# Flexible multivariate linear mixed models for structured multiple traits

Hyeonju Kim[1], Gregory Farage[1], John T. Lovell[2], John K. Mckay[3], Thomas E. Juenger[4], & Śaunak Sen[1]

[1]*Division of Biostatistics, Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, Tennessee, USA.*

[2]*Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA.*

[3]*Department of Bioagricultural Sciences and Graduate Degree Program in Ecology, Colorado State University, Fort Collins, Colorado, USA.*

[4]*Department of Integrative Biology, University of Texas, Austin, Texas, USA.*

**Many genetic studies collect structured multivariate traits that have rich information about the traits encoded in trait covariates, in addition to the genotype and covariate information on individuals. Examples of such data include gene-environment studies where the same genotype/clone is measured in multiple enviroments, and longitudinal studies where a measurement is taken at multiple time points. We present a flexible multivariate linear mixed model suitable for QTL (Quantitative Trait Loci) analysis of structured multivariate traits (flxQTL). Our model can incorporate low- and high-dimensional trait covariates to test the genetic association across structured multiple traits while capturing the correlations due to individual-to-individual similarity measured by genome-wide markers and trait-to-trait sim-**

**ilarity measured by trait covariates.**

Linear mixed models (LMMs) are widely used for genetic mapping; an important feature is their ability to adjust for confounding due to genetic relatedness among individuals. They are used in genomewide association studies (GWAS) [1–7] to identify genetic loci contributing to quantitative traits of interest (QTL). Multivariate LMMs (MLMMs) enhance statistical power over univariate LMMs [8] because they can accumulate signals across traits [2,9] with a common genetic locus. Many multivariate traits are structured. For example, a trait (such as body weight) might be measured weekly; crop yield or biomass may be measured in multiple environments, and gene expression may be measured in different brain regions of heterogeneous rats. In each of these cases, we attain trait covariates that provide information on the trait – for body weight the trait covariate would be time, for crop yield or biomass we might have information on the soil and/or weather conditions at the site, and for gene expression one could extract distance information depending on brain regions of heterogeneous rats.

Multiple traits with spatial or temporal structure in the context of LMMs have been explored in plant and animal genetics [10–14]. For these models, parameter estimation is a multidimensional optimization problem; there is a risk of reaching suboptimal solutions, and it is computationally expensive when fitting multiple traits [3,8]. Existing algorithms using an expectation-maximization (EM) method followed by second order schemes such as Newton-Rapson (NR), average information (AI), etc. provide stable, fast convergence [8,13] (GCTA [15], ASReml [16], WOMBAT [17], GEMMA [8]). However, they are not suitable for managing such traits measured in finite time

points or environments, and the time and environment combination because some of them [15,16] even cannot handle correlation among multiple traits in variance components, and none of them have fixed trait covariates to account for such trait-to-trait structures. GEMMA, WOMBAT may be able to deal with multiple traits without spatial or temporal information but would eventually slow down in estimating a fixed effects matrix and two variance-covariance matrices.

We now introduce a flexible multivariate linear mixed model, using high- and low-dimensional trait covariates, which can test the association between genetic marker(s) and trait(s) (**Supplementary Software** and `https://github.com/hkim89/flxQTL`). Our algorithm, flxQTL, is built on a MLMM that combines an expectation-conditional maximization (ECM) [18] with a Speed restarting Nesterov's accelerated gradient method which requires no second derivative like the NR, AI methods, etc. but provides stable linear convergence [19]. Coupled with the algorithmic efficiency, it reduces the computational burden significantly by introducing low- and high-dimensional trait covariates in a fixed and a random term on the MLMM, respectively. flxQTL is highly adaptable in that one can analyze varied trait structures such as a trait across multi-environments, multi-traits, and time-valued traits for plant, mouse, and human data, etc. with genotype (probability) data of multiple crosses. Our algorithm is implemented in Julia [20], a high-performance programming language, which enables users to add as many (even virtual) cores as possible for additional computational speed. The users take advantage of a high-performance computing (HPC) system to achieve a significant increase in speed or/and to do genome scan with millions of markers, so that practical high-dimensional data analysis is made possible (Online Methods).

The MLMM in flxQTL is modeled by two variance components in which the Kronecker product of two relatedness matrices, i.e. individual-to-individual covariance (genetic relatedness) and trait-to-trait covariance (or climatic relatedness), captures background genotype and trait structures such as infinitesimal random GxE effects (**Supplementary Note**), and that of unstructured environmental (or trait) errors and iid (independently and identically distributed) noise takes all levels of residual errors into account while (low-dimmensional) fixed trait covariates can capture gene by environment interactions (gene by multi-traits, or gene by trait over time, etc.):

$$Y = XBZ' + R + E. \tag{1}$$

Let $Y_{n\times m}$ be a matrix for a trait measured in $n$ individuals across $m$ environments (or over $m$ time points), or phenotypes of $n$ individuals for $m$ traits; $X_{n\times p}$ be $p$ genotype probabilities (or genotype including the intercept) of a marker to be tested without (or with) covariates; $Z_{m\times q}$ be $q$ (low-dimensional) fixed trait covariates such as contrasts, (orthonomal) basis functions, etc. [10] with $B$ effect sizes, $R$ be random effects, and $E$ be residual errors. We use likelihood ratio tests (LRT) to test associations between a marker and a multivariate trait of interest. We express the LRT as a LOD score (taking the log base 10 of the LRT). Cutoffs are obtained from the null distribution of maximum LOD scores by permutation test (Online Methods).

The algorithm, flxQTL, is marked by simplifying parameter estimation by its model structure and the choice of numerical methods. The low-dimensional trait covariates [10] reduce the matrix of fixed effect sizes from $p \times m$, which needs to be estimated in conventional multivariate LMMs [2,8], to $p \times q$ for $p$ genotypes or genotype probabilities in a genetic marker to be tested, $q$ low-dimensional trait covariates (or environmental covariates), and $m$ environments (or sites). The

trait covariates play a substantial role when $m$ gets large. The high-dimensional covariates, moreover, reduce a $m \times m$ random effects covariance matrix to a constant, i.e. overall genetic variance among environments, due to the climatic relatedness matrix (or trait-to-trait covariance) computed from annotated climatic information, etc. With an unstructured environment error (or model error) matrix, only $pq+1+\frac{m(m+1)}{2}$ parameters in total are estimated rather than $pm+m(m+1)$ ($q << m$). flxQTL has been tested with up to 30 traits or a trait across 30 environments. Unlike other second order approximations such as the NR method, the algorithm no longer needs to compute the second derivative to the objective function since the Nesterov's accelerated gradient method is a first order approximation modeled by a second order differential equation. It only requires the first order approximation, here ECM, followed by the acceleration step using a particularly tuned momentum coefficient, which behaves like the second derivative, to update parameters (**Supplementary Note**); therefore, the time complexity of the ECM embedded in the Nesterov's accelerated gradient method with speed restart is simpler than that of the other second order approximations, but its speed is faster than the other first order approximations (Online Methods).

The algorithm, flxQTL, is flexible in handling various trait data, as well as genotype (probability) data of multiple crosses. Not only can the fixed trait covariates be trait contrasts to capture GXE interactions, but they can be orthonormal basis functions, such as B-splines, wavelet, Fourier, etc., to account for a trait trend across the environmental gradient or a trend by elapsed time. In this way, flxQTL can cope with high dimensional traits better than existing MLMM based algorithms by decreasing degrees of freedom for testing. When independence across environments or traits is assumed, the climatic relatedness matrix can be replaced by an identity matrix with which genome

scan runs slightly faster. For genotype data, one can also use genotype probability data for mulitple crosses, so that the information loss could be minimal in QTL analysis.

To evaluate the statistical performance of flxQTL, we simulated data sets similar to the structure of *Arabidopsis thaliana* data [21]. Phenotypes of a trait measured in a combination of 3 years by 2 sites, which can be considered as 6 quantitative traits since *A. thaliana* is an annual species, were simulated by adding genetic and environmental effects as a site by year combination (**Supplementary Figure** 1). We carried out two comprehensive simulation studies on contributions of the climatic relatedness matrix and low-dimensional trait covariates, respectively. A comparison of powers with and without the climatic relatedness was made when including the same trait contrasts (Supplementary Table 1, **Supplementary Figure** 2, 3). We next compared powers with and without low-dimensional trait covariates keeping those climatic relatedness matrices the same (**Supplementary Figure** 4). The detailed data processing and simulation study can be found in Online Methods. The thresholds for the type I error rate were calibrated by the distribution of maximum LOD scores under the null model of existing multiplicative gene and environment random effects. The power was measured under the alternative model based on the thresholds. The overall result showed that including climatic relatedness yielded only subtle differences in power. Analyzing real data leads to the analogous result that genome scans using different climatic information such as daily range soil or air temperature, precipitation, etc. produced similar LOD scores one another, which coincided with those when assuming no climatic relatedness (i.e. the identity relatedness) as the simulation study showed (**Supplementary Figure** 6). In assessing the low-dimensional trait covariates, for each overall genetic variance among environments vary-

ing fixed effect sizes, the power was the highest when including low-dimensional trait covariates (**Supplementary Figure** 4).

Further, we tested flxQTL with empirical data from *A. thaliana*, where the trait is the mean number of fruits per seedling planted in two distinct regions, Sweden and Italy, from July, 2009 to June, 2012. Based on the study to examine the genetic basis of adaptive differentiation between the two populations, both main and interaction effects attained from genome scan in flxQTL are interpretable and by and large are in accordance with the existing results [21]. Further analysis reveals one more significant candidate QTL in three chromosomes each but one less in two chromosomes each, so that we have detected one more QTL than the existing result (The data processing and analysis is detailed in Online Methods, **Supplementary Table** 3,4 and **Supplementary Figure** 5, 14).

We assessed the performance of flxQTL in comparison with GEMMA (v.0.98) on Mouse HS1940 data available with the GEMMA package. GEMMA can implement only up to 3 traits, so that we could make a comparison for up to three traits. We set the low-dimensional covariates and climatic relatedness to be identity matrices in flxQTL since they are not supported by GEMMA. For a fair comparison, we measured computation time per SNP rather than that by completing a whole genome scan due to different dropping criteria for correlated and bad SNPs between the two softwares (Data processing and analysis procedure in Online Methods). flxQTL is about as fast as GEMMA for 3 traits although $-\log_{10} P$ values from flxQTL were slightly smaller than those from GEMMA in some places or vice versa, but this difference was negligible. Notice that estimating the single-valued variance in flxQTL is the mean of diagonal elements in the variance component

in GEMMA (**Supplementary Note**). We also compared the Julia-version MLMM with GEMMA, both of which give rise to the similar results each other.

To gauge the performance of the low-dimensional trait covariates on real data, we analyzed a trait of weekly measured body weights after birth to 16 weeks in 1212 F2-intercrosses between Gough Island mice and WSB/EiJ to investigate genetic basis of adaptation between them [22]. Using a 4 degrees of freedom cubic spline matrix as low-dimensional trait covariates requires neither smoothing step to estimate the first derivative for weekly growth rates nor additional analysis for the weekly growth rates, both of which are essential steps in univariate LMM and existing MLMM algorithms to describe the trait trend for 16 weeks. The result from single-QTL analysis run by flxQTL is overall consistent with the published one by R/qtl [23] that operates under the univariate framework except for Chromosome 6 and 7 perhaps due to polygenic effects. Since a SNP in Chromosome 7 is known to be crucial, we implemented 2-dimensional genome scan for Chromosome 7, as well as Chromosome 8, 10 for comparison. The results showed flxQTL detected several more QTL than R/qtl did despite the former requiring further analysis such as multiple-QTL analysis to finalize them (Online Methods and **Supplementary Figure** 15, 19).

Overall, our results demonstrate how the model and the corresponding algorithm can effectively deal with higher dimensional traits with spatial or temporal sturcture and run fast operation while retaining decent accuracy for parameter estimation. On current simulations and real data analysis, flxQTL using (low-dimensional) fixed trait covariates increases power compared with existing algorithms while leveraging information from climatic relatedness generated by various

annotated environmental information (high-dimensional trait covariates) including an independent case (i.e. identity relatedness matrix) turns out to be insensitive to such environmental information, which suggests genome scan using the identity matrix offer fast data analysis to have an idea for further analysis.

Our algorithm is not free from limitations. The model assumes a joint multivariate normal distribution, i.e. the sum of two Kronecker products of covariance matrices, which requires fully observed genotype and phenotype data for efficient inference [2,8]. If genotype, phenotype data contain missing values, one possibly takes advantage of any software package containing imputation methods for genotype data [23,24], phenotype data [25]. The statistical inference relies on LOD scores and permutation test, which are better than using $-\log_{10} P$ from an asymptotic $\chi^2$-distribution, when the pattern of linkage distribution is complex or extensive. The computational burden of permutations will be notable when matrix dimensions of data get considerably large even working with multi-nodes. It is possible to develop an efficient threshold estimation scheme [26] for such data. Like other existing MLMM based algorithms, flxQTL cannot fit three- or higher-dimensional trait structure, for instance, multiple traits measured over time, across multiple sites, or their combination. flxQTL may be able to fit the multiple traits across sites by juxtaposing each multiple traits into a large matrix and manipulating the fixed trait covariates with contrasts or B-spline basis functions to account for sites. This would eventually require a great deal of physical memory and intensive computation. Possibly one can extend the model to a multi-linear model which performs multi-way array (tensor) decompositions for dimension reduction. With the growth of more complex and high-throughput phenotyping, we expect the use of our models to help shed light on the

gentic control of structured multivariate traits.

