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 containing rich information across traits. We present a flexible multivariate linear mixed model for quantitative trait loci mapping (FlxQTL) for multiple correlated traits that adjusts for genetic relatedness and that models information on multiple environments or multiple timepoints using trait covariates. FlxQTL handles genetic mapping of multivariate traits faster with greater flexibility compared to previous implementations.

Multivariate traits are increasingly common in genetic studies. A trait may be measured in multiple environments, at multiple timepoints (e.g. repeated measures), or under different treatments. Such a trait can be considered as a multivariate trait with spatial, temporal [1,2] or more com-

plex structures. For instance, body weight might be measured weekly, plant crop yield or biomass could be measured in multiple geographic locations for years, (in multi-environment trials, MET), and gene expression may be measured in different brain regions of the same rat. In humans it is common to study genotype-phenotype associations in population-based samples for genome-wide association studies (GWAS); in model organisms complex breeding designs are often used. Linear mixed models (LMMs) [3–6] are employed for controlling confounding due to genetic relatedness among individuals and are used to identify genetic loci contributing to quantitative traits of interest (QTL) [7–13]. Multivariate LMMs (MLMMs) further enhance statistical power over univariate LMMs [14] because they can accumulate signals across traits [8,15] with a common genetic locus.

While recognized as advantageous, fitting MLMMs is generally avoided because it is computationally challenging. Parameter estimation involves multidimensional optimization, there is a risk of reaching suboptimal solutions, and it is computationally expensive [9,14]. Existing algorithms (GCTA [16,17], ASReml [18], WOMBAT [19], GEMMA [14]) use the expectation-maximization (EM) method followed by second order schemes such as Newton-Rapson (NR), Average Information (AI), etc. to provide stable, fast convergence [5,14]. The computational complexity is $O(n^3m^3)$ for EM and $O(n^3m^7)$ for NR, AI (m traits, n individuals). This suggests that using GCTA, WOMBAT, and ASReml is not practical for GWAS with a large number of SNPs and a moderate number of individuals [14]. GCTA can fit only up to 2 traits; ASReml and WOMBAT offer some flexibility in choosing some covariance structures and including modeling fixed effects of covariates. All have limitations in the number of traits they can handle and are much slower than GEMMA. None of these methods offer the ability to model spatial or temporal structure in the trait using trait

covariates.

Here, we introduce FlxQTL, which can test associations between genetic markers and multiple correlated traits (**Supplementary Software** and https://github.com/senresearch/FlxQTL.jl). Our method is a MLMM that models the mean using a bilinear model of individual and trait covariates (**Fig. 1**). The error term is the sum of a genetic random effect and a pure error component. The pure error term is correlated across traits but is independent across individuals. The genetic variance component is assumed to be proportional to the Kronecker product of a genetic kinship matrix and a trait kernel. In the case of MET, that can be interpreted as a random effect term comprised of the sum of many small random GxE effects (**Supplementary Note**). We assume

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

where $Y_{n\times m}$ is the response matrix for a trait measured in n individuals across m environments (or over m time points) or phenotypes of n individuals for m traits, $X_{n\times p}$ is a matrix of p genotype probabilities (or genotypes including the intercept) of a marker to be tested, and $Z_{m\times q}$ is a matrix of q trait covariates such as environment contrasts in MET, spline basis functions for modeling smooth temporal traits. B is a coeffcient matrix to be estimated, R is the matrix of genetic random effects, and E is the residual error matrix. The likelihood ratio test (LRT) to test the association between a marker and a multivariate trait of interest is expressed as a LOD score ($-\log_{10}(LRT)$). Genomewide thresholds for significance may be obtained from the null distribution of maximum LOD scores using a permutation scheme (Online Methods).

FlxQTL simplifies parameter estimation using its model structure (**Fig. 1**) and the choice of

numerical methods. The trait covariates reduce the fixed effects matrix dimension from $p \times m$ for existing MLMMs [8,14], to $p \times q$, which is substantial when m/q is large. The trait kernel reduces a $m \times m$ random effects covariance matrix to a scalar parameter. With an unstructured pure error matrix, only $pq+1+\frac{m(m+1)}{2}$ parameters are estimated, instead of $pm+m(m+1)$ ($q << m$) for other MLMMs. FlxQTL has been tested with up to a 36-dimensional quantitative trait. Unlike NR and AI, a Speed restarting Nesterov's accelerated gradient method requires no second derivative and provides stable linear convergence [20]. We use an expectation-conditional maximization (ECM) [21] step, followed by the acceleration step using a tuned momentum coefficient, which behaves like the second derivative, to update parameters (**Supplementary Note**). These simplifications offer improved complexity compared to existing algorithms (Online Methods).

FlxQTL offers flexibility in handling varied trait structures and can analyze complex crosses if genotype probability data is available. The user can choose contrasts between trait covariates to influence how structured traits are analyzed. For example, GxE interactions can be analyzed using a contrast between sites in the Z matrix. One can use B-spline, wavelet, or Fourier basis functions in the Z matrix to model an environmental gradient or a time trend. By judiciously choosing the Z (trait covariate design) matrix, the user can model high-dimensional structured traits in MET, multiple related traits, and time-valued traits for a variety of model organisms as well as humans. Unlike some programs designed only for human genetics (maximum of three genotype states possible at a locus), if genotype probabilities are available, they can be used for QTL analysis in FlxQTL. This opens up the possibility of using 4-way crosses and multi-parental crosses such as heterogeneous stocks and the Collaborative Cross.

We compared the performance of FlxQTL with that of GEMMA on the Mouse HS1940 data distributed with GEMMA. The trait kernel and trait covariate matrices were set to be identity matrices in FlxQTL since they are not supported by GEMMA. We measured computation times for analyzing 3 traits and then increased the number of traits and the number of SNPs analyzed for both algorithms (Data processing and analysis in Online Methods). The difference in P values between the algorithms was negligible (**Fig. 2b** and **2c**). GEMMA was faster for a smaller number of traits (3 traits), whereas FlxQTL was 1.5-21 times faster than GEMMA when more traits were analyzed (6-12 traits) (**Fig. 2a**). These results show that FlxQTL offers fast implementation for high-dimensional traits.

We evaluated the statistical performance of FlxQTL simulating 6-dimensional quantitative trait datasets using genotypes from *Arabidopsis thaliana* [22] studied in two sites for three years (Online Methods). We carried out two simulation studies on the contributions of a climatic relatedness matrix as a trait kernel and trait covariates, respectively. A comparison of power with, and without the climatic relatedness was made when including the same trait contrasts (**Supplementary Figs. 3-5**). We next compared power with and without trait covariates, allowing main effect QTL and QTL$\times$site interactions, keeping those climatic relatedness matrices the same (**Supplementary Fig. 6**). The thresholds for the type I error rate were calibrated by the distribution of maximum LOD scores under the null model of only multivariate gene and environment random effects. The power was measured under the alternative model of existing large effects of QTL based on the thresholds. The overall result of assessing the trait kernel showed that including climatic relatedness as a trait kernel yielded only subtle differences in power. The result from the added value of

trait covariates varying the strength of the genetic variance component demonstrated that in each case, the power was the highest when including trait covariates (**Supplementary Fig. 6**).

We applied FlxQTL on the data of recombinant inbred lines from *A. thaliana* by crossing populations, where the trait was the mean number of fruits per seedling planted in two distinct regions, Sweden and Italy, from July, 2009 to June, 2012. Due to the lack of formal statistical frameworks for QTL analysis in such reciprocal transplant experiments, Ågren et al. [22] performed six separate univariate QTL scans and dissected the effects post hoc. FlxQTL can directly test for the presence of genetic drivers of adaptation in replicated reciprocal transplants (the data processing and analysis is detailed in Online Methods, **Supplementary Tables 1** and **2** and **Supplementary Figs. 11-15**).

FlxQTL was applied to a more complex dataset, a four-way outbred mapping population in outcrossing switchgrass (*Panicum virgatum*). The trait was flowering time measured across 10 sites spanning a latitudinal gradient for 4 years (2016-2019). We considered it as a 36-dimensional multivariate trait since 4 site-year combination had missing data. These traits are highly correlated due to shared genetic and environmental contributions. We scanned for loci showing multivariate effects and picked the top three QTL. For those QTL, we performed an environment scan, for monthly averaged photoperiod in different sites and years. We found that the photoperiod influences one locus (QTL@6.6012 on Chromosome 5N) much more compared to the two other QTL (**Supplementary Fig. 16** and Data processing and analysis in Online Methods). This suggests one locus is more responsive to photoperiod and has implications for breeding.

As an example of a longitudinal trait, we analyzed body weight measured weekly from birth to 16 weeks in 1212 F_2 intercrosses between Gough Island mice and WSB/EiJ [2]. We used cubic splines with 4 degrees of freedom as the trait covariates. FlxQTL captured the weight trend and corresponding weight growth rate simultaneously and yielded the result similar to R/qtl [23]. For two chromosomes, we performed two-dimensional genome scan for further investigation and found several more QTL. (Online Methods and **Supplementary Figs. 17- 21**).

