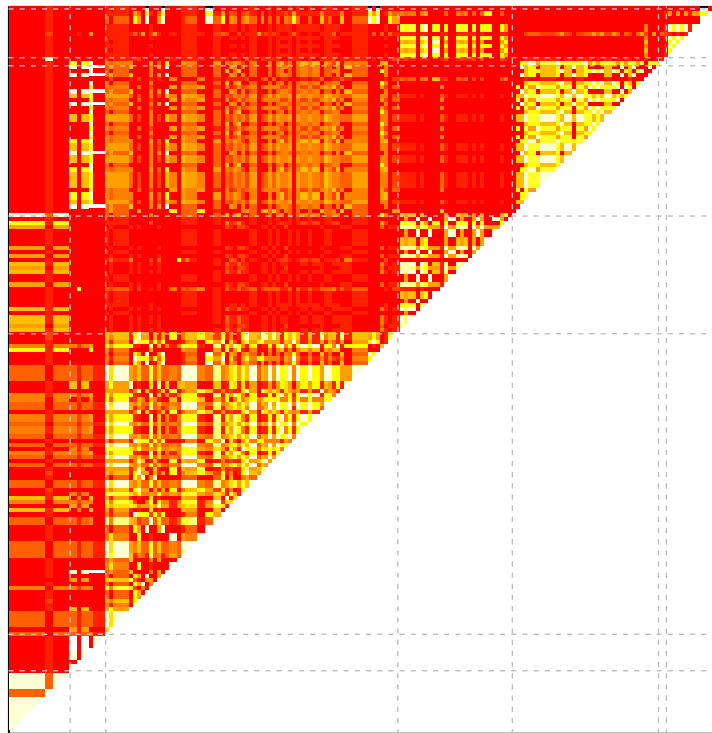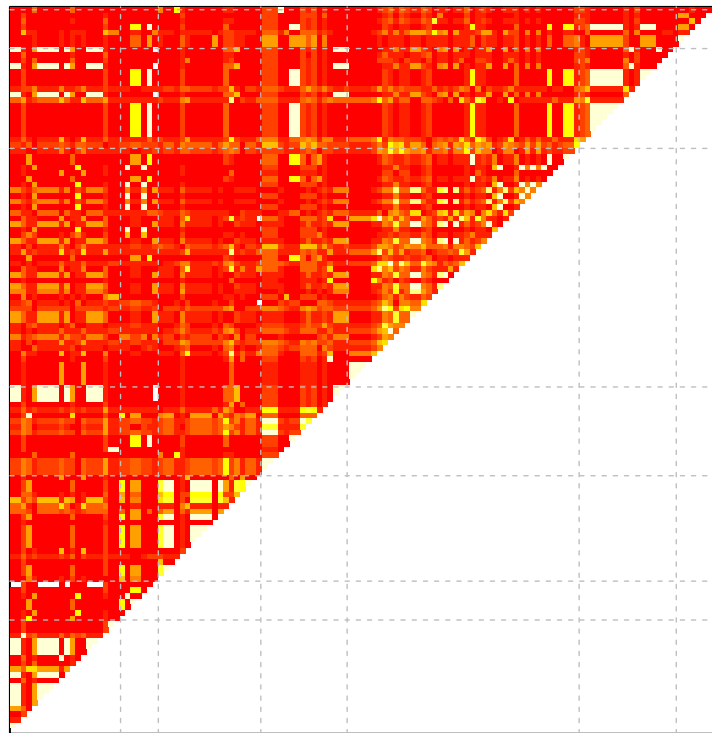
False posivite rates

Clade A   **Clade B**   **Clade C**

0   1