**Acknowledgements** The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No DE-AC02-05CH11231. We thank Jon Ågren and Grey Monroe for providing climate data from the *Arabidopsis thaliana* reciprocal transplant experiments. *A.thaliana* fitness: https://datadryad.org/stash/downloads/file_stream/72691 F2 intercrosses between Gough Island mice and WSB/EiJ: https://phenome.jax.org/projects/Payseur1

**Competing Financial Interests** The authors declare that they have no competing financial interests.

**Author Contributions**

## ONLINE METHODS

**Code availability.** A Julia software implementation of flxQTL is available at https://github.com/hkim89/flxQTL.

**flxQTL.** Let $Y_{n\times m}$ be a trait for $n$ individuals across $m$ environments (or over $m$ time points), or phenotypes of $n$ individuals for $m$ traits. $X_{n\times p}$ is $p$ genotypes including the intercept or genotype probabilities of a marker to be tested without (or with) covariates, $Z_{m\times q}$ is $q$ low-dimensional fixed trait covariates such as treatments (or contrasts), (orthonormal) basis functions, etc [10]. flxQTL is then constructed as follows:

$$Y = XBZ' + R + E. \tag{2}$$

One can observe that the size of $B$ is $p \times q$ and is smaller than that of $B$ in conventional MLMMs, where $Z = I_m$, $q = m$. This dimension reduction is easily discovered ($q << m$) when working with higher dimensional trait data, so as to resolve degrees-of-freedom problems in regard with such data for hypothesis testing.

Two independent random effects follow multivariate normal distributions, such that

$$R \sim MVN(0, \tau^2 K_C, K_G), \quad E \sim MVN(0, \Sigma, I_n). \tag{3}$$

In the first variance component, $K_G$ is individual-to-individual similarity (a genetic relatedness or kinship matrix) computed by background genetic markers. $K_C$ is trait-to-trait similarity (climatic relatedness) generated by high-dimensional trait covariates such as annotated climatic and/or environmental information in regard with multiple sites or environments. The information could be, for instance, maximum, minimum, average, or daily range temperature, precipitation, soil or its related data, etc. The Kronecker product of the two relatedness matrices produces infinitesimal random effects; for infinitesimal GxE random effects, it describes all levels of associations between QTL and multiple environments simultaneously. $\tau^2$ is an overall genetic variance among environments. In the second, $\Sigma$ is unstructured environment errors, which assumes to be correlated among environments, but assumes iid noise between individuals. The two relatedness matrices are precomputed from corresponding data sets, so that only constant genetic variance and environment error matrix are estimated. The multi-level GxE random structure is verified in **Supplementary Note**. The MLMM in flxQTL is then modeled by multivariate normal distribution whose covariance matrix is the sum of two Kronecker products indicating two independent random effects:

$$Y^v \sim N((X \otimes Z)B^v, \tau^2 K_C \otimes K_G + \Sigma \otimes I_n). \tag{4}$$

LOD scores are calculated by log-likelihood difference between the full model and null, scaled by negative common logarithm. At least 1000 permutations are implemented to obtain empirical LOD thresholds from the null distribution of maximal LOD scores in following steps: transforming the correlated trait to the iid by dividing corresponding individual-wise variance-covariance matrices, shuffling the iid trait by individuals, transforming back to the correlated trait, and then scanning genetic markers to pick up a maximal LOD score for each permutation. Note that LOD scores using low-dimensional trait covariates are lower than those implemented by conventional MLMMs (i.e. no low- dimensional covariates included), but this is not a problem since the cutoff will be adjusted in permutations.

*Computational costs.* When it comes to algorithmic implementation, the eigen decomposition is applied to the two relatedness matrices only once for computational efficiency, which requires $O(n^3 + m^3) \approx O(n^3)$, $(m << n)$ as existing MLMMs. A LOCO (Leave One Chromosome Out) scheme [27],[1],[28],[3] also works well because $O(dn^3) \approx O(n^3)$ $(d << n)$ for $d$ chromosomes in genotype data, but in practice, genome scan with the LOCO scheme takes longer depending on types of data (about 30 % longer ) than that without LOCO. For $t_0 >> t$, where $t_0, t$ are the maximal number of iterations of ECM and ECM embedded in Speed restarting Nesterov's scheme, respectively, the computational complexity of two step implementation that runs ECM for finding 'good' initial values followed by ECM embedded in Speed restarting Nesterov's scheme is $O((c_1qn^2 + (m^3 + c_1m^2)n)(t_0 + t))$ per marker for $n$ individuals, $c_1$ covariates. The whole time complexity including the eigen decomposition and rotation by corresponding eigen vectors is now $O(n^3 + (m + c_1)n^2 + s(c_1qn^2 + (m^3 + c_1m^2)n)(t_0 + t))$ with $s$ markers.

*Julia implementation.* For faster genome scan with sufficiently numerous markers, one can add large number of cores in multi-nodes such as the HPC system launching Julia. Julia then distributes each core to a marker to do genome scan independently. One can also employ MKL.jl, a Julia software package, to easily switch the BLAS library to the Intel MKL library in Julia for faster runtime on Intel hardware.

**Data processing and analysis procedure.** *Mouse data.* For mouse_HS1940 data, we only selected markers (SNPs) in Chromosome 1 for genome scan and computed a centered genetic relatedness matrix by leaving Chromosome 1 information out (LOCO) (1458 SNPs from total 12226 SNPs) in GEMMA to use the same genetic kinship matrix for a fair comparison due to no LOCO option for the MLMM in GEMMA. We filtered one of two SNPs out for identical two consecutive SNPs and excluded SNPs having zero variances to eliminate a singularity problem. In total, 950, 967 SNPs in Chromosome 1 (1940 individuals) for GEMMA, flxQTL respectively were employed for comparison since the dropping criteria for correlated or bad SNPs differ from ours, and GEMMA starts genome scan with processing SNPs, which is not under our control to figure out which SNPs were dropped. Three traits were imputed using R-mice package with a 'pmm' option (predicted mean method). For analysis, flxQTL ran genome scan by adding 8 cores in Julia since GEMMA performs genome scan based on maximum 8 threads, which cannot be customized by users.

For F2 intercrosses between Gough Island mice and WSB/EiJ, we computed genotype probabilities from R/qtl adding pseudo markers in every 1cM between two markers while excluding sex chromosomes for minimal loss of information in QTL analysis instead of using genotypes. Traits

were imputed using R-mice package with a ‘pmm’ option and then were standardized by overall mean and standard deviation because of their time-valued measures. We used 4 degrees of freedom cubic spline from R-splines package to produce low-dimensional trait covariates ($Z$), which can capture the trends of weekly body weights and corresponding weight growth rates simultaneously. Here, we set the climatic relatedness to be an identity matrix since GxE interactions are not appropriate in this case and computed a linear genetic relatedness matrix from genotype probabilities due to its statistical interpretability. The thresholds at $\alpha = 0.1, 0.05$ were estimated by 1000 permutations.

*Arabidopsis thaliana data.* We employed R/qtl to add pseudo markers in every 1cM between two markers and dropped one of two consecutive markers when they are identical. Phenotypes of a trait were imputed using a ‘pmm’ option in R-mice package and were standardized by columns since the plant is annual. The data comprises 699 markers (or SNPs) across 5 chromosomes by 400 lines (or individuals) for 6 traits. A matrix ($Z$) sets to be trait (orthonormal) contrasts, which consist of overall site (1’s) and site difference (1’s for Sweden and -1’s for Italy), so that multiplying the site difference by QTL yields QTL$\times$site or GxE interactions. We computed both a genetic and a climatic relatedness matrices using Manhattan distance; thus, subtracting a marker-distance mean between two individuals from 1 yields a probability of genetic distance between two individuals. The climatic relatedness matrix was computed in a similar fashion. However, the difference between with and without climatic relatedness matrix appears to be minimal in this analysis. We performed 1-, 2-dimensional QTL analyses with the LOCO option followed by multiple QTL analysis, i.e. a stepwise model selection approach by forward selection and backward

elimination adding or dropping one or two SNPs for each scan, respectively. The 95% of cutoff was estimated by permutation test and was used for a penalty term in the stepwise model selection since the false positive rate is maintained at the rate of $\alpha$ in the case of no QTL and with the search restricted to models with no more than one QTL [29].

**Simulation study.** We generated simulated data sets of a quantitative trait for 3 years and 2 sites (or 6 quantitative traits) by simulating random GxE effects based on genotype data and annotated daily soil minimum and maximum data from the Arabidopsis data [21]. We rearranged soil daily range temperatures (difference between maximum and minimum temperatures) into a 3 year by 2 location combination, i.e. high-dimensional trait covariates of 365 days by 6 environments, and standardized them by overall mean and standard deviation to keep their inherent information unchanged. To generate null trait data sets, we selected to standardize about 3.5% of genetic markers (about 25 markers of total 699 markers) and 8.2% of 365-day soil daily range temperatures (about 30 days), both of which constitute multiplicative GxE random effects while the overall variance among environments ($\tau^2$) varies from small to large. The remaining errors indicating unmeasured environmental errors and iid noise were also added to the null trait data. Both random and error terms were respectively standardized. To add large effect QTL to the null data, we sampled one genetic marker apart from the selected many small effect QTL and generated large effect QTL multiplied by fixed effect sizes that vary from small to large and a (orthonormal) contrast matrix ($Z$) mentioned in the **Data processing and analysis procedure** (**Supplementary Note**). Here we only considered one large QTL effect fixed when simulating data sets in order to eliminate complications that may arise for including an additional QTL. We carried out 1000 simulations

under the null to attain a threshold at $\alpha = 0.05$ from the distribution of maximum LOD scores and measured power under the alternative for each instance given the climatic relatedness. The first simulation study to assess power for the effect of climatic relatedness varied with values of the overall genetic variance among environments ($\tau^2$) for each fixed effect size matrix ($B$), which in total resulted in 100,000 experiments. We then narrowed down to the feasible range of fixed effect sizes by excluding the effect sizes producing zeros and ones of power. In this range of fixed effect sizes, we ran simulations to establish a comparison between inclusion and exclusion of the climatic relatedness matrix ($K_C$) changing $\tau^2$. Since corresponding results showed slight distinctions in power, we tackled unusual scenarios in genome scan, where $K_C$'s are an autocorrelated matrix, an extreme case of positive definite matrix, etc. The results revealed analogous power to those shown in prior experiments. The second simulation study to assess the effect of low-dimensional trait covariates ($Z$) was implemented in the similar manner. In the given range of fixed effect sizes, we compared three cases: contrasts (flxQTL), an identity matrix ($Z = I$ in flxQTL), Julia-version MLMM. Here, the Julia-version conventional MLMM was coded as an alternative of GEMMA to use as a reference for no low- and high-dimensional trait covariates (i.e. an identity trait covariates) because both run by different algorithms under the same model while yielding very similar results. It is also easier to draw a comparison with the Julia-version MLMM than to do with GEMMA since GEMMA uses different data formats, criteria on excluding correlated or bad markers, etc.

1. Lippert, C. *et al.* Fast linear mixed models for genome-wide association studies. *Nature methods* **8**, 833 (2011).

2. Casale, F. P., Rakitsch, B., Lippert, C. & Stegle, O. Efficient set tests for the genetic analysis of correlated traits. *Nature methods* **12**, 755–758 (2015).

3. Loh, P.-R. *et al.* Efficient bayesian mixed-model analysis increases association power in large cohorts. *Nature genetics* **47**, 284 (2015).

4. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics* **11**, 459–463 (2010).

5. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nature genetics* **42**, 348 (2010).

6. Yu, J. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics* **38**, 203–208 (2006).

7. Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nature genetics* **42**, 355 (2010).

8. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature methods* **11**, 407–409 (2014).

9. Korte, A. *et al.* A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature genetics* **44**, 1066 (2012).

10. Xiong, H. *et al.* A flexible estimating equations approach for mapping function-valued traits. *Genetics* **189**, 305–316 (2011).

11. Garin, V., Malosetti, M. & van Eeuwijk, F. Multi-parent multi-environment qtl analysis: an illustration with the eu-nam flint population. *TAG. Theoretical and Applied genetics. Theoretische und Angewandte Genetik* (2020).

12. Smith, A., Cullis, B. & Thompson, R. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* **57**, 1138–1147 (2001).

13. Smith, A. B., Cullis, B. R. & Thompson, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *The Journal of Agricultural Science* **143**, 449–462 (2005).

14. Cullis, B. R., Smith, A. B., Beeck, C. P. & Cowling, W. A. Analysis of yield and oil from a series of canola breeding trials. part ii. exploring variety by environment interaction using factor analysis. *Genome* **53**, 1002–1016 (2010).

15. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. Gcta: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88**, 76–82 (2011).

16. Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J. & Thompson, R. Asreml user guide release 1.0 (2002).

17. Meyer, K. Wombata tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (reml). *Journal of Zhejiang University Science B* **8**, 815–821 (2007).

18. Meng, X. L. & Rubin, D. B. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80**, 267–278 (1993).

19. Su, W., Boyd, S. & Candes, E. A differential equation for modeling Nesterov's accelerated gradient method: Theory and Insights. In *Advances in Neural Information Processing Systems*, 2510–2518 (2014).

20. Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. Julia: A fresh approach to numerical computing. *SIAM review* **59**, 65–98 (2017).