Our results demonstrate that the model and estimation algorithm are effective in analyzing higher dimensional structured traits with fast parameter estimation. On the simulation study, FlxQTL increases power making use of information from trait covariates. Our method appears to be insensitive to a choice of trait kernel matrices in our studies so far, and this may be worth further investigation.

The underlying trait model assumes a joint multivariate normal distribution and is not ideal for traits such as count data with lots of zeros, which do not fit this scenario. Some traits may need transformation to better fit the multivariate normal assumption. Our method assumes that the multivariate trait has no missing data and that genotypes are either completely observed or imputed, or genotype probabilities are used. If phenotype data contain missing values, one can use imputation [24]. With the growth of more complex and high-throughput phenotyping, we expect the use of our models to help shed light on the genetic control of structured multivariate traits.

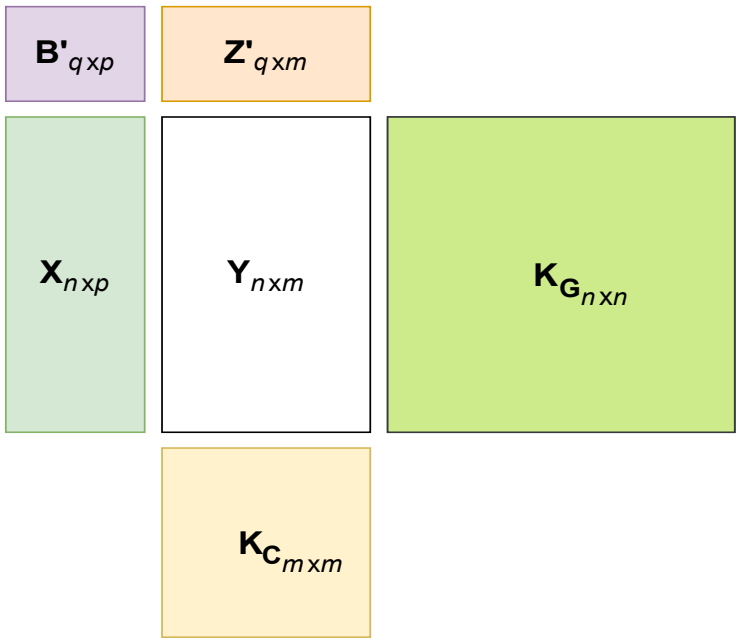


Figure 1: FlxQTL model components and their dimensions. The trait matrix ($Y_{n\times m}$) has information for n individuals (row) and m traits (column). In each dimension (by row and column), we have two paired components corresponding to fixed and random term respectively. The two components by row are a matrix of p genotypes of a marker ($X_{n\times p}$) to be tested and a kinship matrix (K_G). The two components by column are a matrix of q trait covariates $Z_{m\times q}$, and a trait kernel K_C. $B_{p\times q}$ is a matrix of fixed effects to be estimated. The two kernel matrices, K_G and K_C contribute to the variance component, and the two design matrices, X and Z, contribute to the mean components.

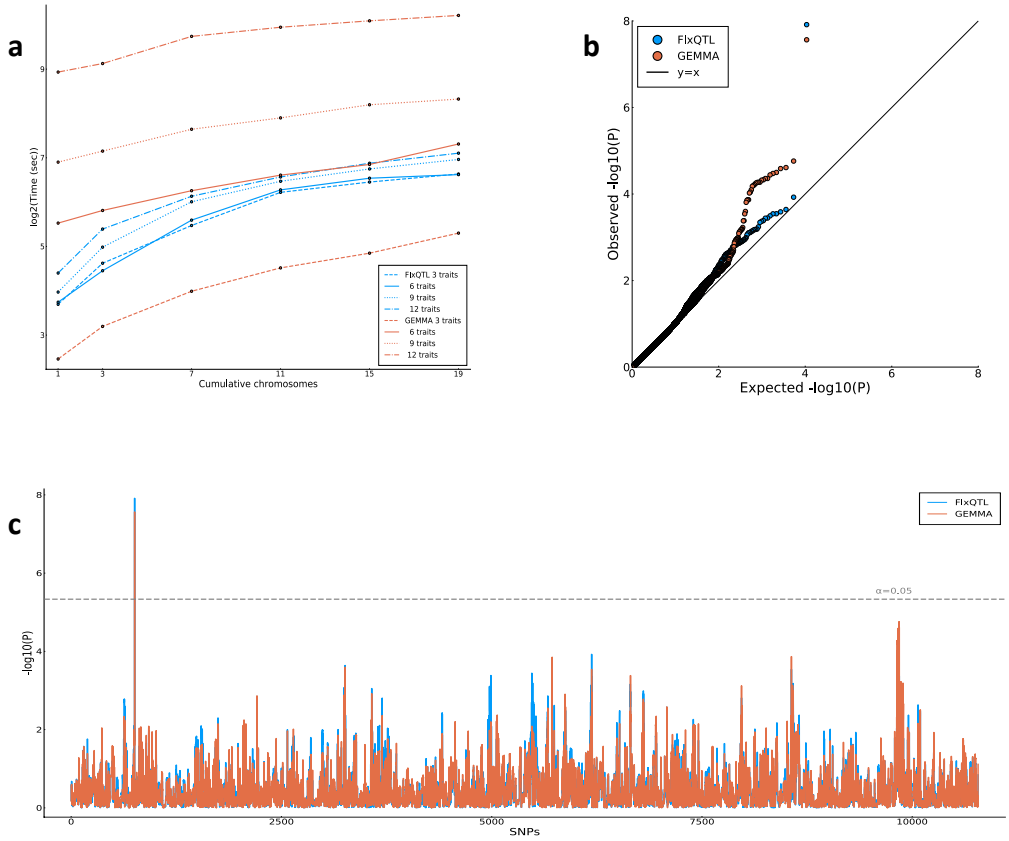


Figure 2: Comparison of FlxQTL and GEMMA using Mouse HS1940 data (n=1940 individuals, p=10783 SNPs) and simulated data. (**a**) Computation times in seconds on logarithmic scale (base 2) for parameter estimation as a function of the number of traits and SNPs. (**b**,**c**) Comparison of P values from the two algorithms using a quantile-quantile plot (**b**) and a genome scan plot (**c**) for 3 traits. The horizontal line is a Bonferroni adjusted threshold at the significance level of $\alpha = 0.05$. Computing platforms: Linux Debian 4.19.37-5 (OS), Intel(R) Xeon(R) CPU E5-2630 2.4GHz (CPU), FlxQTL (Julia v.1.5.3 with 32 cores), GEMMA (v.0.98.3 with maximum 32 threads)

METHODS

Methods and any associated references are available in the Online Methods.

ONLINE METHODS

Code availability. A Julia software implementation of FlxQTL is available at https://github.com/senresearch/FlxQTL.jl.

FlxQTL model. Let $Y_{n\times m}$ be a trait measured in n individuals across m environments (or over m time points), or phenotypes of n individuals for m correlated traits. $X_{n\times p}$ is a matrix of p genotypes including the intercept (or genotype probabilities) of a marker to be tested and can also optionally include individual level covariates such as sex for animals or cytoplasm origin for plants. $Z_{m\times q}$ is a matrix of q trait covariates such as (site) contrasts, (orthonormal) basis functions, etc. Our model for the trait is:

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

The two independent random effects follow matrix variate normal distributions:

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

One can easily see that compared to standard multivariate regression, where $Z = I_m$, $q = m$, our model reduces the size of B from $p\times m$ to $p\times q$. This dimension reduction is consequential when $q \ll m$, as is often the case when working with higher dimensional trait data.

In the first variance component, K_G is a genetic relatedness (kinship) matrix between in-

dividuals computed by background genetic markers, and K_C is a trait kernel generated by extra information associated with the traits; in MET, the trait kernel can be a relatedness matrix derived from high-dimensional environment (or site) information such as daily minimum, or maximum temperature, precipitation, etc. We show that for MET, the variance component, R^v, composed of many infinitesimal $G\times E$ effects, has variance proportional to $K_C \otimes K_G$ (**Supplementary Note**). τ^2 is a scalar parameter for reducing the dimension of a covariance matrix, which is unknown in other studies, by assuming it is proportional to K_C. If no information is available on the multiple traits, then $K_C = I$ is advisable.

