SM3 – Simulation Study of Genome Scans for Selection (FDR and Power)

We applied FDR control algorithms to three model-free hypothesis testing approaches to evaluate the sensitivity of methods to false discoveries in data sets simulated for discrete and continuous populations. The 3 methods included F_{ST} outlier approaches, linear and logistic regression models and latent factor models.

Discrete population simulations. To evaluate correction/calibration methods in genome scans based on population differentiation measures, we used the computer program `ms` to perform coalescent simulations of neutral and outlier SNP loci for discrete populations in 2-island models (Hudson, 2002). The justification for the use of Wright's island models is that loci with selection coefficient s, have an effectively reduced migration rate, m_{s}, which could be interpreted as evidence of divergent selection (Bazin *et al.* 2010). Simulated data sets included diploid individual genotypes for 100 samples from each island. Individual genotypes consisted of 1,000 SNPs, and the proportion of loci under selection was set to 10%. We used 3 values of the effective migration rate $4Nm = 5, 10, 20$ to simulate varying levels of neutral divergence. Outlier loci were generated using the values $4Nm_{\mathrm{s}} = .1$ and $.5$.

Continuous population simulations. We simulated 165 genotypes from populations that underwent a demographic range expansion 1,000 generations ago using the program SPLATCHE (Currat *et al.* 2004). The simulations implemented a non-equilibrium stepping-stone model based on a rectangular array of demes that was colonized from a geographic origin located south of the area. For each deme, the migration rate was equal to $m = .4$, the expansion rate was equal to $r = .4$, and the carrying capacity was equal to $C = 100$. We simulated genetic variation at 1,000 SNPs with 10% outlier loci, and we sampled one diploid individual from each deme. To simulate genetic variation at adaptive loci, we created artificial ecological gradients as described in (Frichot *et al.* 2015a). A reference ecological gradient was defined to be parallel to the main axis of expansion. Considering 17 distinct angles ranging from -90 to +90 degrees, the gradient

directions were rotated by angles of 11.25 degrees from their previous position. Using the Haldane transform, we reproduced clinal allele frequency patterns as expected under spatially varying selection intensities at 100 loci (Haldane 1948). Those data consisted of subsamples from the simulated data used in (Frichot *et al.* 2015a), where a full description of the simulation design is provided.

Estimation algorithms. In our analyses, we used the `R` functions `lm` (ANOVA), `glm` the base package and the function `lfmm` from the package LEA to estimate regression coefficients, z-scores and p-values for locus-specific effects (Frichot *et al.* 2015b). Considering squared z-scores, p-values were obtained from chi-square distributions with one degree of freedom. This value corresponded to the fact that 2 populations were used in PD methods, and that a single ecological variable was used in EA methods. Linear model and generalized linear model results were similar and reported under the term *linear model*.

In implementing `lfmm`, the Markov chain Monte Carlo procedure was executed five times using run-lengths of 10,000 cycles and burn-in periods of 5,000 cycles in each run. The resulting z-scores were then combined using a robust variant of the Stouffer method (Brown 1975, Whitlock 2005, see Text). For each locus, we computed the median values of z^2 scores, and corrected this statistic using the genomic inflation factor. We chose the number of factor corresponding to the best estimate of the number of cluster from the sNMF ancestry estimation program (Frichot *et al.* 2014), and displayed the histogram of the p-values for graphical checks. For PD methods using LFMM, the degrees of freedom were set to $d = K - 1$ where K is the number of genetic cluster estimated from the data ($d = 1$).