21. Ågren, J., Oakley, C. G., McKay, J. K., Lovell, J. T. & Schemske, D. W. Genetic mapping of adaptation reveals fitness tradeoffs in arabidopsis thaliana. *Proceedings of the National Academy of Sciences* **110**, 21077–21082 (2013).

22. Gray, M. M. *et al.* Genetics of rapid and extreme size evolution in island mice. *Genetics* **201**, 213–228 (2015).

23. Broman, K. W. & Sen, Ś. *A Guide to QTL Mapping with R/qtl*, vol. 46 (Springer, 2009).

24. Broman, K. W. *et al.* R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics* **211**, 495–502 (2019).

25. Buuren, S. v. & Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in r. *Journal of statistical software* 1–68 (2010).

26. Feingold, E., Brown, P. O. & Siegmund, D. Gaussian models for genetic linkage analysis using complete high-resolution maps of identity by descent. *American journal of human genetics* **53**, 234 (1993).

27. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nature genetics* **46**, 100 (2014).

28. Listgarten, J. *et al.* Improved linear mixed models for genome-wide association studies. *Nature methods* **9**, 525 (2012).

29. Manichaikul, A., Moon, J.-Y., Sen, Ś., Yandell, B. S. & Broman, K. W. A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. *Genetics* **181**, 1077–1086 (2009).

30. Henderson, C. R. Applications of linear models in animal breeding. *University of Guelph Press, Guelph* **11**, 652–653 (1984).

31. Ledoit, O. & Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis* **88**, 365–411 (2004).

32. Schäfer, J. & Strimmer, K. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology* **4** (2005).

## 1 Supplementary Note: flxQTL for gxe detection

Condsider a multivariate linear mixed model [30],

$$Y_0 = X_0 B Z_0' + R_0 + E_0, \tag{5}$$

or its vectorized form,

$$Y_0^v = (Z_0 \otimes X_0)B^v + R_0^v + E_0^v,$$

where

$$R_0^v \sim MVN(0, \tau^2 K_C \otimes K_G), \quad E_0^v \sim MVN(0, \Sigma_0 \otimes I_n). \tag{6}$$

$Y_0$ is a matrix of $n \times m$ multivariate reponses (phenotypes or a trait in multiple environments), where $n$ is the number of individuals(or lines), $m$ is the number of environments(or sites). $X_0$ is $n \times p$ genotypes (or genotype probabilities) for a candidate marker including intercepts and/or covariates. $Z_0$ is $m \times q$ fixed column covariates in regard with environments (or sites). $B$ is $p \times q$ fixed effects including main and interaction effects. $R_0$ is $n \times m$ gene by environment random effects. $E_0$ is $n \times m$ genetic variability among individuals that is not explained by $R$ and residual errors in which $I_n$, $\Sigma_0$ are respective $n \times n$ indentity and $m \times m$ matrices. $\tau^2$ is an overall genetic variance among environments. $K_G$ is a $n \times n$ genetic background (or a genetic kinship) matrix, and $K_C$ is a $m \times m$ environmental background (or weather profiles) matrix. Examples of $C$ could be air, soil temperatures, water availiability (or draught) through growing season profile, or soil microbial composition. Note that we confine the relatedness matrices to full-rank and symmetric positive definite matrices. For a semi-positive definite relatedness matrix, one could take advantage of a Shrinkage approach [31,32] to force it to be positive definite without loss of information. Thus, $K_G = U_G \Lambda_G U_G'$ and $K_C = U_C \Lambda_C U_C'$. Multiplying $U_C' \otimes U_G'$ to the model yields

$$Y = XBZ' \quad + \quad R + E, \tag{7}$$

$$R^v \sim MVN(0, \tau^2 \Lambda_C \otimes \Lambda_G), \qquad E^v \sim MVN(0, \Sigma \otimes I_n), \tag{8}$$

where $Y_{n\times m} = [\mathbf{y}_1', \mathbf{y}_2', \cdots, \mathbf{y}_n']'$, $X_{n\times p} = [\mathbf{x}_1', \mathbf{x}_2', \cdots, \mathbf{x}_n']'$, $R_{n\times m} = [\mathbf{r}_1', \mathbf{r}_2', \cdots, \mathbf{r}_n']'$, and $E_{n\times m} = [\mathbf{e}_1', \mathbf{e}_2', \cdots, \mathbf{e}_n']'$.

The corresponding individual vector is now,

$$\mathbf{y}_i = ZB'\mathbf{x}_i + \mathbf{r}_i + \mathbf{e}_i, \;\; \mathbf{r}_i \sim N(0, \tau^2\lambda_i\Lambda_C), \;\; \mathbf{e}_i \sim N(0, \Sigma), \tag{9}$$

and its variance is $V_i = \tau^2\lambda_i\Lambda_C + \Sigma$ $(i = 1, \ldots, n)$.

**Verification of infinitesimal gxe random terms** Assume that a gene by environment random term has a form of $R_0 = GB_{rand}C'$, where $C_{m\times q_1}$ is a $q_1$ random climate (or weather) factor matrix, $G_{n\times p_1}$ is a $p_1$ background genotype matrix, and $B^v_{rand} \sim MVN(0, \tau^2 I_{q_1} \otimes I_{p_1})$. Then, the distribution of the gene by envrionment random effects in (6) is obtained in the following way.

$$\begin{aligned}
E(R_0^v R_0^{v\prime}) &= E((C \otimes G)B^v_{rand}{B^v_{rand}}'(C' \otimes G')) \\
&= (C \otimes G)E(B^v_{rand}{B^v_{rand}}')(C' \otimes G') \\
&= (C \otimes G)(\tau^2 I_{q_1} \otimes I_{p_1})(C' \otimes G') \\
&= \tau^2(CC' \otimes GG') \\
&= \tau^2 K_C \otimes K_G,
\end{aligned}$$

where C and G are scaled by $\sqrt{m}$, $\sqrt{n}$, respectively.

**Expectation Conditional Maximization (ECM) algorithm** In the ECM algorithm [18], the joint log-likelihood function, assuming the random term $R$ to be missing, is

$$l(Y, R|B, \tau^2, \Sigma)$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}\{-m\log(2\pi) - \frac{m}{2}\log\tau^2 - \frac{m}{2}\log(\lambda_i) - \frac{1}{2}\log|\Lambda_C| - \frac{1}{2}\log|\Sigma| \\
&\quad -\frac{1}{2}\mathbf{e}_i'\Sigma^{-1}\mathbf{e}_i - \frac{1}{2}\mathbf{r}_i'(\tau^2\lambda_i\Lambda_C)^{-1}\mathbf{r}_i\}. \qquad (10)
\end{aligned}
$$

In the expectation step, the conditional distribution of $\mathbf{r}_i$ given $\mathbf{y}_i$ ($i = 1, \cdots, n$), $B^{(t)}$,$\tau^{2(t)}$, and $\Sigma^{(t)}$ at the current iteration $t$ is [30]

$$
\mathbf{r}_i|\mathbf{y}_i, B^{(t)}, \tau^{2(t)}, \Sigma^{(t)} \sim MVN(\hat{\mathbf{r}}_i^{(t)}, \hat{\Theta}_i^{(t)}), \qquad (11)
$$

where

$$
\begin{aligned}
V_i^{(t)} &= \tau^{2(t)}\lambda_i\Lambda_C + \Sigma^{(t)}, \quad \hat{\mathbf{r}}_i = \tau^{2(t)}\lambda_i\Lambda_C(V_i^{(t)})^{-1}(\mathbf{y}_i - ZB^{(t)\prime}\mathbf{x}_i), \\
\hat{\Theta}_i^{(t)} &= \tau^{2(t)}\lambda_i\Lambda_C - \tau^{2(t)}\lambda_i\Lambda_C(V_i^{(t)})^{-1}\tau^{2(t)}\lambda_i\Lambda_C. \qquad (12)
\end{aligned}
$$

The expectation of the loglikelihood function in terms of $R$ given $Y$, $B^{(t)}$,$\tau^{2(t)}$, and $\Sigma^{(t)}$ is obtained as follows.

$$
\begin{aligned}
&E\left(l(Y, R|B, \tau^2, \Sigma)|R|Y, B^{(t)}, \tau^{2(t)}, \Sigma^{(t)}\right) \\
&= \sum_{i=1}^{n}\{-m\log(2\pi) - \frac{m}{2}\log\tau^2 - \frac{m}{2}\log(\lambda_i) - \frac{1}{2}\log|\Lambda_C| - \frac{1}{2}\log|\Sigma| \\
&\quad -\frac{1}{2}(\mathbf{y}_i - ZB'\mathbf{x}_i)'\Sigma^{-1}(\mathbf{y}_i - ZB'\mathbf{x}_i) - \frac{1}{2}\hat{\mathbf{r}}_i^{(t)\prime}(\tau^{-2}\lambda_i^{-1}\Lambda_C^{-1} + \Sigma^{-1})\hat{\mathbf{r}}_i^{(t)} \\
&\quad +\hat{\mathbf{r}}_i^{(t)\prime}\Sigma^{-1}(\mathbf{y}_i - ZB'\mathbf{x}_i) - \frac{1}{2}tr((\tau^{-2}\lambda_i^{-1}\Lambda_C^{-1} + \Sigma^{-1})\hat{\Theta}_i^{(t)})\}. \qquad (13)
\end{aligned}
$$

In the conditional maximization step, one can update $B^{(t+1)}$ conditional on $\tau^{2(t)}$ and $\Sigma^{(t)}$, followed by updating $\{\tau^{2(t+1)}, \Sigma^{(t+1)}\}$ conditional on $B^{(t+1)}$, $\tau^{2(t)}$, and $\Sigma^{(t)}$ [8].

$$
\begin{aligned}
B^{(t+1)} &= (X'X)^{-1}X'(Y - \hat{R}^{(t)})\Sigma^{-1}Z(Z'\Sigma^{-1}Z)^{-1}, \qquad (14) \\
\tau^{2(t+1)} &= \frac{1}{nm}\sum_{i=1}^{n} tr(\lambda_i^{-1}\Lambda_C^{-1}(\hat{\mathbf{r}}_i^{(t)}\hat{\mathbf{r}}_i^{(t)\prime} + \hat{\Theta}_i^{(t)})), \qquad (15)
\end{aligned}
$$

$$\Sigma^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\mathbf{e}}_i^{(t)}\hat{\mathbf{e}}_i^{(t)\prime} + \hat{\Theta}_i^{(t)}), \tag{16}$$

$$\text{where} \quad \hat{\mathbf{e}}_i^{(t)} = \mathbf{y}_i - ZB^{(t+1)\prime}\mathbf{x}_i - \hat{\mathbf{r}}_i^{(t)}. \tag{17}$$

**Log-likelihood function** In the expectation step, the conditional distribution of $\mathbf{r}_i$ given $\mathbf{y}_i$ ($i = 1, \cdots, n$), $B^{(t)}$,$\tau^{2(t)}$, and $\Sigma^{(t)}$ at the current iteration $t$ is [30]

$$\begin{aligned}
& l(Y, R|B, \tau^2, \Sigma) = -nm\log(2\pi) \\
& -\frac{1}{2}\log|(U_C \otimes U_G)(\tau^2\Lambda_C \otimes \Lambda_G)(U_C' \otimes U_G')| - \frac{1}{2}\log|(U_C \otimes U_G)(\Sigma \otimes I_n)(U_C' \otimes U_G')| \\
& -\frac{1}{2}(Y_0 - X_0BZ_0' - R_0)^{v\prime}((U_C \otimes U_G)(\Sigma \otimes I_n)(U_C' \otimes U_G'))^{-1}(Y_0 - X_0BZ_0' - R_0)^v \\
& -\frac{1}{2}R_0^{v\prime}((U_C \otimes U_G)(\tau^2\Lambda_C \otimes \Lambda_G)(U_C' \otimes U_G'))^{-1}R_0^v \\
= \; & -nm\log(2\pi) - \frac{nm}{2}\log(\tau^2) - \frac{m}{2}\log|\Lambda_G| - \frac{n}{2}\log|\Lambda_C| - \frac{1}{2}\log|\Sigma \otimes I_n| \\
& +2\log|(U_C' \otimes U_G')| - \frac{1}{2}(Y - XBZ' - R)^{v\prime}(\Sigma \otimes I_n)^{-1}(Y - XBZ' - R)^v \\
& -\frac{1}{2}R^{v\prime}(\tau^2\Lambda_C \otimes \Lambda_G)^{-1}R^v \\
= \; & \sum_{i=1}^{n}\{-m\log(2\pi) - \frac{m}{2}\log\tau^2 - \frac{m}{2}\log(\lambda_i) - \frac{1}{2}\log|\Lambda_C| - \frac{1}{2}\log|\Sigma| \\
& -\frac{1}{2}\mathbf{e}_i'\Sigma^{-1}\mathbf{e}_i - \frac{1}{2}\mathbf{r}_i'(\tau^2\lambda_i\Lambda_C)^{-1}\mathbf{r}_i\}.
\end{aligned} \tag{18}$$

**Derivation of conditional expected log-likelihood and its closed form solutions** To obtain closed-form solutions of $B^{(t+1)}$, $\tau^{2(t+1)}$, and $\Sigma^{(t+1)}$ to the conditional expected loglikelihood in (13), one differentiates (13) with respect to each parameter after integrating out $\mathbf{r}_i$ ($i = 1, \ldots n$) in the joint log-likelihood function, (10).