The second variance component assumes correlations among traits but independent and identically distributed (iid) noise between individuals. The common error covariance matrix is Σ which assumes to be unstructured (no constraints). The two kernel matrices are precomputed from *a priori* information, so that only τ^2 and Σ are estimated. The MLMM in FlxQTL is then modeled (in vectorized form) by a multivariate normal distribution whose covariance matrix is the sum of two Kronecker products for the two independent random effects:

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

where the superscript v denotes the vectorization of a matrix. LOD scores ($-\log_{10}(LRT)$) are calculated by log-likelihood difference between the full model and null. We use permutations (at least 1000) to calculate empirical genomewide LOD thresholds as follows: 1. The traits are rotated by the eigenvectors of K_G (and $K_C(\neq I)$). 2. The rotated traits are divided by the square root of their variance-covariance matrix to make them iid. 3. They are shuffled by row and transformed back by multiplying by the square root of the covariance matrix. 4. A genome scan is performed

on this shuffled trait, and the maximum LOD for each permutation is stored. Note that LOD scores using trait covariates (Z) are lower than those implemented by conventional MLMMs (i.e. $Z = I$), but the genomewide significance threshold is also correspondingly lower.

Computational costs. The eigen decomposition of the kernel matrices is done only once; this has computational complexity $O(n^3 + m^3) \approx O(n^3)$ $(m << n)$, the same as existing MLMMs. A LOCO (Leave One Chromosome Out) scheme [7,9,25] has complexity $O(dn^3) \approx O(n^3)$ $(d \ll n)$ for d chromosomes and does not change complexity appreciably. In practice, genome scans with the LOCO scheme take longer depending on the data (about 30 % longer) than that without LOCO. For $t_0 \gg t$, where t_0, t are the maximal number of iterations of ECM and ECM embedded in the Speed restarting Nesterov's scheme, respectively, the computational complexity of two step implementation, which runs ECM for finding 'good' initial values followed by ECM embedded in the 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 computational 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. FlxQTL is implemented in Julia [26], a relatively new programming language, well-suited for developing and implementing algorithms for large datasets. Julia has a just-in-time (JIT) compiler, which allowed us to prototype algorithms with an ease comparable to scripted languages like R, and with a speed approaching compiled languages. Julia has native support for distributed and parallel computing, and users can take advantage of multiple cores or a high-performance computing (HPC) system to achieve a significant speedup. Genome scan with

millions of markers involving moderately high-dimensional traits is possible.

Data processing and analysis. *Mouse HS1940 data.* SNPs were filtered out by minor allele frequency (MAF) less than 2% since we found GEMMA automatically processed the data with that criterion. For a fair comparison, we used a centered genetic relatedness matrix computed by GEMMA, but it needed some adjustment by adding very small value $(0.00001 * I)$ to make sure the matrix should be positive definite. The trait data consisted of 3 traits with missing values and was imputed using R package, mice, with a 'pmm' option (predicted mean method). To compare the computation time between FlxQTL and GEMMA as a function of the number of traits and SNPs, we created new trait (triplets) by shuffling the trait data by individual preserving the correlation among three traits and added them to the existing traits. This produced 6, 9, and 12 simulated traits. We also generated 19 cumulative genotype datasets from the mouse data, that is, starting from a chromosome 1 SNP dataset (950 SNPs), building up a new dataset with 4-chromosome increments and reaching the whole SNP dataset (10783 SNPs). The median time for the genome scan without LOCO was obtained from 64 runs on each simulated trait data.

Gough Island mouse data. This is a F_2 intercross between Gough Island mice and WSB/EiJ. We computed genotype probabilities, excluding sex chromosomes, from R/qtl adding pseudo markers in every 1cM between two markers. Traits were imputed using R-mice package with the 'pmm' option and then were standardized by overall mean and standard deviation. We employed a 4 degrees-of-freedom cubic spline from R-splines package to produce trait covariates (Z) to capture the trend of weekly body weights and corresponding weight growth rate simultaneously. Here, we set the trait kernel to be an identity matrix. A genetic relatedness matrix was computed from

genotype probabilities using a linear kernel. The thresholds at $\alpha = 0.1, 0.05$ were estimated from 1000 permutations. The results from the one-dimensional multivariate genome scan by FlxQTL was largely consistent with that [2] from R/qtl using univariate linear regression, except for Chromosome 6 and 7 perhaps due to polygenic effects. Since one QTL in Chromosome 7 was known to be crucial, we implemented a two-dimensional genome scan for Chromosome 7, as well as 8 and 10 for comparison. FlxQTL detected several more QTL than R/qtl without requiring further analysis such as multiple-QTL analysis.

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. The trait data were imputed using the 'pmm' option in R-mice package and were standardized by column. A total of 699 markers across 5 chromosomes were genotyped for 400 recombinant inbred lines, and a fitness trait was measured at two sites (Italy and Sweden) for three years (6-dimensional multivariate trait). The trait covariate matrix (Z) set to be a matrix with an intercept column, and a contrast between sites (1's for Sweden and -1's for Italy). We computed both a genetic and a climatic relatedness matrix using Manhattan distance. The climatic relatedness matrix was computed using daily range soil or air temperature, precipitation, etc. However, the difference between with and without climatic relatedness matrix appears to be small in our simulation studies (**Supplementary Fig. 7**). 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 QTL for each scan, respectively. The 95% 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 α in the case of no QTL and with the search restricted to models with no more than one QTL [27]. Our result revealed one more significant QTL in three chromosomes each but one less in two chromosomes each, by and large, agreeing with the existing result, with improved interpretation and without requiring secondary analysis of QTL across chromosomes [22].

Switchgrass (Panicum virgatum) data. Genotype probabilities were computed in every 1cM between two markers via R/qtl. Traits and climate information data had no missing values, so that the whole dataset comprises 6118 genetic markers by 750 individuals for 36 quantitative traits of flowering time as a combination of 10 latitudinal sites and 4 years from 2016 to 2019, as well as a matrix of 12 by 36-monthly photoperiod. Three significant QTL (two in Chromosome 5N and one in Chromosome 4K) were selected by 1D-genome scan with the LOCO option. For each QTL, an environment scan was performed by regressing each monthly photoperiod factor under the alternative of existing an environmental factor that affects a QTL after the null scan. For computational efficiency, traits and monthly photoperiod data were centered by column means and scaled by overall standard deviations of the centered data.

Simulation study. We simulated data with the same overall structure and features of the Arabidopsis dataset [22]. A 6-dimensional multivariate trait was simulated with random GxE effects based on genotype data and daily soil minimum and maximum temperature data from those sites and years. Soil daily range temperatures (the difference between maximum and minimum temperatures) were rearranged into a 3 year by 2 location combination, i.e. 365 days by 6 environments, and were standardized by overall mean and standard deviation. To generate null trait datasets, we

selected at random about 3.5% of the genetic markers (about 25 markers of total 699 markers) and 8.2% of 365 soil daily range temperatures (about 30 days) and gave their interaction effect sizes drawn from a normal distribution with zero mean and variance τ^2 that varies from small to large (**Supplementary Note**). An iid noise indicating random error was added to the null trait data; the covariance matrix of the error term was varied. Both random and error terms were respectively standardized. To generate a fixed effect QTL to the null data, we sampled one genetic marker not selected among the random effects above and multiplied it by fixed effect sizes varying from small to large and the orthonormal site contrast matrix (Z). For simplicity, we only considered one fixed effect 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 trait kernel. The first simulation study to assess power for the effect of climatic relatedness varied with values of τ^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. Within that range, simulations were performed to establish a comparison between inclusion and exclusion of the climatic relatedness matrix (K_C) changing τ^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 trait covariates (Z) was implemented in the similar manner. In the given range of fixed effect sizes, we compared three cases: non-identity contrasts (site contrasts in FlxQTL), an identity contrast ($Z = I$ in FlxQTL), Julia ver-

sion MLMM. Note that the Julia version of MLMM was developed as an alternative of GEMMA for ease of comparison with our method.

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

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

3. 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).

4. 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).

5. 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).

6. 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).

299 7. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nature*
300 *methods* **8**, 833 (2011).

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

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

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

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

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

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

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

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

16. 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).

17. Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M. & Wray, N. R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540–2542 (2012).

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

19. Meyer, K. Wombat-A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *Journal of Zhejiang University Science B* **8**, 815–821 (2007).

20. 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).

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

22. 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).

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

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

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

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

27. 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).

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

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

30. 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).

Acknowledgements This work was funded by NIH grants R01GM123489 (S.S., H.K.) R01AI121144, and P30DA044223 (S.S.); Department of Energy award DESC0014156, IOS0922457 and IOS1444533 (T.E.J.); NSF Awards DEB1022196 and DEB1556262 (J.K.M.). The work conducted by J.L. was funded by

354 the US Department of Energy Joint Genome Institute, which is supported by the Office of Science of the US
355 Department of Energy under Contract No DE-AC02-05CH11231. We thank Jon Ågren and Grey Monroe
356 for providing climate data from the *Arabidopsis thaliana* reciprocal transplant experiments; K. Broman and
357 B. Payseur for the Gough Island mouse data.

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

359 **Author Contributions** S.S. conceived the idea; H.K. and S.S. designed the study. H.K. wrote the first
360 draft of the paper, implemented the algorithms and analyzed the data. G.F. contributed to developing the
361 software, J.K.M., T.E.J. and J.L. contributed data and interpreted the results, and all authors revised the
362 manuscript.