FDR and power study. For ANOVA, linear regression models and LFMM, we produced lists of candidate loci based on the FDR control algorithm (Benjamini and Hochberg 1995). In ANOVA and generalized linear regression models, we observed that the classical method for computing the genomic inflation factor, λ, as the median of z-scores divided by 0.456 was unpractical, and led to tests lacking power to detect selection. For each simulated data set, we optimized λ so that observed FDRs were close to their expected values. This procedure, which is based on the

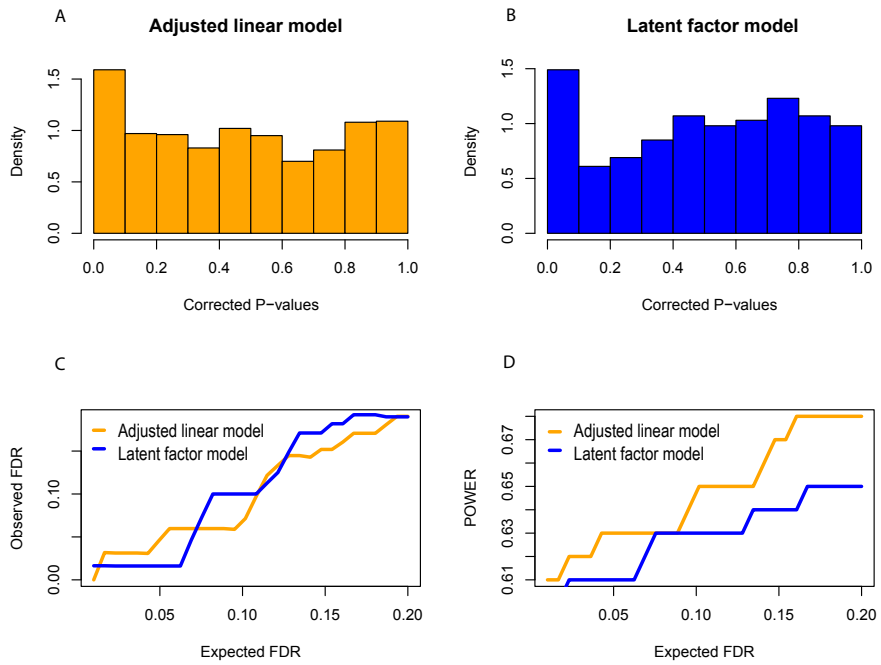

Figure 1: **Population differentiation tests.** FDR control in a particular simulated example data set ($4mN = 10$, $4m_sN = .1$). A) Histogram of corrected test p-values in linear models (optimal correction). B) Histogram of corrected test p-values in latent factor models ($K = 1$). C) Observed FDR as a function of their expected levels for the optimal linear model and for LFMM. D) Power as a function of expected FDR levels.

inspection of candidates obtained by the Benjamini-Hochberg algorithm and the evaluation of the number of adaptive loci among the candidates, provides an optimal control of the FDR in linear models. The resulting tests were termed *adjusted linear model* tests. This procedure is not applicable to real cases, because it assumes that the lists of truly selectively neutral and adaptive loci are known. Here this approach was used to obtain an empirical bound on the power of methods to detect selection. Better approximations of λ than the genomic inflation factor could be obtained by using graphical methods based on histograms of p-values or on the application of local FDR and empirical-null testing methods to z-scores. For LFMM, we followed the approach described in the documentation of the LEA package, and described in the previous paragraph. We eventually evaluated the sensitivity (power) of each method by computing the proportion of loci with positive tests among the list of truly adaptive loci.

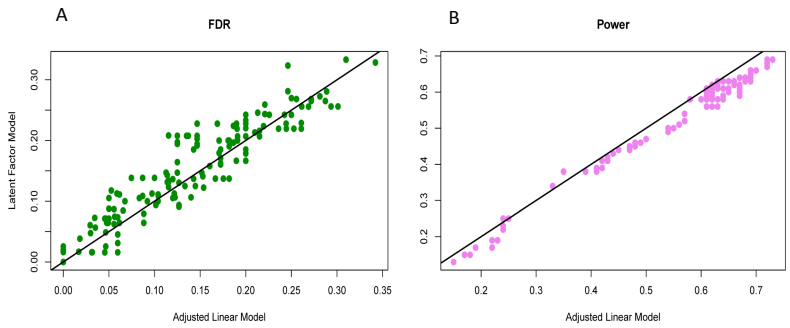

Figure 2: **Population differentiation tests (All data sets)**. Relative performances of latent factor models and optimally adjusted linear models. A) Observed FDRs. B) Power. These values corresponded to expected FDR levels between 0.01 and 0.3. Optimal test corrections in linear models used the list of truly adaptive loci.