In (10), taking expectation with respect to $\mathbf{r}_i$ ($i = 1, \cdots, n$) yields

$$\begin{aligned}
El &= E\left(l(Y, R|B, \tau^2, \Sigma)|R|Y, B^{(t)}, \tau^{2(t)}, \Sigma^{(t)}\right) \\
&= \sum_{i=1}^{n}\{-m\log(2\pi) - \frac{m}{2}\log\tau^2 - \frac{m}{2}\log(\lambda_i) - \frac{1}{2}\log|\Lambda_C| - \frac{1}{2}\log|\Sigma| \\
&\quad -\frac{1}{2}E\left((\mathbf{y}_i - ZB'\mathbf{x}_i - \mathbf{r}_i)'\Sigma^{-1}((\mathbf{y}_i - ZB'\mathbf{x}_i - \mathbf{r}_i)|R\right) - \frac{1}{2}E\left(\mathbf{r}_i'(\tau^2\lambda_i\Lambda_C)^{-1}\mathbf{r}_i|R\right)\}.
\end{aligned}$$

For simplicity, let $E(\cdot|R|Y, B^{(t)}, \tau^{2(t)}, \Sigma^{(t)}) = E(\cdot|R)$. The last two terms then can be simplified as follows:

$$\begin{aligned}
&\sum_{i=1}^{n}\{E(-\frac{1}{2}(\mathbf{y}_i - ZB'\mathbf{x}_i)'\Sigma^{-1}(\mathbf{y}_i - ZB'\mathbf{x}_i) + \mathbf{r}_i'\Sigma^{-1}(\mathbf{y}_i - ZB'\mathbf{x}_i) \\
&-\frac{1}{2}\mathbf{r}_i'\Sigma^{-1}\mathbf{r}_i - \frac{1}{2}\mathbf{r}_i'(\tau^2\lambda_i\Lambda_C)^{-1}\mathbf{r}_i|R)\} \\
= &\sum_{i=1}^{n}\{-\frac{1}{2}(\mathbf{y}_i - ZB'\mathbf{x}_i)'\Sigma^{-1}(\mathbf{y}_i - ZB'\mathbf{x}_i) + E(\mathbf{r}_i'|R)\Sigma^{-1}(\mathbf{y}_i - ZB'\mathbf{x}_i) \\
&-\frac{1}{2}E(\mathbf{r}_i'(\Sigma^{-1} + (\tau^2\lambda_i\Lambda_C)^{-1})\mathbf{r}_i|R)\} \\
= &\sum_{i=1}^{n}\{-\frac{1}{2}(\mathbf{y}_i - ZB'\mathbf{x}_i)'\Sigma^{-1}(\mathbf{y}_i - ZB'\mathbf{x}_i) + \hat{\mathbf{r}}_i^{(t)'}\Sigma^{-1}(\mathbf{y}_i - ZB'\mathbf{x}_i) \\
&-\frac{1}{2}tr((\Sigma^{-1} + (\tau^2\lambda_i\Lambda_C)^{-1})E(\mathbf{r}\mathbf{r}'|R))\}. \\
= &\sum_{i=1}^{n}\{-\frac{1}{2}(\mathbf{y}_i - ZB'\mathbf{x}_i)'\Sigma^{-1}(\mathbf{y}_i - ZB'\mathbf{x}_i) + \hat{\mathbf{r}}_i^{(t)'}\Sigma^{-1}(\mathbf{y}_i - ZB'\mathbf{x}_i) \\
&-\frac{1}{2}tr((\Sigma^{-1} + (\tau^2\lambda_i\Lambda_C)^{-1})(\hat{\mathbf{r}}_i^{(t)}\hat{\mathbf{r}}_i^{(t)'} + \hat{\Theta}_i^{(t)}))\} \\
= &\sum_{i=1}^{n}\{-\frac{1}{2}(\mathbf{y}_i - ZB'\mathbf{x}_i)'\Sigma^{-1}(\mathbf{y}_i - ZB'\mathbf{x}_i) + \hat{\mathbf{r}}_i^{(t)'}\Sigma^{-1}(\mathbf{y}_i - ZB'\mathbf{x}_i) \\
&-\frac{1}{2}tr((\Sigma^{-1} + (\tau^2\lambda_i\Lambda_C)^{-1})\hat{\Theta}_i^{(t)}) - \frac{1}{2}\hat{\mathbf{r}}_i^{(t)'}(\Sigma^{-1} + (\tau^2\lambda_i\Lambda_C)^{-1})(\hat{\mathbf{r}}_i^{(t)})\}.
\end{aligned}$$

Now one can find closed-form solutions of $B$, $\tau^2$, and $\Sigma$ to the expected log-likelihood function using properties of the trace, calculus for matrices. For the solution for $B$,

$$El = \sum_{i=1}^{n}\{-\frac{1}{2}(\mathbf{y}_i - ZB'\mathbf{x}_i)'\Sigma^{-1}(\mathbf{y}_i - ZB'\mathbf{x}_i) + \hat{\mathbf{r}}_i^{(t)'}\Sigma^{-1}(\mathbf{y}_i - ZB'\mathbf{x}_i)\} + C,$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}\{-\frac{1}{2}(\Sigma^{-\frac{1}{2}}\mathbf{y}_i - \Sigma^{-\frac{1}{2}}ZB'\mathbf{x}_i)'(\Sigma^{-\frac{1}{2}}\mathbf{y}_i - \Sigma^{-\frac{1}{2}}ZB'\mathbf{x}_i) \\
&\quad +(\Sigma^{-\frac{1}{2}}\hat{\mathbf{r}}_i^{(t)})'(\Sigma^{-\frac{1}{2}}\mathbf{y}_i - \Sigma^{-\frac{1}{2}}ZB'\mathbf{x}_i)\} + C,
\end{aligned}
$$

where $C$ is a constant.

$$
\frac{\partial El}{\partial B'^v} = \sum_{i=1}^{n}((\mathbf{x}_i' \otimes \Sigma^{-\frac{1}{2}}Z)'(\Sigma^{-\frac{1}{2}}\mathbf{y}_i - (\mathbf{x}_i' \otimes \Sigma^{-\frac{1}{2}}Z)B'^v) - (\mathbf{x}_i' \otimes \Sigma^{-\frac{1}{2}}Z)'\Sigma^{-\frac{1}{2}}\hat{\mathbf{r}}_i^{(t)}) = 0
$$

$$
\sum_{i=1}^{n}(\mathbf{x}_i\mathbf{x}_i' \otimes Z'\Sigma^{-1}Z)B'^v = \sum_{i=1}^{n}(\mathbf{x}_i \otimes Z'\Sigma^{-\frac{1}{2}})(\Sigma^{-\frac{1}{2}}\mathbf{y}_i - \Sigma^{-\frac{1}{2}}\hat{\mathbf{r}}_i^{(t)})
$$

$$
B'^v = ((X'X)^{-1}X' \otimes (Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-\frac{1}{2}})(\Sigma^{-\frac{1}{2}}(Y' - \hat{R}^{(t)'}))^v
$$

$$
B' = (Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1}(Y' - \hat{R}^{(t)'})X(X'X)^{-1}
$$

$$
\therefore B = (X'X)^{-1}X'(Y - \hat{R}^{(t)})\Sigma^{-1}Z(Z'\Sigma^{-1}Z)^{-1}
$$

$$
\frac{\partial El}{\partial \tau^2} = \sum_{i=1}^{n}\{-\frac{m}{2}\tau^{-2} + \frac{1}{2}\hat{\mathbf{r}}_i^{(t)'}(\tau^{-4}\lambda_i^{-1}\Lambda_C^{-1})\hat{\mathbf{r}}_i^{(t)} + \frac{1}{2}tr(\tau^{-4}\lambda_i^{-1}\Lambda_C^{-1}\hat{\Theta}_i^{(t)})\} = 0 \qquad (19)
$$

$$
mn = \sum_{i=1}^{n} tr((\tau^{-2}\lambda_i^{-1}\Lambda_C^{-1})\hat{\mathbf{r}}_i^{(t)}\hat{\mathbf{r}}_i^{(t)'} + (\tau^{-2}\lambda_i^{-1}\Lambda_C^{-1})\hat{\Theta}_i^{(t)})
$$

$$
mn\tau^2 = tr(\sum_{i=1}^{n}\lambda_i^{-1}\Lambda_C^{-1}(\hat{\mathbf{r}}_i^{(t)}\hat{\mathbf{r}}_i^{(t)'} + \hat{\Theta}_i^{(t)}))
$$

$$
\therefore \tau^{2(t+1)} = \frac{1}{nm}\sum_{i=1}^{n} tr(\lambda_i^{-1}\Lambda_C^{-1}(\hat{\mathbf{r}}_i^{(t)}\hat{\mathbf{r}}_i^{(t)'} + \hat{\Theta}_i^{(t)}))
$$

$$
\begin{aligned}
\frac{\partial El}{\partial \Sigma^v} &= \sum_{i=1}^{n}\{-\frac{1}{2}tr(\Sigma^{-1}\frac{\partial \Sigma}{\partial \Sigma^v}) + \frac{1}{2}(\mathbf{y}_i - ZB^{(t+1)'}\mathbf{x}_i)'(\Sigma^{-1}\frac{\partial \Sigma}{\partial \Sigma^v}\Sigma^{-1})(\mathbf{y}_i - ZB^{(t+1)'}\mathbf{x}_i) \\
&+ \hat{\mathbf{r}}_i^{(t)'}(\Sigma^{-1}\frac{\partial \Sigma}{\partial \Sigma^v}\Sigma^{-1})\hat{\mathbf{r}}_i^{(t)} + \frac{1}{2}tr(\Sigma^{-1}\frac{\partial \Sigma}{\partial \Sigma^v}\Sigma^{-1}\hat{\Theta}_i^{(t)}) \\
&- \hat{\mathbf{r}}_i^{(t)'}(\Sigma^{-1}\frac{\partial \Sigma}{\partial \Sigma^v}\Sigma^{-1})(\mathbf{y}_i - ZB^{(t+1)'}\mathbf{x}_i)\} = 0
\end{aligned}
$$

$$\sum_{i=1}^{n}\{tr(-I_m + \Sigma^{-1}(\mathbf{y}_i - ZB^{(t+1)'}\mathbf{x}_i)(\mathbf{y}_i - ZB^{(t+1)'}\mathbf{x}_i)' + \Sigma^{-1}\hat{\mathbf{r}}_i^{(t)}\hat{\mathbf{r}}_i^{(t)'} + \Sigma^{-1}\hat{\Theta}_i^{(t)} - 2\Sigma^{-1}(\mathbf{y}_i - ZB^{(t+1)'}\mathbf{x}_i)\hat{\mathbf{r}}_i^{(t)'})\} = 0$$

$$n \cdot tr(\Sigma) = \sum_{i=1}^{n} tr(\hat{\mathbf{e}}_i^{(t)}\hat{\mathbf{e}}_i^{(t)'} + \hat{\Theta}_i^{(t)})$$
$$\therefore \Sigma^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\mathbf{e}}_i^{(t)}\hat{\mathbf{e}}_i^{(t)'} + \hat{\Theta}_i^{(t)}),$$

where $\hat{\mathbf{e}}_i^{(t)} = \mathbf{y}_i - ZB^{(t+1)'}\mathbf{x}_i - \hat{\mathbf{r}}_i^{(t)}$.

**ECM embeded in Nesterov's accelerated gradient with speed restart**

**Algorithm** : ECM embeded in Speed Restarting Nesterov's Scheme [19]

**input** : $B^{(0)} \in \mathbb{R}^{q\times p}$, $\tau^{2(0)} \in \mathbb{R}$, $\Sigma^{(0)} \in \mathbb{R}^{m\times m}$, $B^{(-1)} = B^{(0)}$, $\tau^{2(-1)} = \tau^{2(0)}$, $\Sigma^{(-1)} = \Sigma^{(0)}$, $t_{min} \in \mathbb{N}$

$j \leftarrow 1$

**while** $t = 1, \parallel B^{(t)} - B^{(t-1)} \parallel + \parallel \tau^{2(t)} - \tau^{2(t-1)} \parallel + \parallel \Sigma^{(t)} - \Sigma^{(t-1)} \parallel > \epsilon$ for small enough $\epsilon$

**E-step** : Integrate out the unknown complete data loglikelihood by $\mathbf{r}_i|\mathbf{y}_i, B^{(t-1)}, \tau^{2(t-1)}, \Sigma^{(t-1)} \sim MVN(\hat{\mathbf{r}}_i^{(t-1)}, \hat{\Theta}_i^{(t-1)})$ as in (11).