1 Supplementary Figures

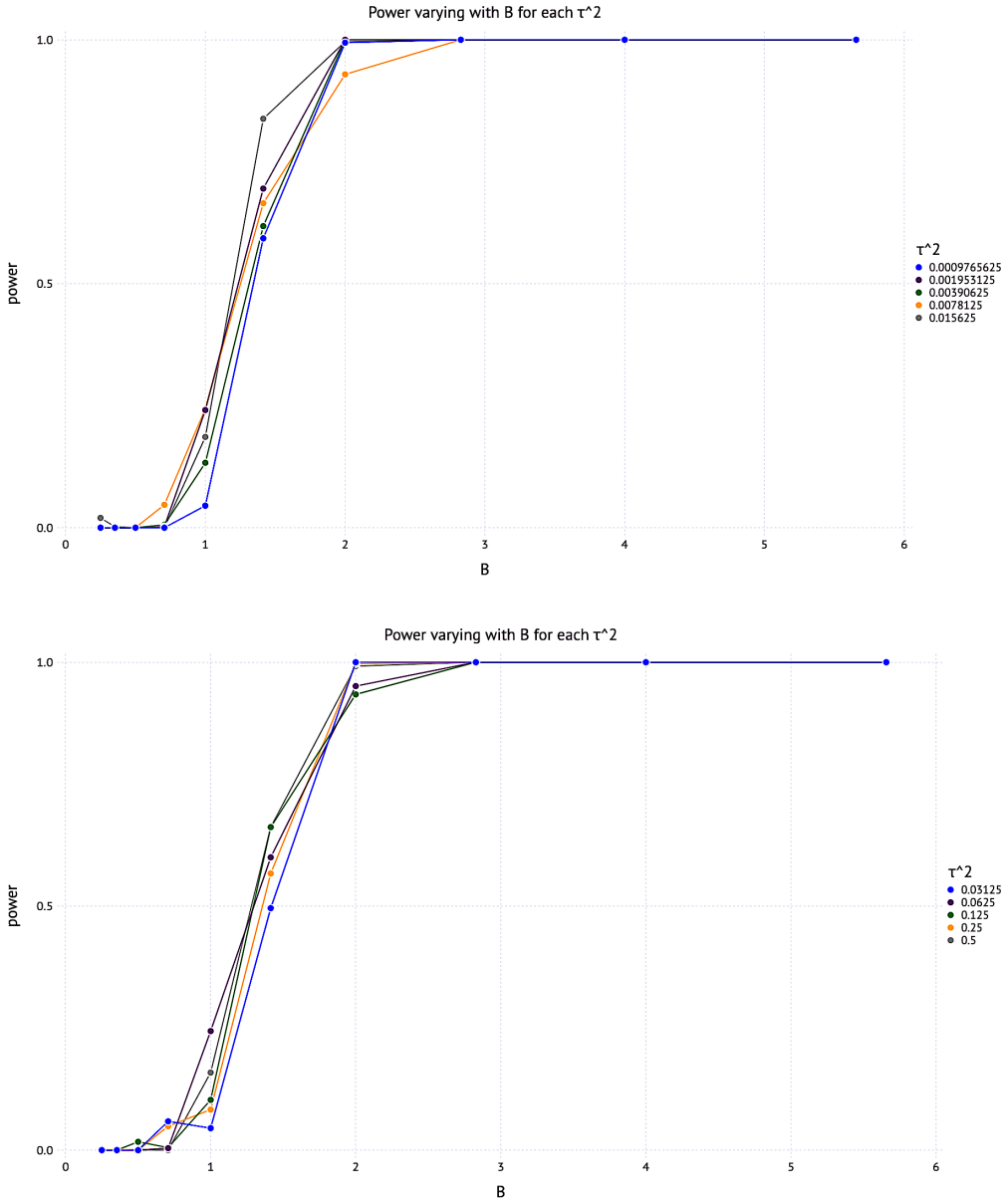


Figure 3: Power varied by τ^2 on the simulated data from Arabidopsis thaliana. The climatic relatedness matrix (K_C) as a trait kernel was computed by 3 year by 2 site soil range temperature data. True effect size matrices B's varied by the following formula: $(\sqrt{2})^{j-1}B_0$ $(j = 1 \cdots 10)$ for B_0=[0.25 -0.25;0.25 0.25], randomly selected from $\{\pm 0.25\}$.

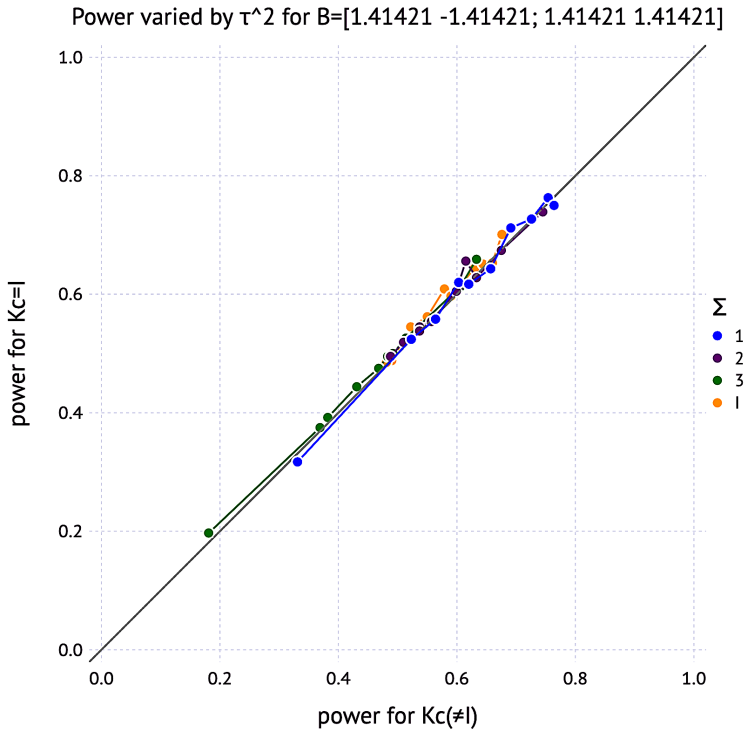


Figure 4: Power comparison of a non-identity K_C and an identity $K_C(= I)$ on the simulated data from Arabidopsis thaliana depending on unstructured environment errors (Σ). K_C was computed by 3 year by 2 site soil daily range temperature data. 6×6 matrices of Σ_i for $i = 1, 2, 3$ are selected using following 3×3 sub-matrices: A, C (1's on the diagonal, a, c on off-diagonal entries, respectively), B (b in all entries). Σ_1 = [A B;B A], where $a = 0.4$, $b = 0.07$, Σ_2 =[A B;B A], where $a = 0.3$, $b = 0.07$, Σ_3=[A B;B C], where $a = 0.3$, $b = 0.0$, $c = 0.1$. I is an identity matrix. The result shows subtle differences in power depending on Σ's