Results for discrete population simulations. To evaluate the performances of LFMM and optimally tuned ANOVA tests, we first compared the observed levels of FDR to their expected levels using simulations of discrete populations. Corrected ANOVA tests used inflation factors that were optimized to produce the best match to their expected levels of FDR (Figure 1C), which resulted in a flat histogram for test p-values (Figure 1A). Without optimization, latent factor models exhibited levels of observed FDR close to the values observed for optimally adjusted linear models (Figure 2A). The results show that latent factor models, which do not compute any direct measures of population differentiation, can automatically correct for the inflation of test statistics due to population structure (Figure 1C and Figure 2A). Next, we compared the power of linear and latent factor models to detect outlier loci. Latent factor models exhibited performances that were close to those of optimally adjusted population differentiation methods (Figure 2B and Figure 1D). Power was generally higher for lower values of the ratio m_{s}/m, corresponding to higher relative levels of selection intensity. Overall, those results provide evidence that EA methods can reach quasi-optimal performances when they are used to perform population differentiation tests

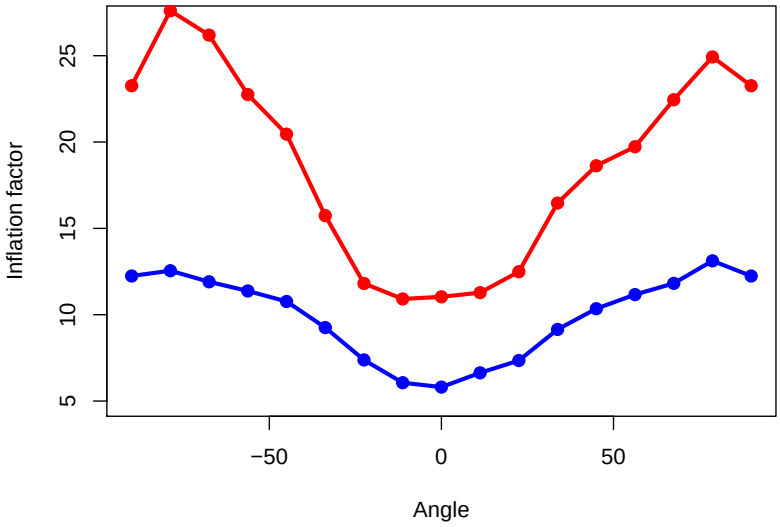

Figure 3: **Genomic inflation factors over-estimated optimal inflation factors in linear models**. Genomic inflation factors for 17 simulated continuous population genetic data sets (red curve). Optimal correction factors for the same data sets (blue curve). Optimal correction factors assumed that the list of truly adaptive loci was available.

in genome scans for selection.

Results for continuous population simulations. In a second series of simulations, we investigated FDR control for linear and latent factor models in the case of spatially continuous populations and association with environmental gradients. For linear models, the genomic inflation factor over-estimated inflation due to population structure, and led to overly conservative tests. Values of inflation factors for which tests matched their expected FDR levels were about two-fold lower than values computed by genomic inflation factors (Figure 3). Latent factor models exhibited levels of observed FDR close to the values observed for optimally adjusted linear models, and the LFMM tests had power close to optimally adjusted linear model tests (Figure 4). Figure 5 provides evidence of FDR control in linear and latent factor models for one particular data set.