**CM-step** : Update each parameter by maximization over the full parameter space as in (14), (15), and (16).

**Nesterov's Acceleration** :

$$
\begin{aligned}
B^{(t)} &\leftarrow B^{(t)} + \frac{j-1}{j+2}(B^{(t)} - B^{(t-1)}) \\
\tau^{2(t)} &\leftarrow \tau^{2(t)} + \frac{j-1}{j+2}(\tau^{2(t)} - \tau^{2(t-1)}) \\
\Sigma^{(t)} &\leftarrow \Sigma^{(t)} + \frac{j-1}{j+2}(\Sigma^{(t)} - \Sigma^{(t-1)})
\end{aligned}
$$

**if** $\| B^{(t)} - B^{(t-1)} \| + \| \tau^{2(t)} - \tau^{2(t-1)} \| + \| \Sigma^{(t)} - \Sigma^{(t-1)} \| < \| B^{(t-1)} - B^{(t-2)} \| + \| \tau^{2(t-1)} - \tau^{2(t-2)} \| + \| \Sigma^{(t-1)} - \Sigma^{(t-2)} \|$ & $t \geq t_{min}$

**then** $j \leftarrow 1$

**else** $j \leftarrow j + 1$

**end if**

**end while**

## 2 Supplementary Figures

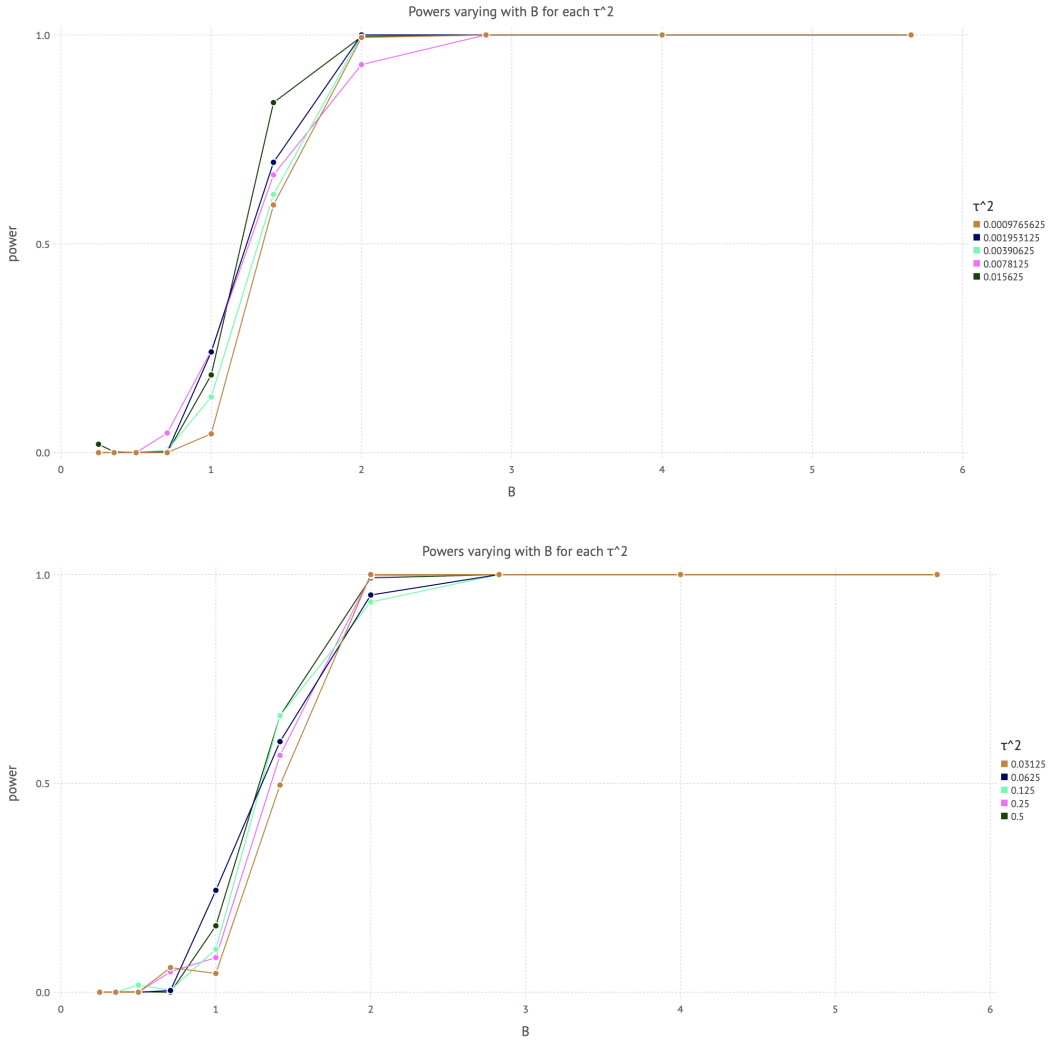



Figure 1: Power varied by $\tau^2$ on simulated data using Arabidopsis thaliana and 3 year by 2 site soil daily range temperature data. Entries of $B$ are randomly selected from $\{\pm 0.25\}$ multiplied by $\sqrt{2}$ raised to the power per $\tau^2$.

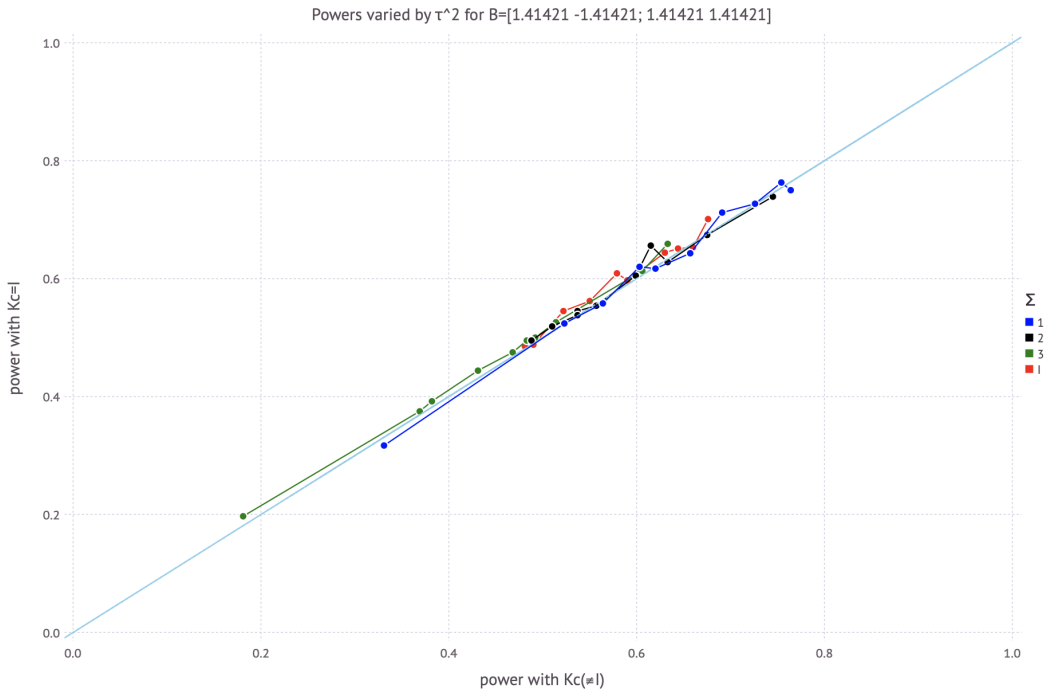



Figure 2: Power comparison of the existence of climatic relatedness ($K_C$) with its non-existence ($K_C = I$) on simulated data using Arabidopsis thaliana and 3 year by 2 site soil daily range temperature data, varying unstructured environment errors ($\Sigma$). Entries of $B$ are randomly selected from $\{\pm 0.25\}$ multiplied by $\sqrt{2}$ raised to the power per $\tau^2$.

## 3 Supplementary Tables

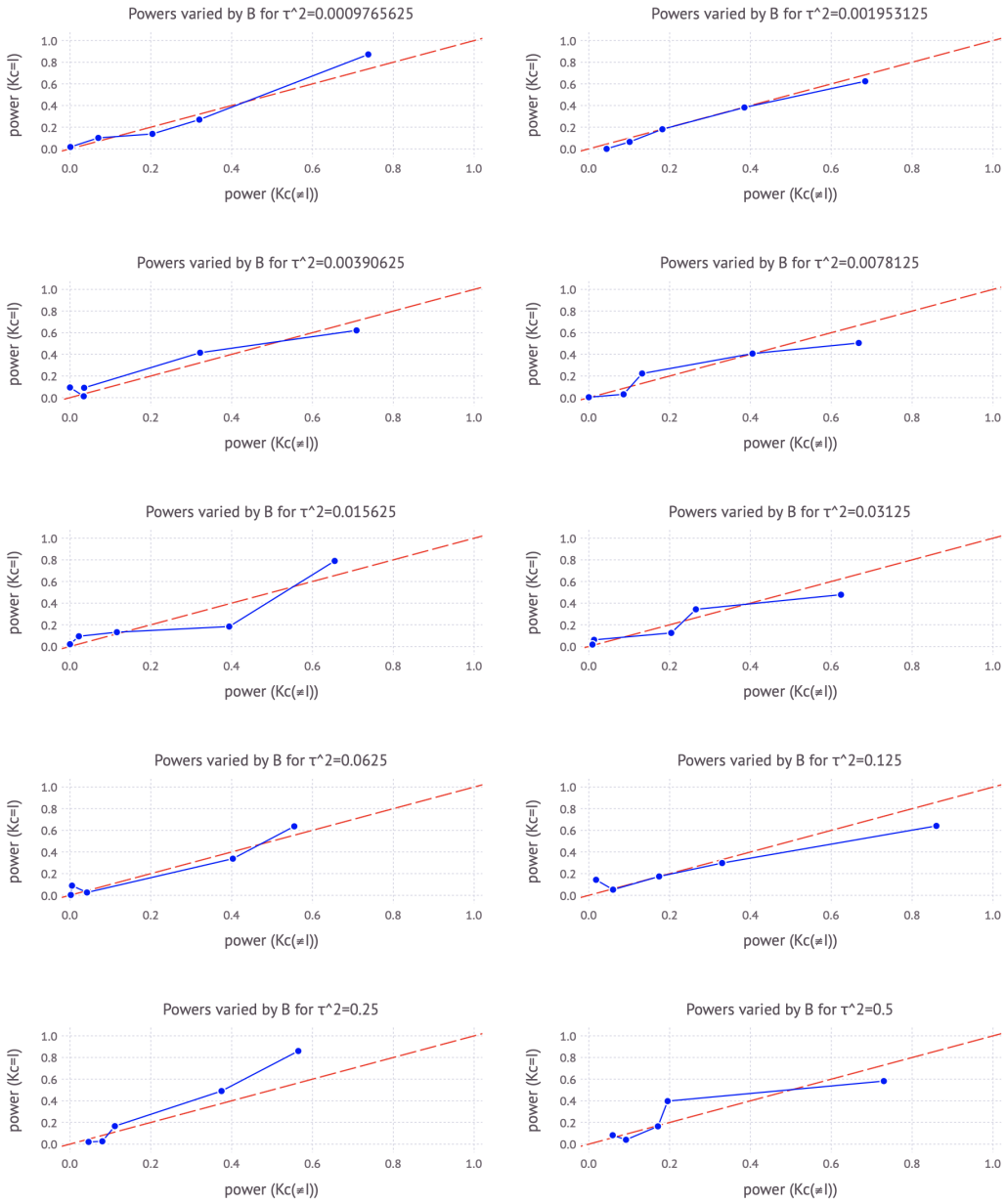



Figure 3: Contribution of climatic relatedness ($K_C$) to power on simulated data using Arabidopsis thaliana and 3 year by 2 site soil daily range temperature data. $K_C = I$ indicates not using the climatic relatedness. The range of true effect size matrices ($B$) narrows down based on power measured in Supplementary Figure 1.

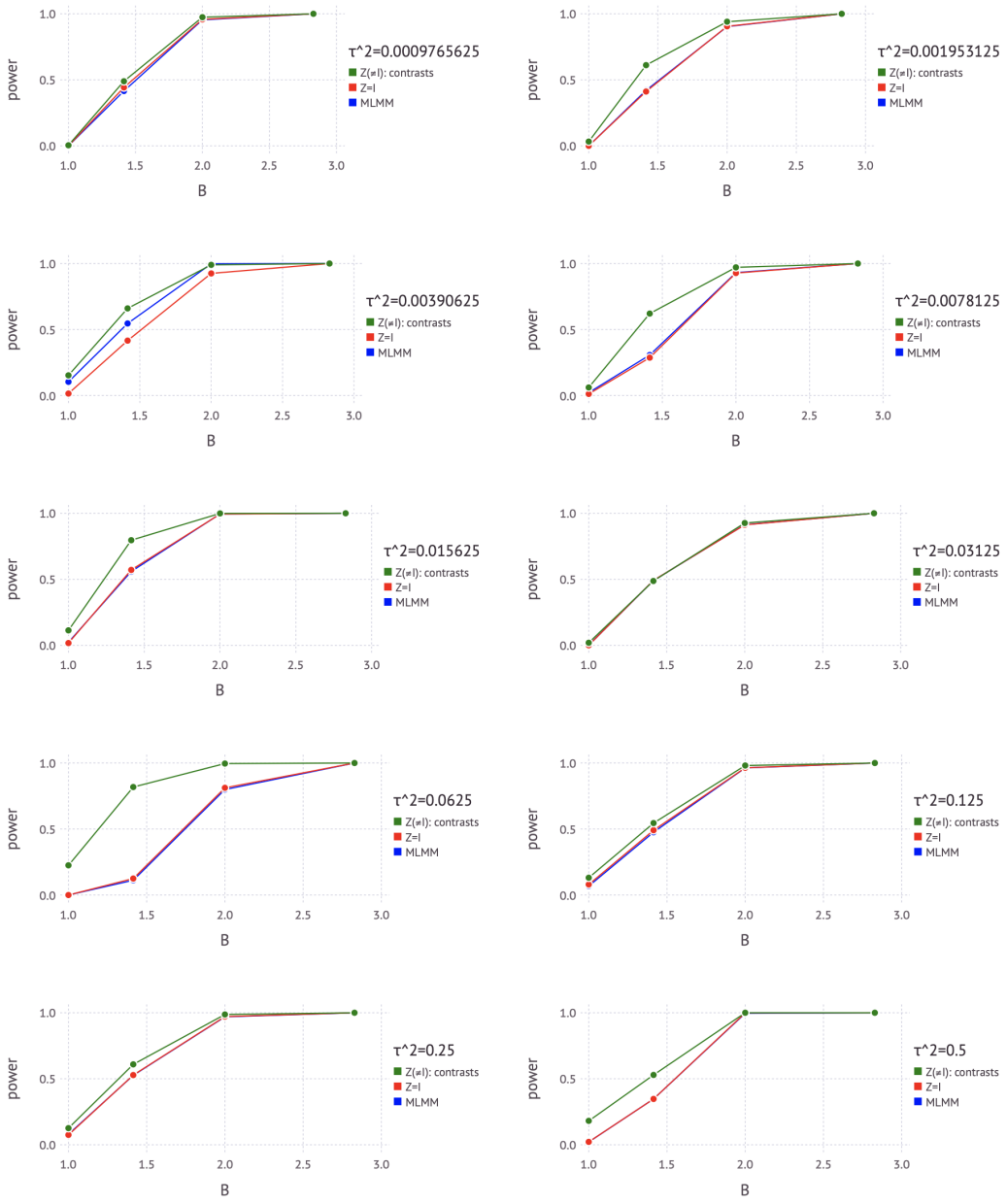



Figure 4: Contribution of low-dimensional phenotypic covariates ($Z$) to power on simulated data using Arabidopsis thaliana and 3 year by 2 site soil daily range temperature data. $Z$: contrasts, $Z = I$ indicate employing the covariates ($Z$) or not in flxQTL, and 'MLMM' indicates a Multivariate LMM implemented in Julia (Julia-version Multivariate LMM) as an alternative for GEMMA.