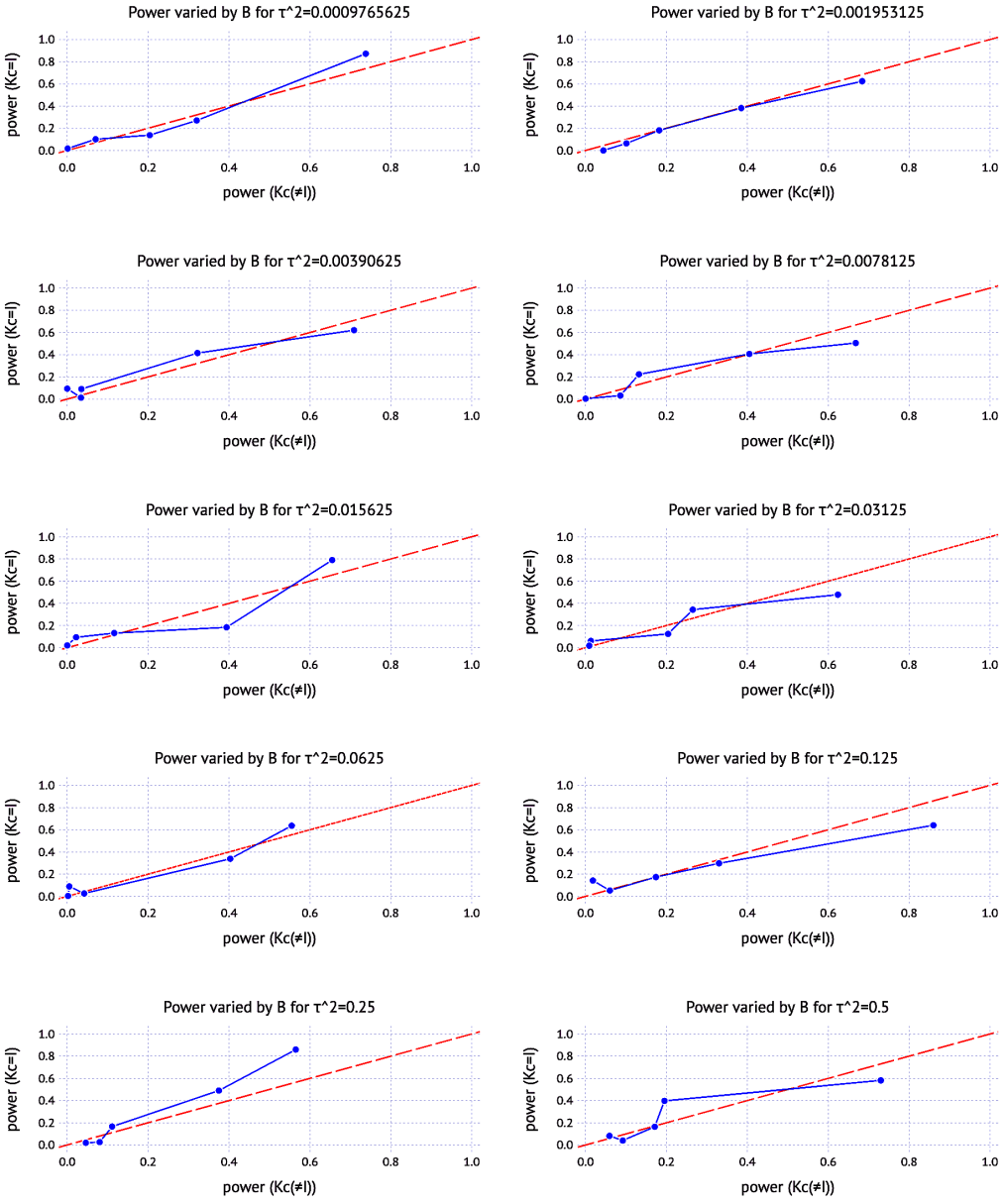


Figure 5: Comparison of power between a non-identity $K_C(\neq I)$ and an identity $K_C(= I)$ on the simulated data from Arabidopsis thaliana. The non-identity K_C was computed by 3 year by 2 site soil daily range temperature data.

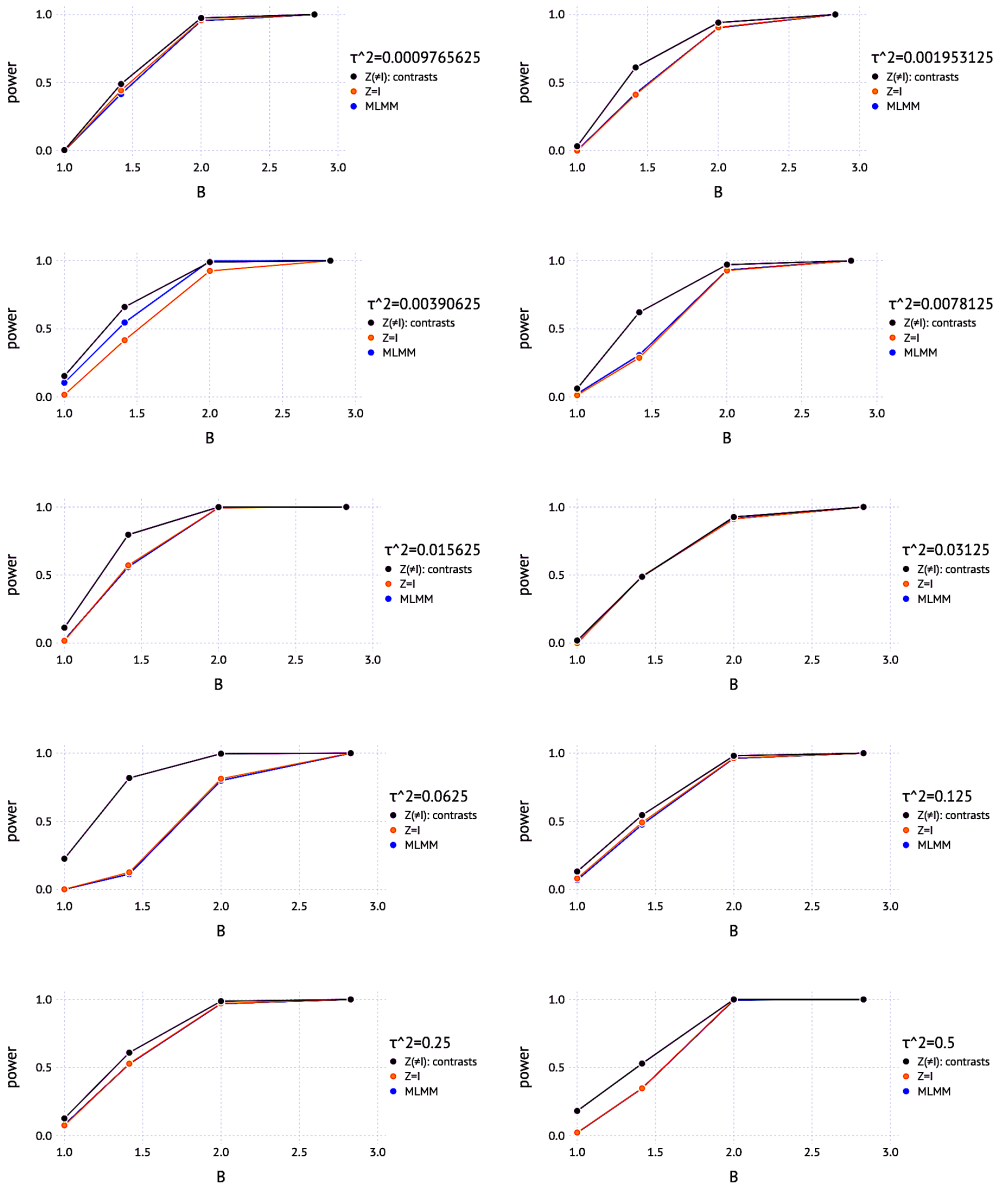


Figure 6: Contribution of trait covariates (Z) to power on the simulated data from the Arabidopsis thaliana. K_C was computed by 3 year by 2 site soil daily range temperature data. 1. $Z(\neq I)$: a 6×2 matrix of contrasts including the intercept and the site contrast of -1's and 1's for Italy and Sweden, respectively. 2. $Z = I$: a 6×6 identity matrix. Both were run in FlxQTL. 3. MLMM: the julia-version MLMM as an alternative for GEMMA.

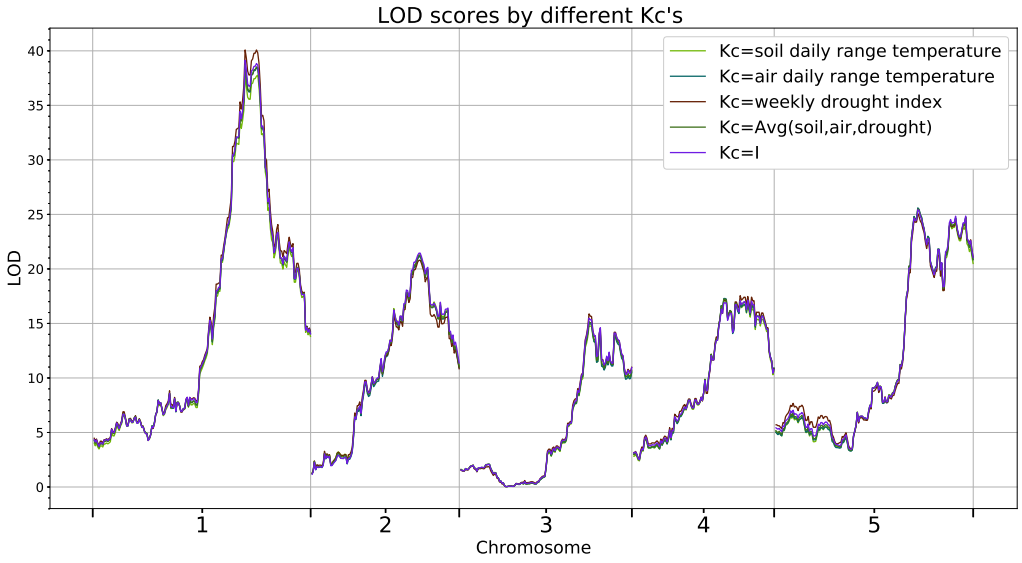


Figure 7: Multivariate genome scan for Arabidopsis thaliana: Comparison of LOD scores depending on various climatic relatedness matrices (K_C). The trait was fitness (mean number of fruits per seedling planted) considered as 6 sites (a 3 year by 2 site combination). K_C's were respectively generated by using soil/air daily range temperatures, weekly drought indices in two sites (Sweden and Italy) for 3 years. The trait covariates (Z) were contrasts containing 1's for overall mean for 6 sites and -1's (Italy), 1's (Sweden) for mean difference between sites to measure GxE interactions. The result appears to be insensitive to the choice of K_C's.