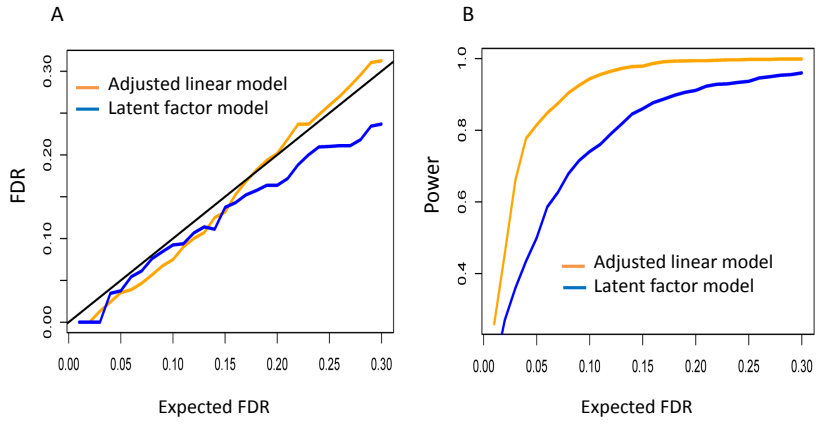

Figure 4: Continuous population models and association tests: Relative performances of latent factor models and optimally adjusted linear models. A) Observed FDRs. B) Power. These values correspond to expected values of FDRs between 0.01 and 0.3. Optimal test corrections in linear models made use of the list of truly adaptive loci.

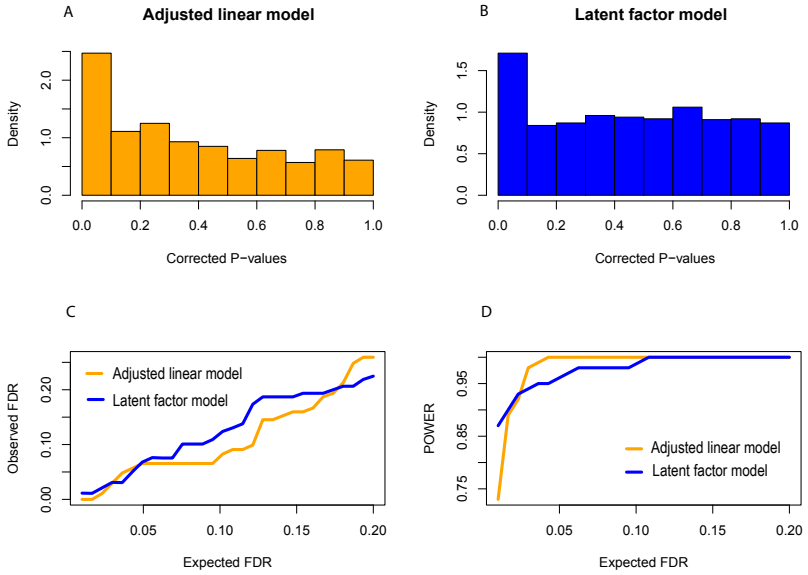

Figure 5: Continuous population models and association tests: FDR control in a particular simulated example data set (angle of 45 degrees). A) Histogram of corrected test p-values in linear models. B) Histogram of corrected test p-values in latent factor models ($K = 6$). C) Observed FDR as a function of their expected values. D) Power as a function of expected FDR values.

References

Bazin E, Dawson KJ, Beaumont MA (2010). Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics* 185(2):587-602.

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* 57: 289-300.

Brown M (1975). A method for combining non-independent, one-sided tests of significance. *Biometrics* 31 (4):987-992.

Currat M, Ray N, Excoffier L (2004). SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Molecular Ecology Notes* 4:139-142.

Hudson RR (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18(2):337-338.

Frichot E, Mathieu F, Trouillon T, Bouchard G, François O (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics* 196:973-983.

Frichot E, Schoville SD, de Villemereuil P, Gaggiotti OE, François O (2015a). Detecting adaptive evolution based on association with ecological gradients: Orientation matters! *Heredity* 115:22-28.

Frichot E, François O (2015b). LEA: an R package for landscape and ecological association studies. *Methods in Ecology and Evolution* 6 (8): 925-929.

Haldane JBS (1948). The theory of a cline. *Journal of Genetics* 48:277-284.

Whitlock MC (2005). Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *Journal of Evolutionary Biology* 18(5): 1368-1373.