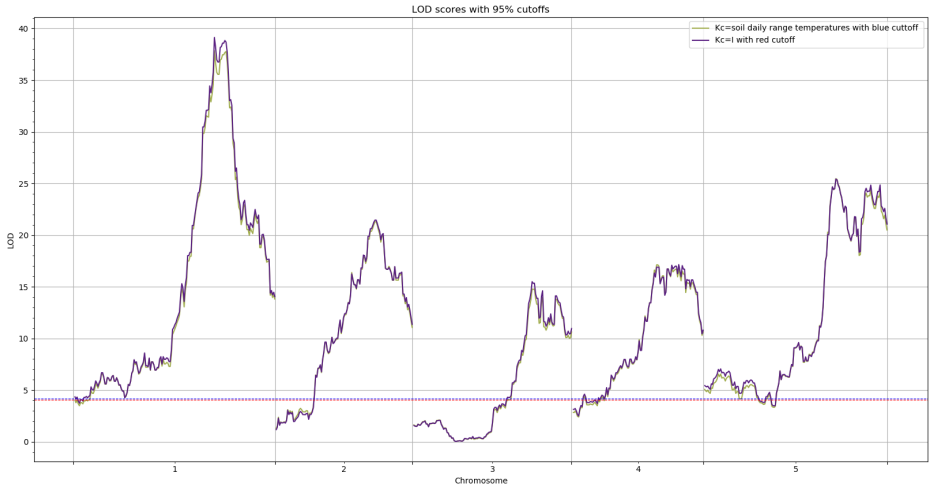



Figure 5: Multivariate genome scan for Arabidopsis thaliana: Comparison of LOD scores including a climatic relatedness ($K_C$) or not and corresponding 95% thresholds computed by 1000 permutations. Phenotypes of fitness (mean number of fruits per seedling planted) are considered as 6 environments, i.e. a 3 year by 2 site combination. The climatic relatedness matrix $K_C$ is computed by using soil daily range temeperatures in two sites (Sweden and Italy) for 3 years. The low-dimensional phenotypic covariates ($Z$) are contrasts containing 1's for overall mean for 6 environments and -1's (Italy), 1's (Sweden) for mean difference between sites to measure gene by environment interaction.

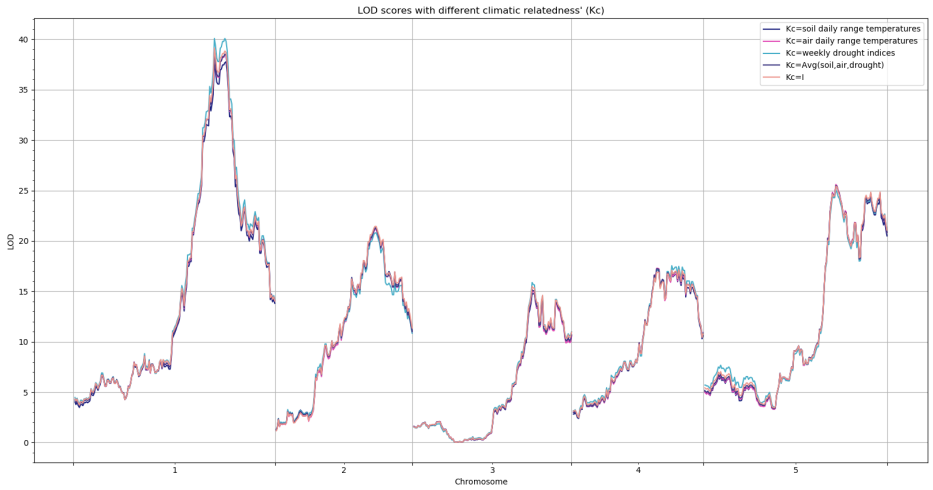



Figure 6: Multivariate genome scan for Arabidopsis thaliana: Comparison of LOD scores including various climatic relatedness matrices ($K_C$). Phenotypes of fitness (mean number of fruits per seedling planted) are considered as 6 environments, i.e. a 3 year by 2 site combination. The climatic relatedness matrices $K_C$ are respectively computed by using soil/air daily range temperatures, weekly drought indices in two sites (Sweden and Italy) for 3 years. The low-dimensional phenotypic covariates ($Z$) are contrasts containing 1's for overall mean for 6 environments and -1's (Italy), 1's (Sweden) for mean difference between sites to measure gene by environment interaction.

With Supplementary Figure 5, both point out that results are insensitive to types of climatic relatedness matrices ($K_C$'s).

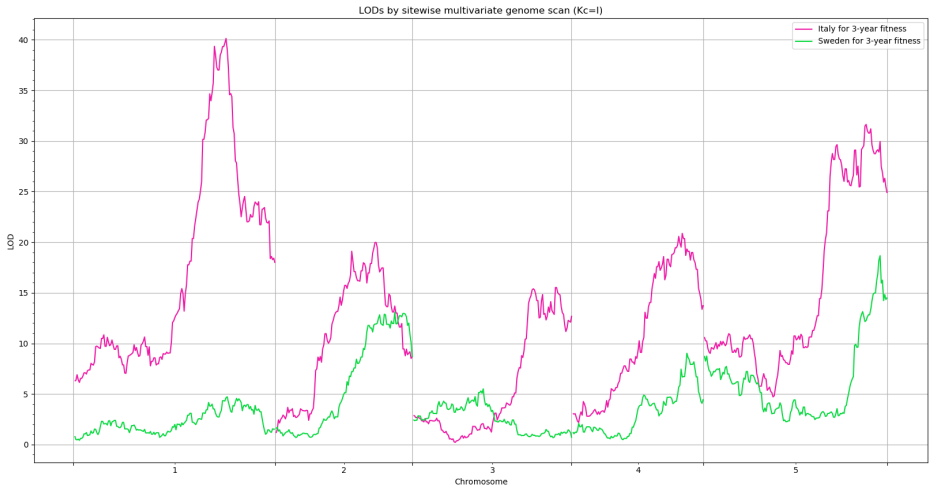



Figure 7: Site-wise multivariate genome scan for Arabidopsis thaliana: Comparison of LOD scores between Italy and Sweden, where $K_C$ is not included according to Supplementary Figures 5, 6.

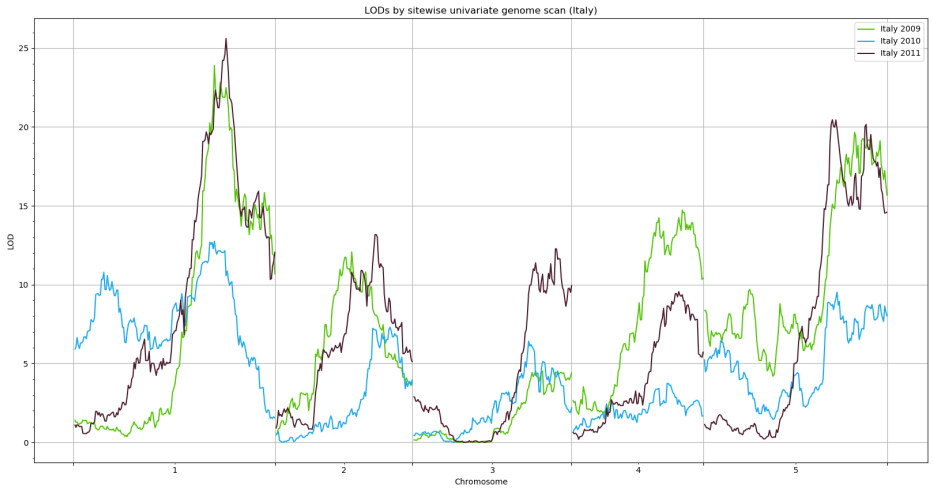



Figure 8: Univariate genome scans for Arabidopsis thaliana in Italy from 2009 to 2011.

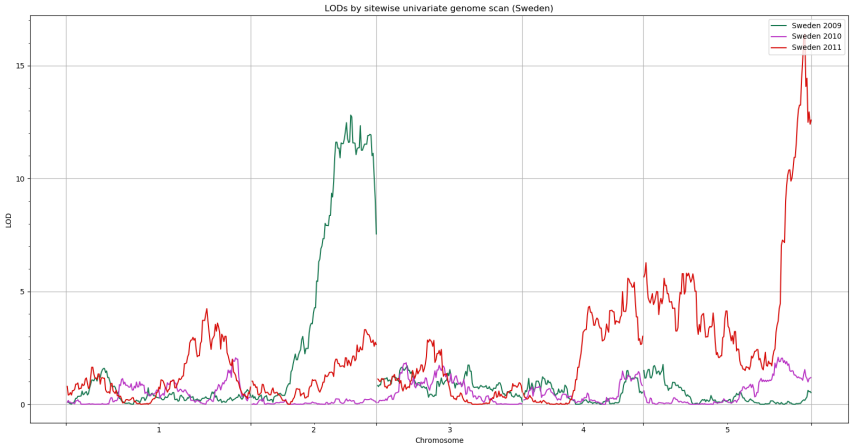



Figure 9: Univariate genome scans for Arabidopsis thaliana in Sweden from 2009 to 2011.

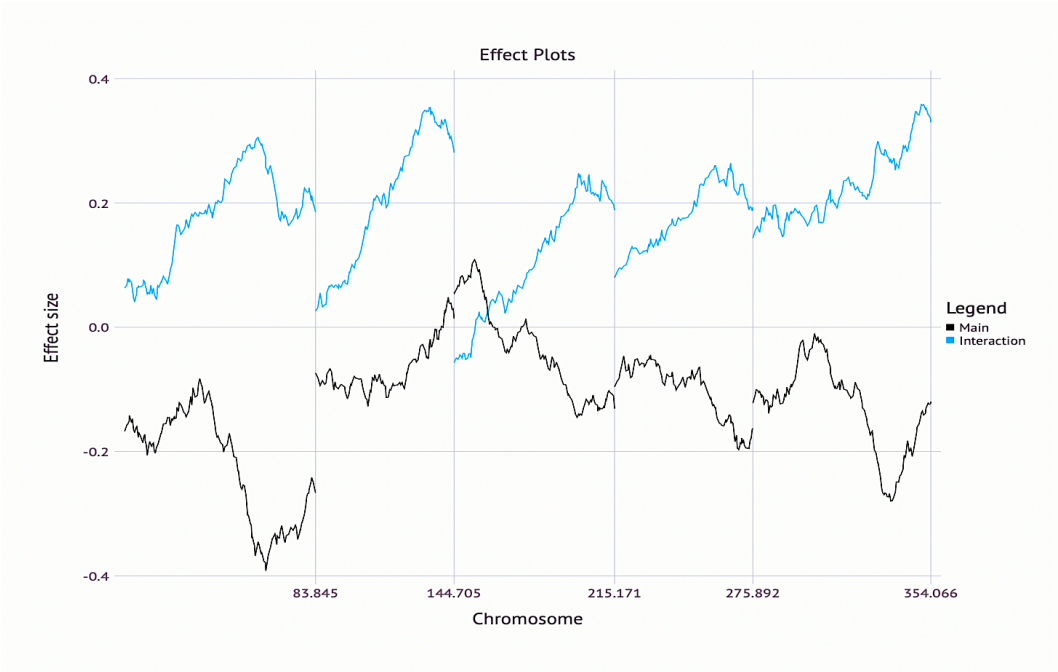



Figure 10: Effect plots for Arabidopsis thaliana with $K_C = I$ for 6 environments as described in Supplementary Figure 5.