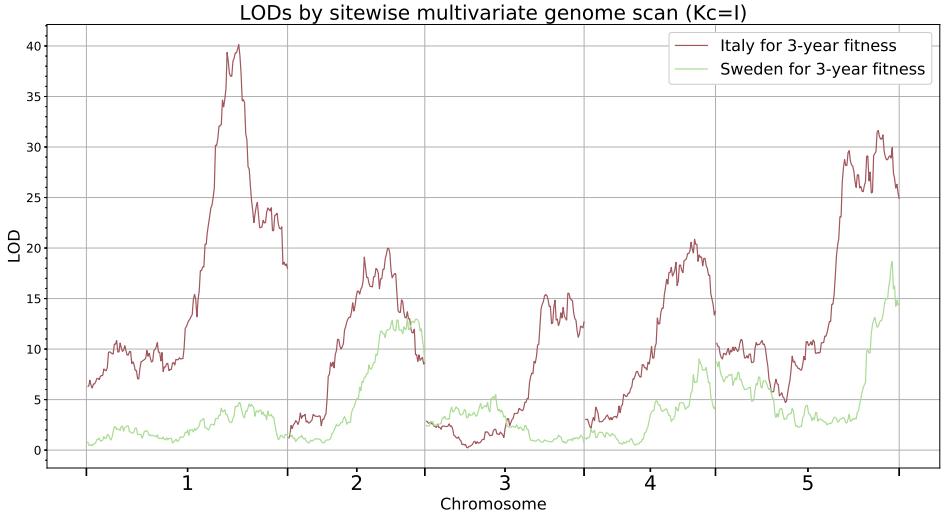


Figure 8: Site-wise multivariate genome scan for Arabidopsis thaliana: Comparison of LOD scores for Italy and Sweden, where $K_C = I$ due to the result in Supplementary Figure 7.

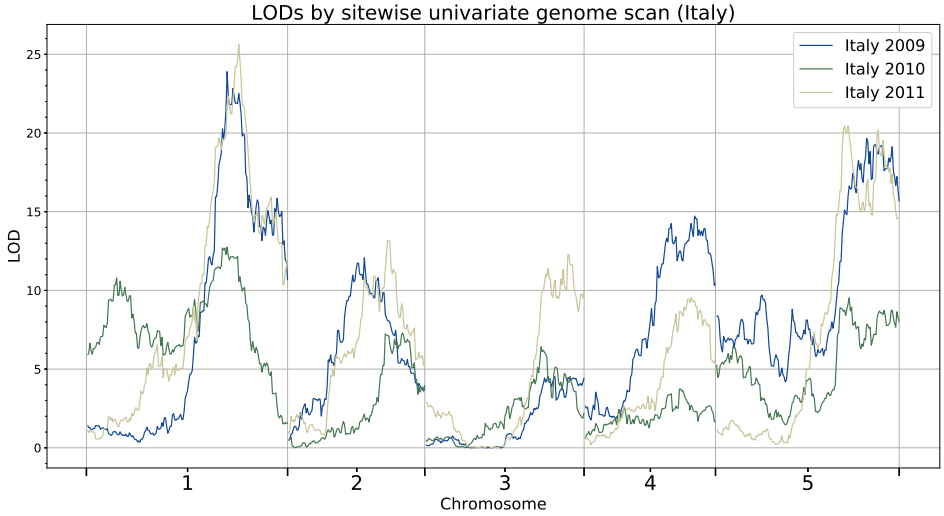


Figure 9: Univariate genome scan for Arabidopsis thaliana in Italy from 2009 to 2011.

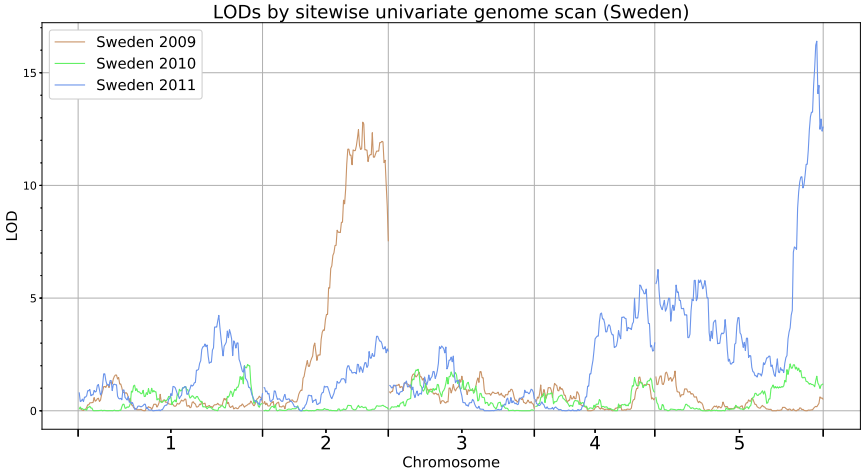


Figure 10: Univariate genome scan for Arabidopsis thaliana in Sweden from 2009 to 2011.

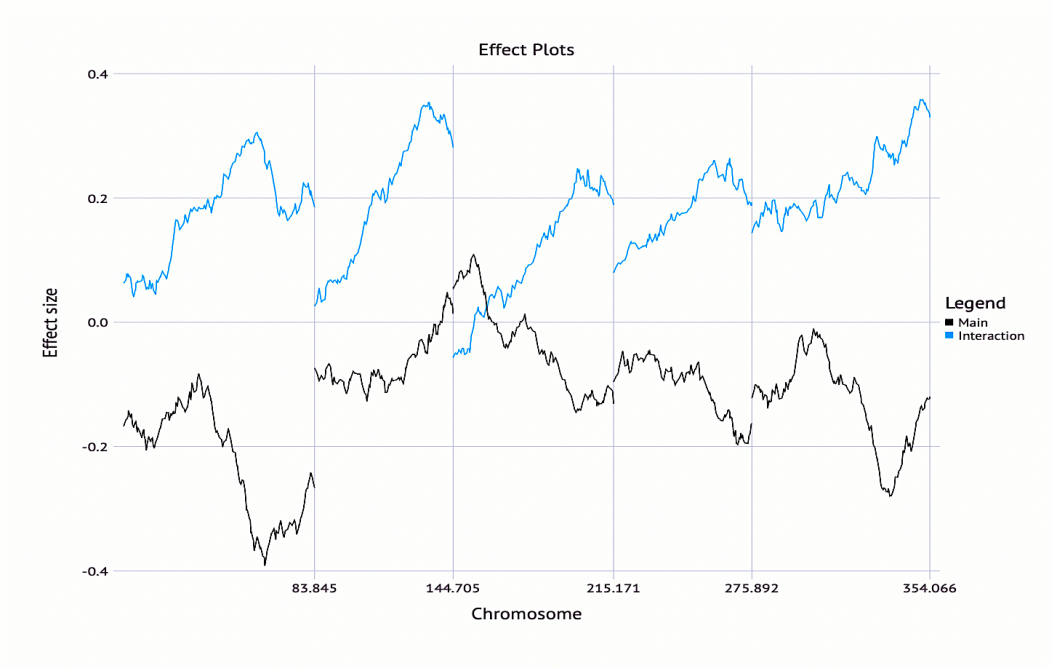


Figure 11: Effect plots for Arabidopsis thaliana with $K_C = I$ for 6 quantitative traits as described in Supplementary Figure 7. Negative values in most main effects imply that the Swedish allele, on average, underperforms. Positive values in most interaction effects indicates home allele advantage; that is, the Swedish genotype in Sweden performs better than that in Italy.

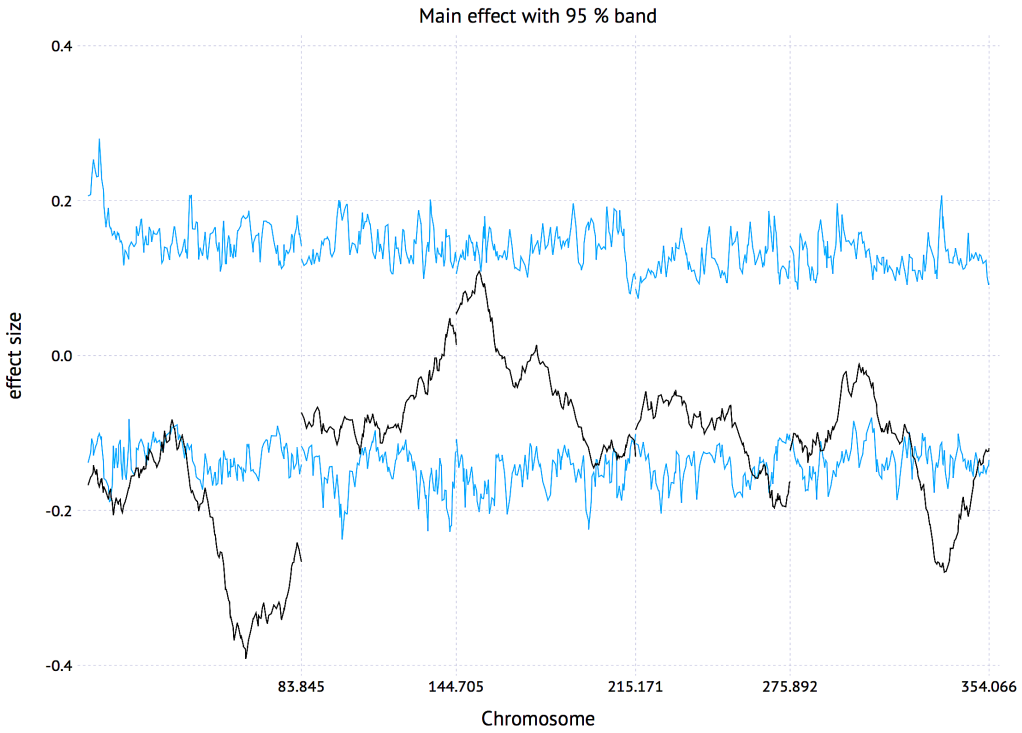


Figure 12: Main effects with 95% band for Arabidopsis thaliana. The band was obtained by 100 permutations

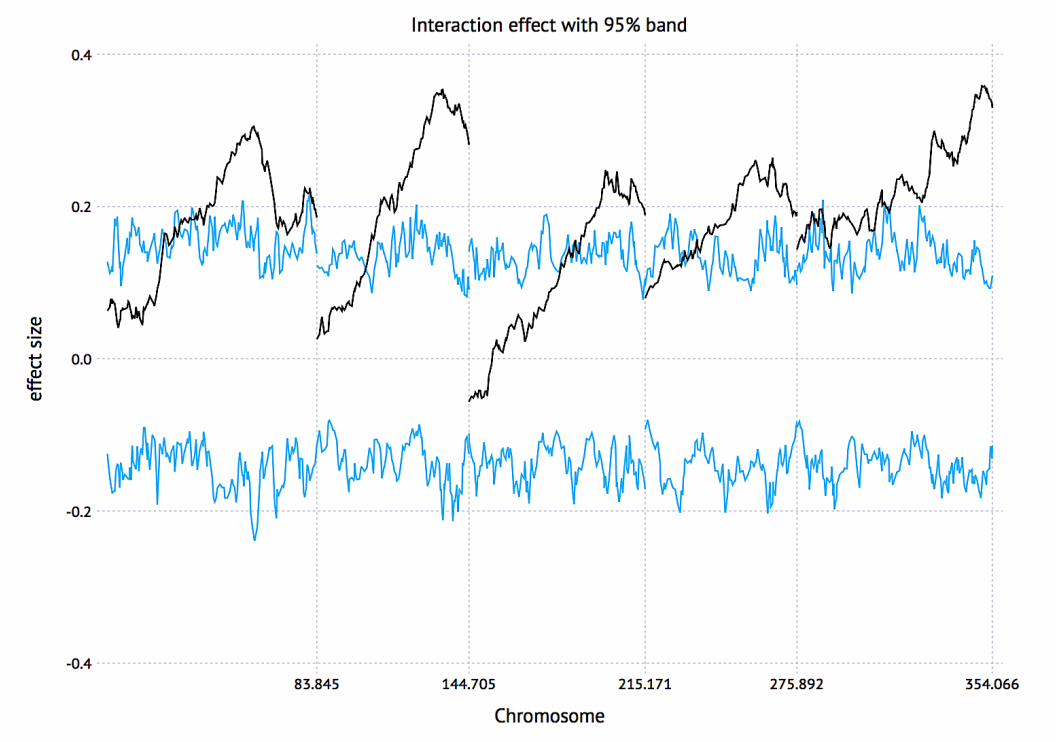


Figure 13: Interaction effects with 95% band for Arabidopsis thaliana. The band was obtained by 100 permutations

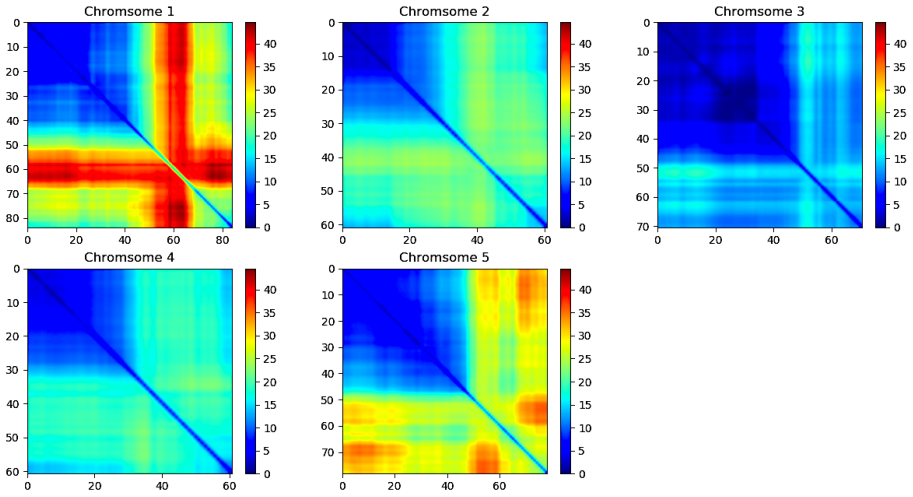
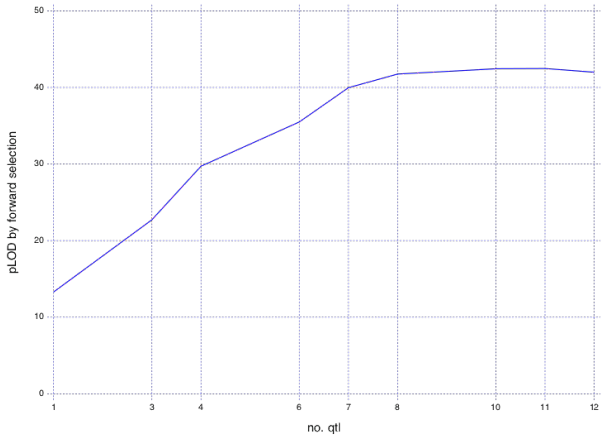
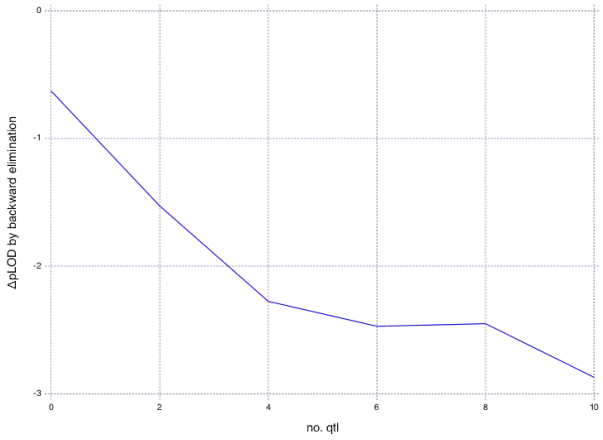


Figure 14: 2-dimensional (2D) multivariate genome scan for Arabidopsis thaliana.