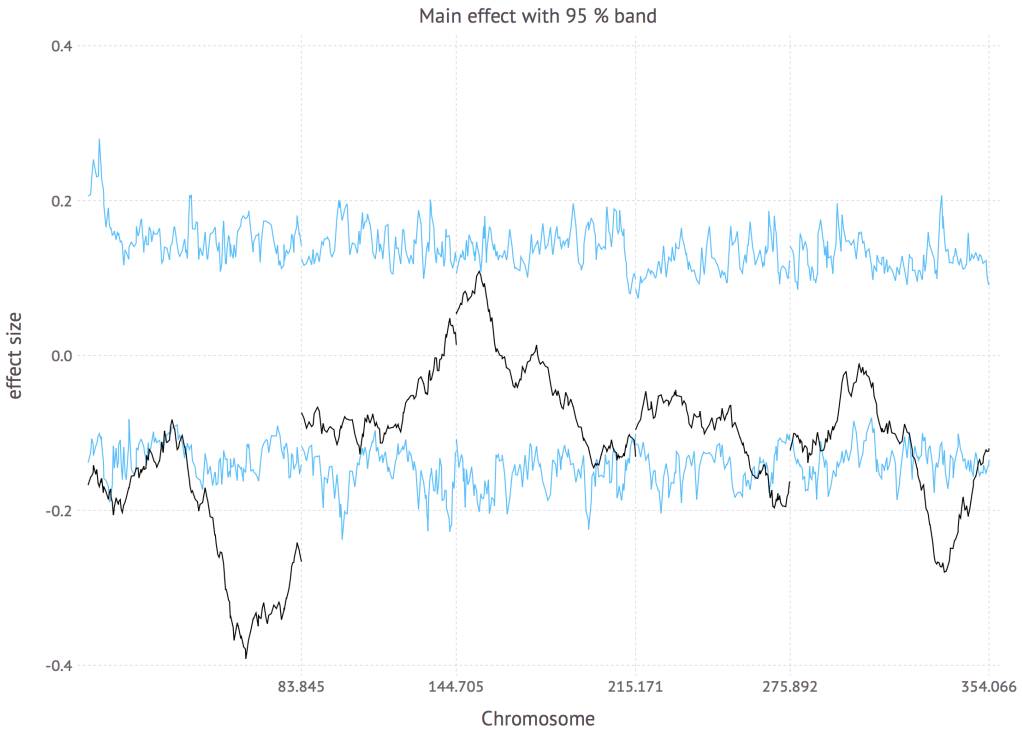



Figure 11: Main effects with 95% band for Arabidopsis thaliana. The band is obtained by 100 permutations

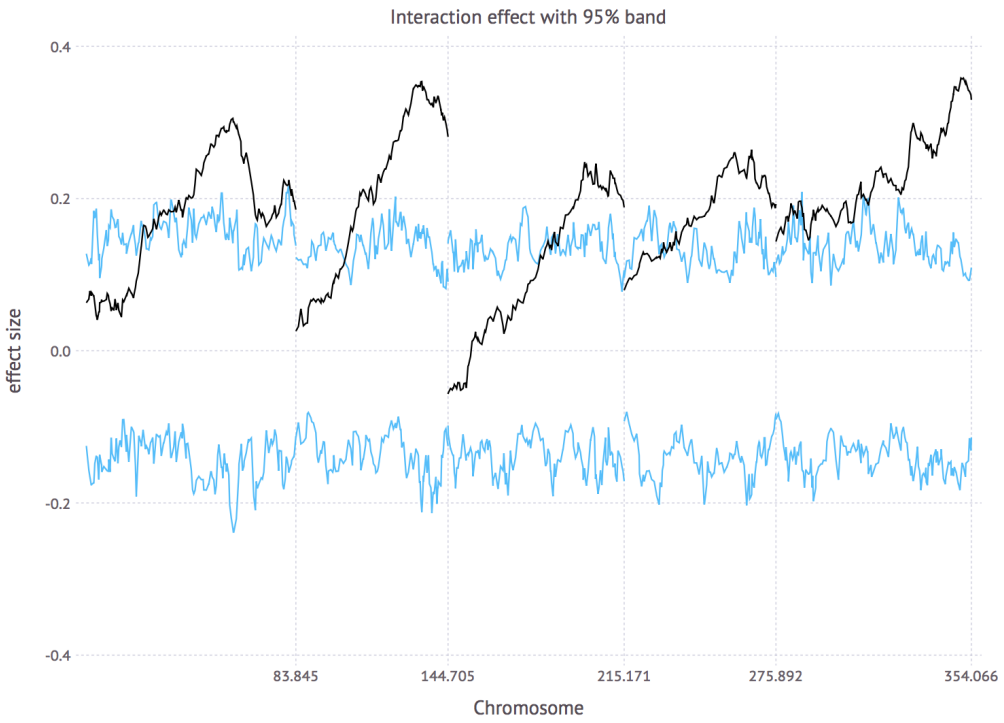



Figure 12: Interaction effects with 95% band for Arabidopsis thaliana. The band is obtained by 100 permutations

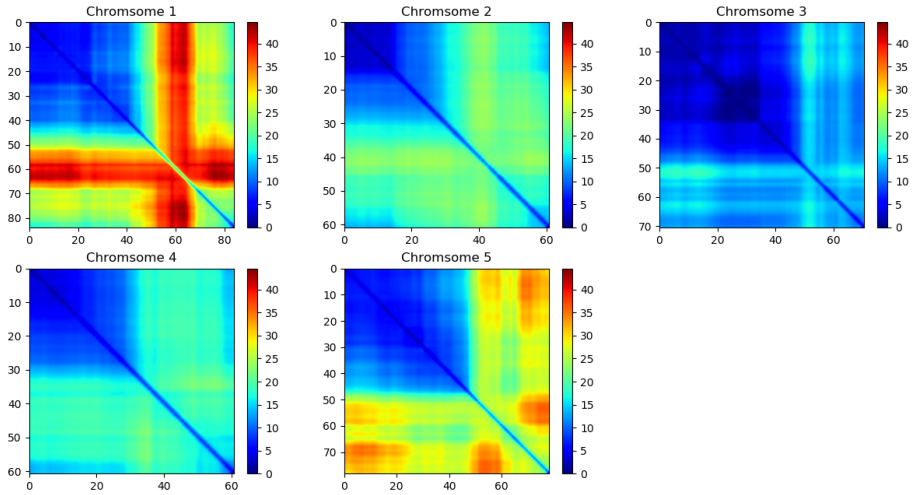
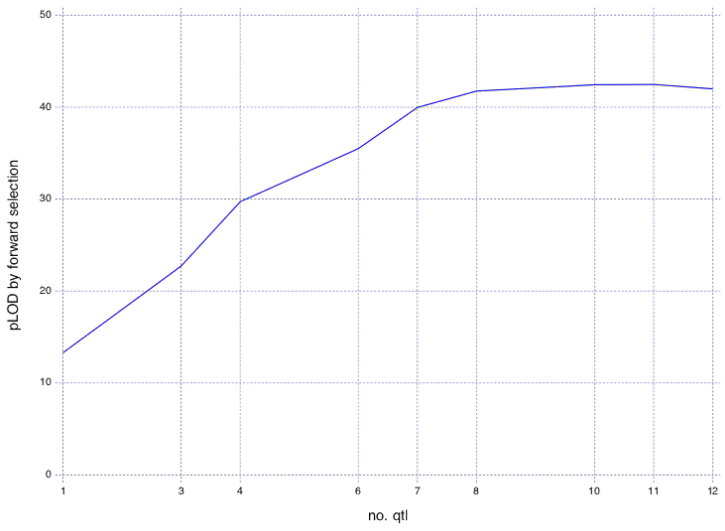
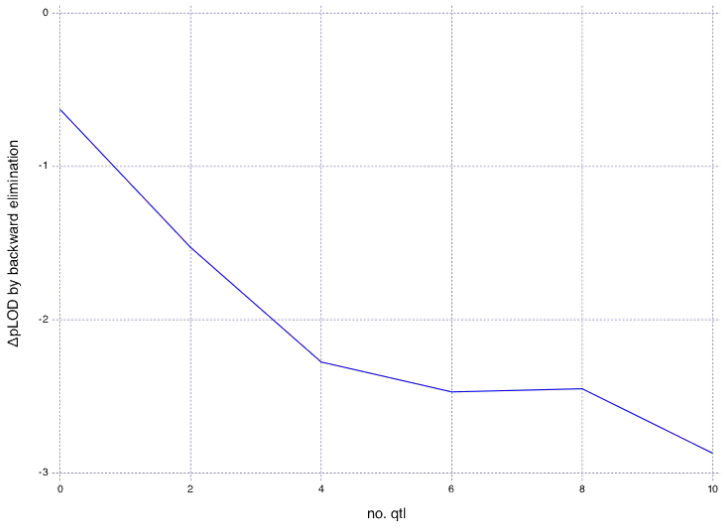



Figure 13: 2-d multivariate genome scan for Arabidopsis thaliana.




(a) Foward selection




(b) Backward elimination

Figure 14: QTL selection for Arabidopsis thaliana (multiple QTL analysis).

A model selection approach by forward selection and backward elimination is performed to finialize significant QTL. A selection criterion is the penalized LOD score for additive models equivalent to BIC [29]. 1-d and 2-d genome scans are performed each step in both schemes based on QTL selected from 1-d and 2-d multivariate genome scans in Supplementary Figure 5, 13.

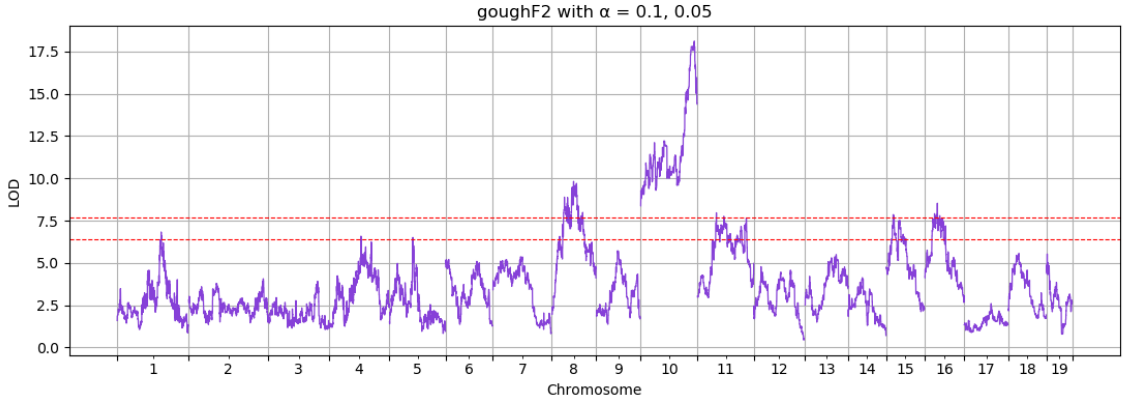



Figure 15: Application to functional data (F2 intercrosses between Gough Island mice and WSB/EiJ): LOD scores and thresholds at $\alpha = 0.1, 0.05$

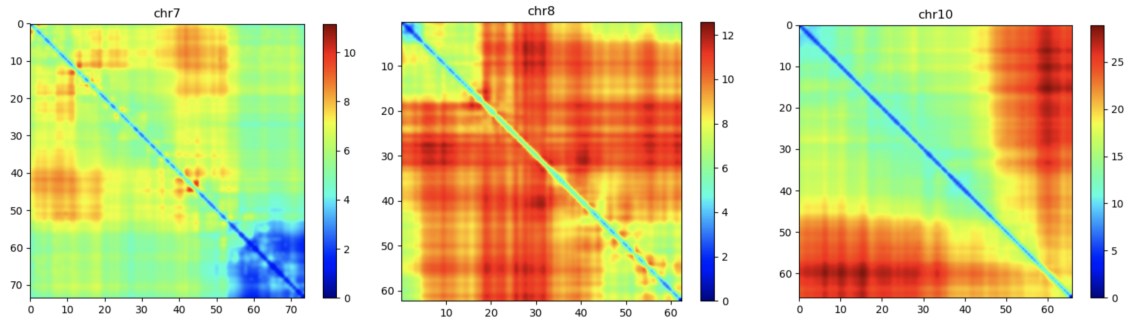



Figure 16: Application to functional data (F2 intercrosses between Gough Island mice and WSB/EiJ):LOD scores by 2D genome scan

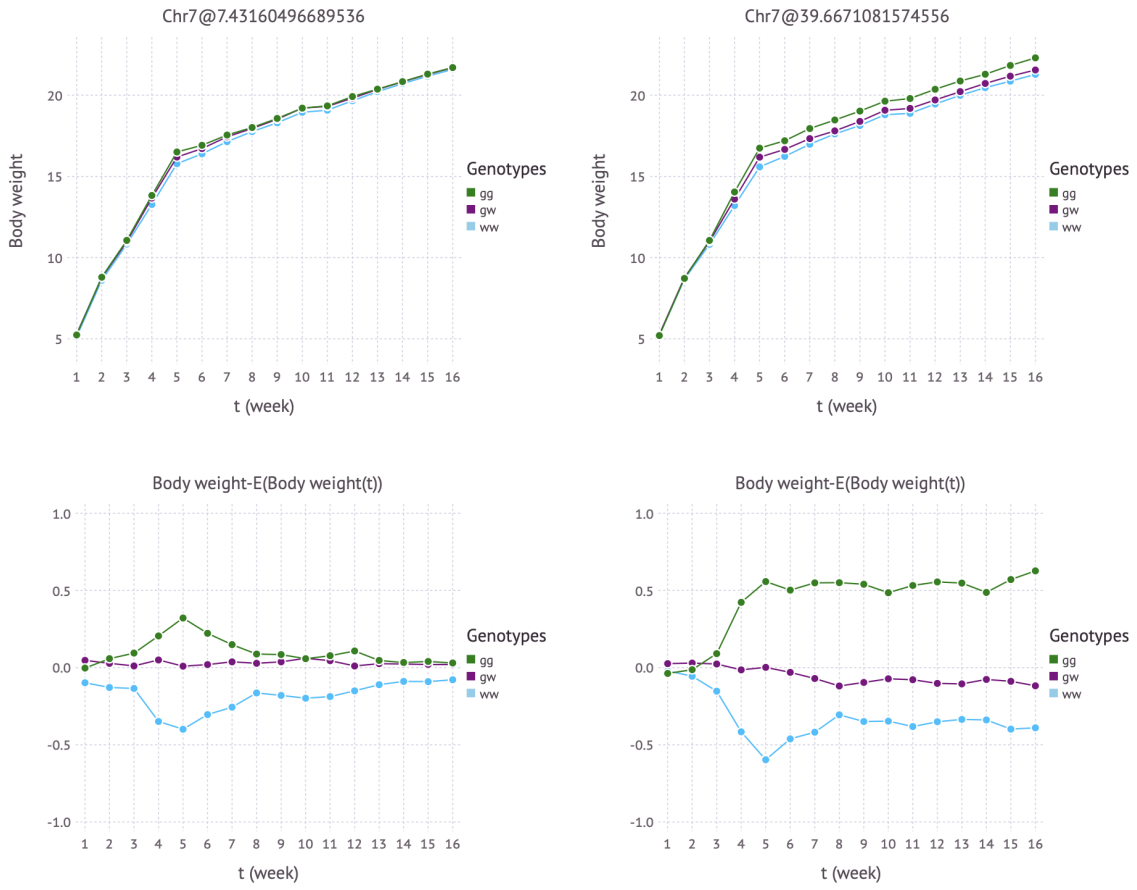



Figure 17: Application to functional data (F2 intercrosses between Gough Island mice and WSB/EiJ): Body weight effects by genotypes and corresponding deviations from weekly overall mean in Chromosome 7.

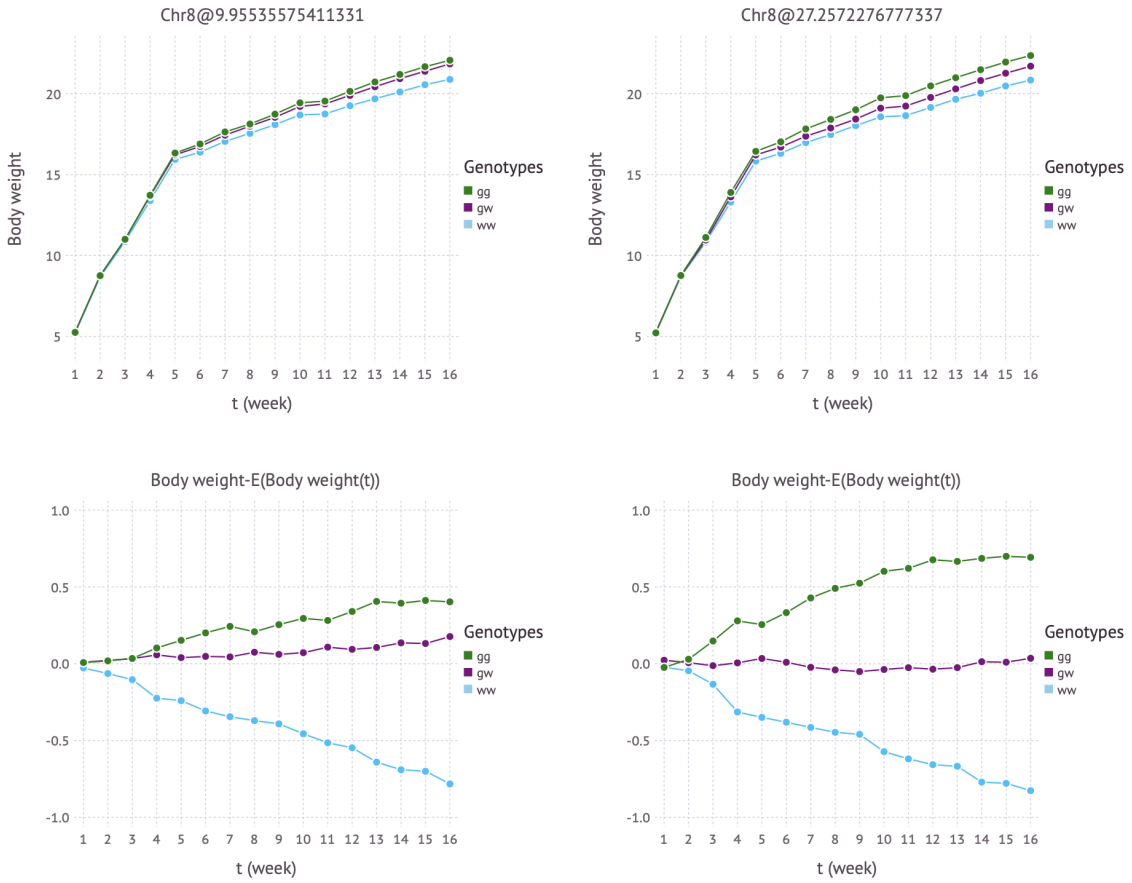



Figure 18: Application to functional data (F2 intercrosses between Gough Island mice and WSB/EiJ): Body weight effects by genotypes and corresponding deviations from weekly overall mean in Chromosome 8.

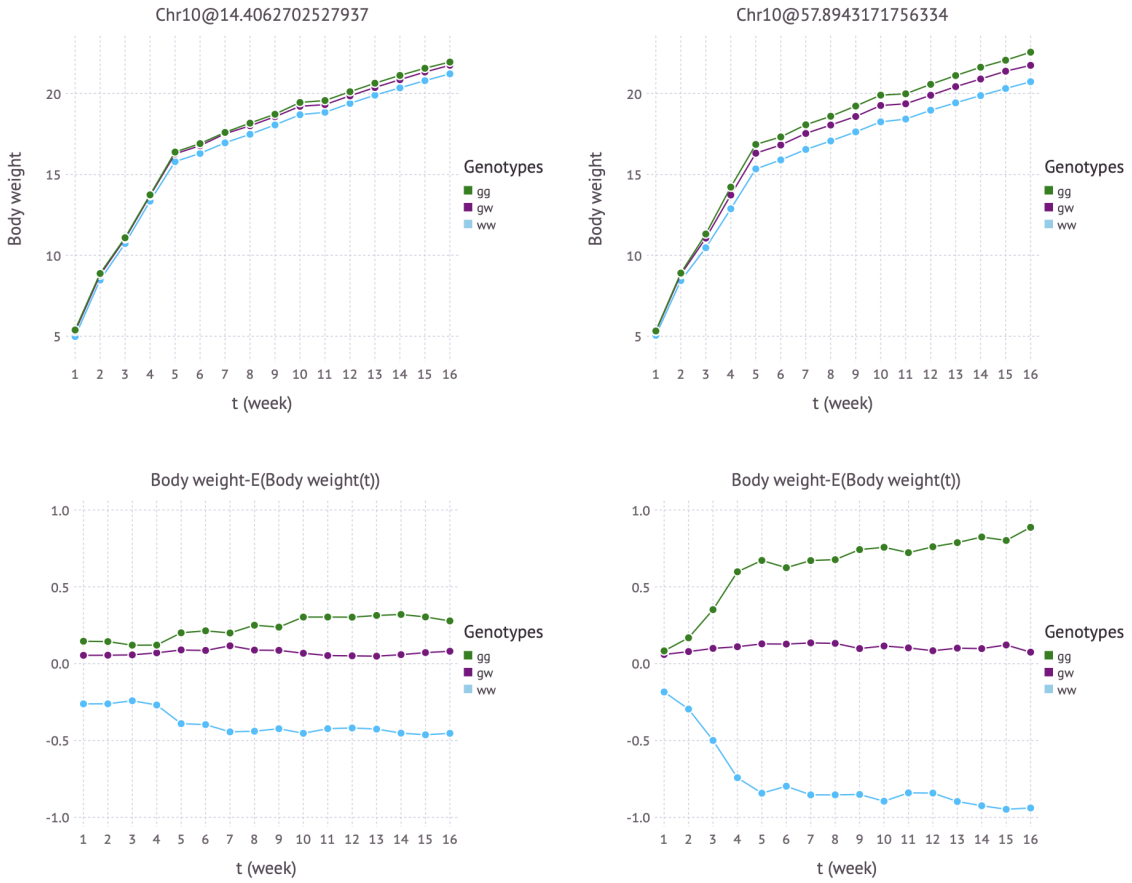



Figure 19: Application to functional data (F2 intercrosses between Gough Island mice and WSB/EiJ): Body weight effects by genotypes and corresponding deviations from weekly overall mean in Chromosome 10.

| $\tau^2$ | $\Sigma_1$ $(K_C \neq I)$ | $\Sigma_1$ $(K_C = I)$ | $\Sigma_2$ $(K_C \neq I)$ | $\Sigma_2$ $(K_C = I)$ | $\Sigma_3$ $(K_C \neq I)$ | $\Sigma_3$ $(K_C = I)$ | $\Sigma = I$ $(K_C \neq I)$ | $\Sigma, K_C = I$ |
|---|---|---|---|---|---|---|---|---|
| 0.00097656 | 0.726 | 0.727 | 0.675 | 0.674 | 0.369 | 0.375 | 0.481 | 0.485 |
| 0.00195313 | 0.62 | 0.617 | 0.537 | 0.538 | 0.492 | 0.5 | 0.676 | 0.701 |
| 0.00390625 | 0.691 | 0.712 | 0.633 | 0.628 | 0.431 | 0.444 | 0.59 | 0.597 |
| 0.0078125 | 0.564 | 0.558 | 0.488 | 0.495 | 0.181 | 0.197 | 0.522 | 0.545 |
| 0.015625 | 0.523 | 0.524 | 0.745 | 0.739 | 0.633 | 0.659 | 0.66 | 0.653 |
| 0.03125 | 0.603 | 0.62 | 0.51 | 0.519 | 0.468 | 0.475 | 0.644 | 0.651 |
| 0.0625 | 0.657 | 0.643 | 0.557 | 0.554 | 0.382 | 0.392 | 0.63 | 0.644 |
| 0.125 | 0.764 | 0.75 | 0.599 | 0.605 | 0.514 | 0.526 | 0.55 | 0.562 |
| 0.25 | 0.754 | 0.763 | 0.615 | 0.656 | 0.606 | 0.613 | 0.49 | 0.488 |
| 0.5 | 0.331 | 0.317 | 0.537 | 0.545 | 0.483 | 0.495 | 0.579 | 0.609 |

Table 1: Power comparison of the existence of climatic relatedness ($K_C$) with its non-existence ($K_C = I$) in Supplementary Figure 2.

$6 \times 6$ matrices of $\Sigma_i$ for $i = 1, 2, 3$ are structured using following $3 \times 3$ sub-matrices: $A$, $C$ (1's on the diagonal, $a$, $c$ on off-diagonal entries, respectively), $B$ ($b$ in all entries). $\Sigma_1$ = [A B;B A], where $a = 0.4$, $b = 0.07$, $\Sigma_2$ =[A B;B A], where $a = 0.3$, $b = 0.07$, $\Sigma_3$=[A B;B C], where $a = 0.3$, $b = 0.0$, $c = 0.1$. $I$ is an identity matrix.

| Algorithm | runtime for SNPs in Chr 1 | runtime per SNP |
|---|---|---|
| flxQTL (Julia) | 13.16 s | 0.0136 s |
| GEMMA (C++) | 12.71 s | 0.0134 s |

Table 2: Comparison of performance time between flxQTL and GEMMA using Mouse_HS1940 data (n=1940,m=3, p=967 for flxQTL, and p=950 for GEMMA due to different bad marker dropping rules.). Version information: OS: Linux Debian 4.19.37-5, CPU: Intel(R) Xeon(R) CPU E5-2630 2.4GHz, Software: flxQTL (Julia v.1.5.0, adding 8 cores), GEMMA (v.0.98, max 8 threads).

| Adding 1 or 2 QTL | $\Delta$ pLOD | pLOD | Deleting 1 or 2 QTL | $\Delta$ pLOD |
|---|---|---|---|---|
| 10 | 0.692 | 42.442 | 10 | -2.872 |
| 11 | 0.028 | 42.471 | 8 | -2.451 |
| 12 | -0.480 | 41.991 | 6 | -2.471 |

Table 3: QTL selection for Arabidopsis thaliana (multiple QTL analysis) as in Supplementary Figure 14. 11 QTL can be finalized as significant QTL since the highest change ($\Delta$ pLOD = -2.872) in the backward elimination step when dropping one QTL from 11 QTL selected from the forward selection step yields.

| Chromosome | no. of QTL | position (cM) |
|---|---|---|
| 1 | 2 | 57, 77 |
| 2 | 2 | 133, 140 |
| 3 | 2 | 152, 199 |
| 4 | 2 | 259, 272 |
| 5 | 3 | 280,333,353 |

Table 4: Significant QTL for Arabidopsis thaliana as described in Supplementary Figure 14 and Supplementary Table 3.