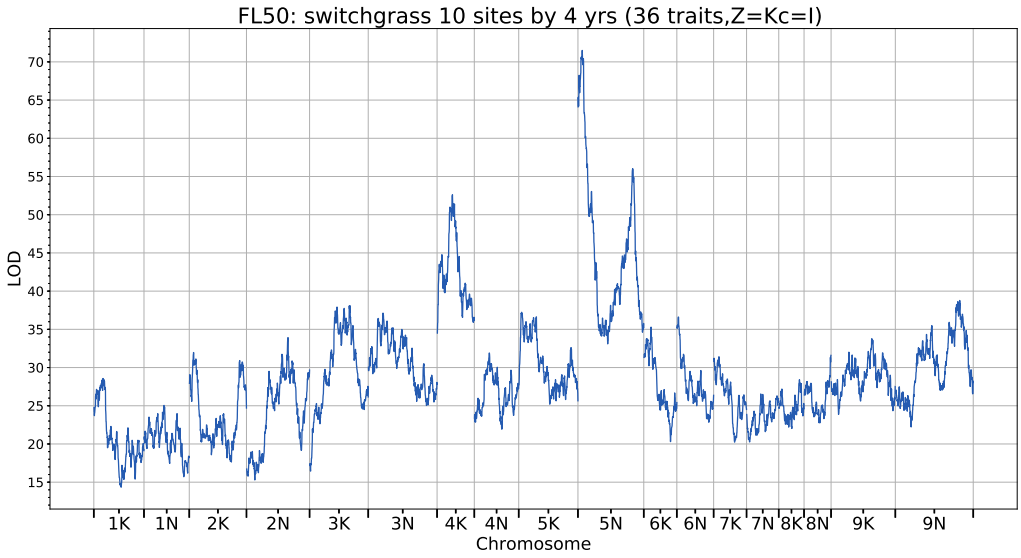


(a) Foward selection

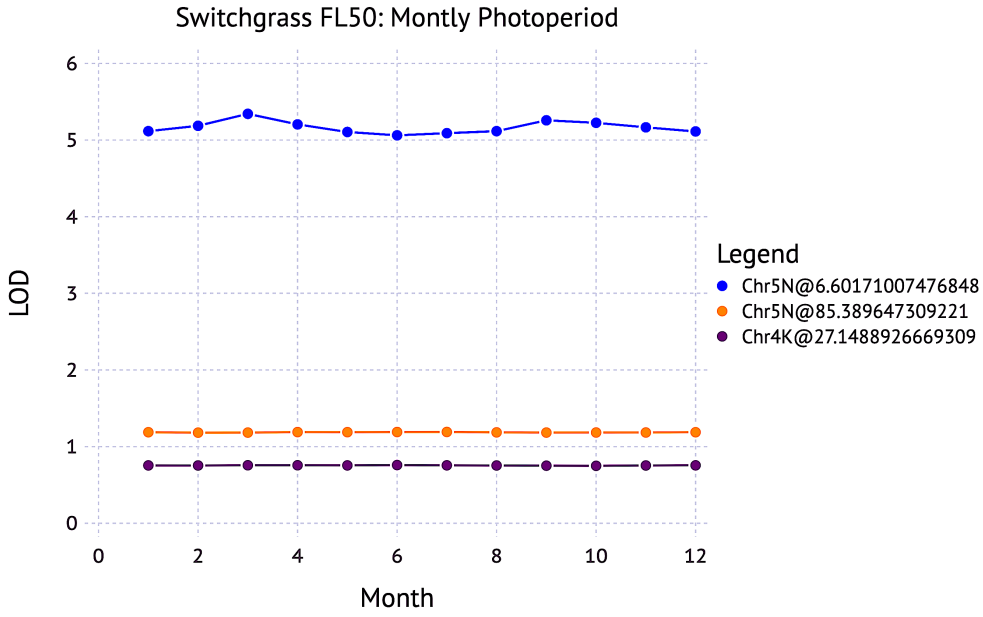


(b) Backward elimination

Figure 15: QTL selection for Arabidopsis thaliana (multiple QTL analysis). A model selection approach by forward selection and backward elimination was performed to finalize significant QTL from candidate QTL selected in Supplementary Figure 7, 14. A selection criterion is the penalized LOD score for additive models equivalent to BIC [27].



(a) 1-dimensional (1D) genome scan



(b) Environment scan by photoperiod

Figure 16: Switchgrass FL50. The environment scan (b) was done for the top 3 QTL selected from the genome scan (a)

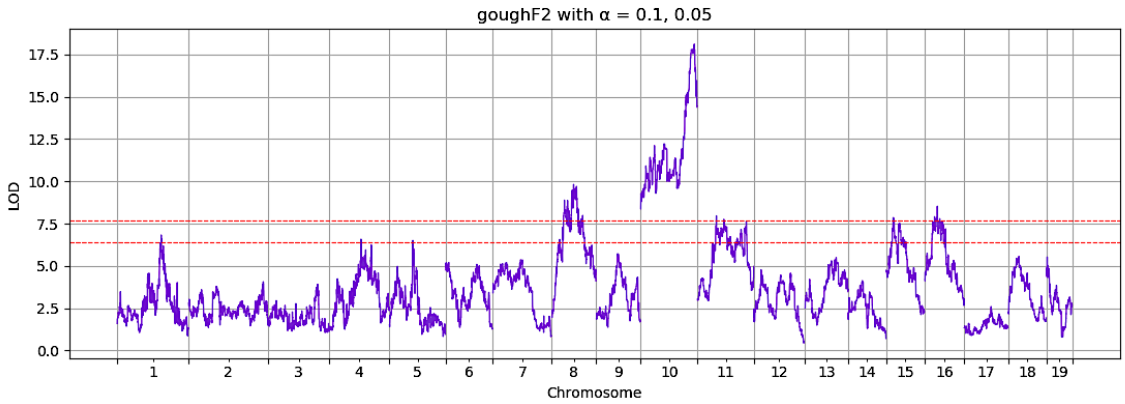


Figure 17: Application to the longitudinal trait data (F2 intercrosses between Gough Island mice and WSB/EiJ): LOD scores and thresholds at $\alpha = 0.1, 0.05$

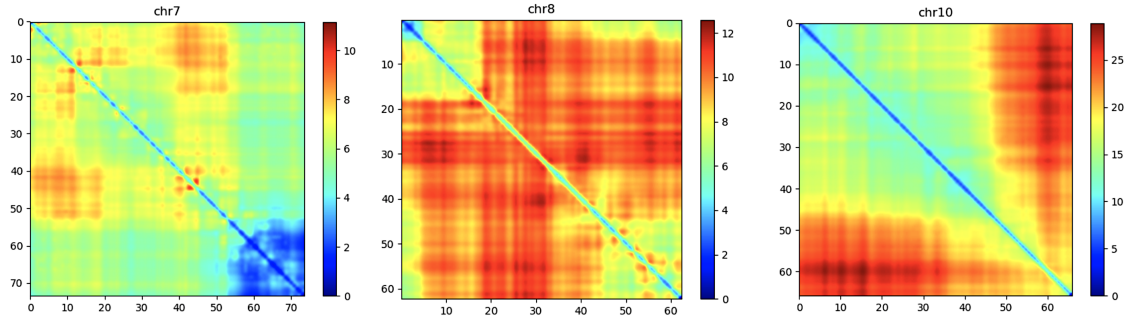


Figure 18: Application to the longitudinal trait data (F2 intercrosses between Gough Island mice and WSB/EiJ): LOD scores by 2D genome scan

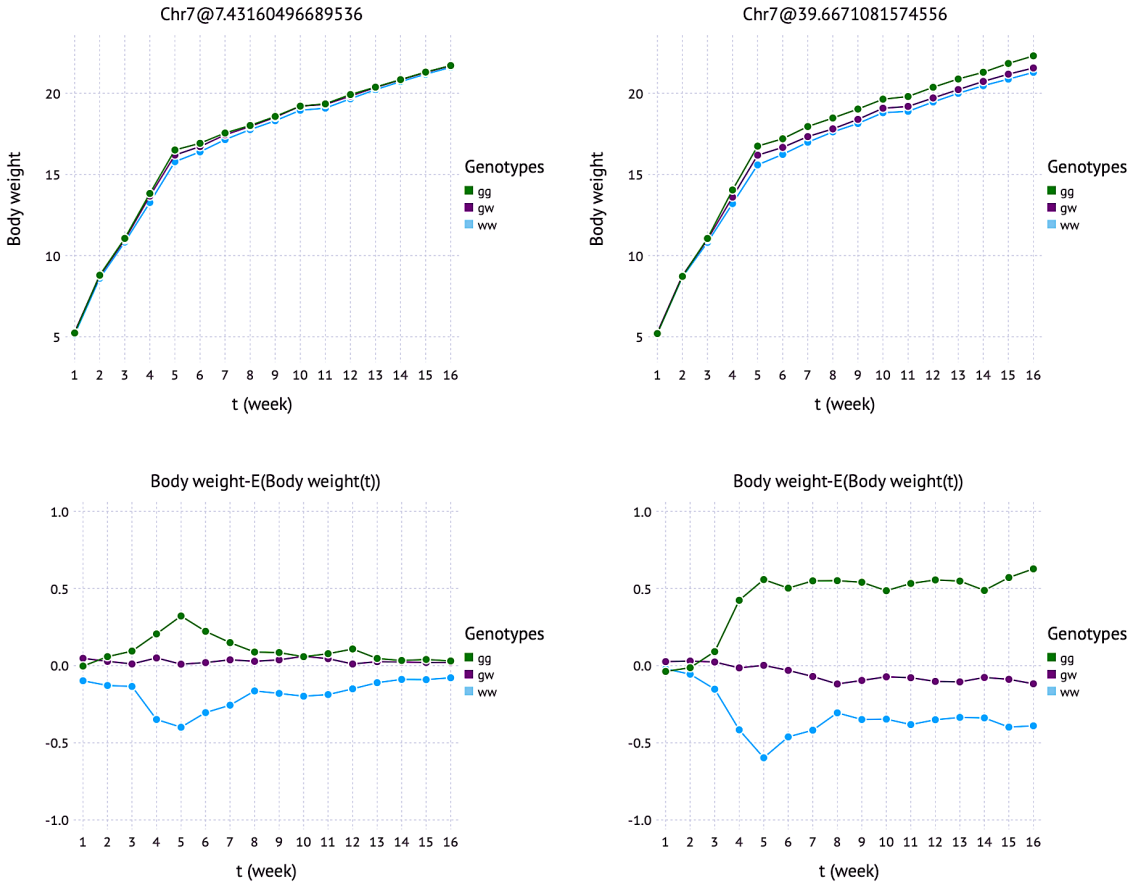


Figure 19: Application to the longitudinal trait data (F2 intercrosses between Gough Island mice and WSB/EiJ): Body weight effects by genotypes and corresponding deviations from weekly overall mean in Chromosome 7.

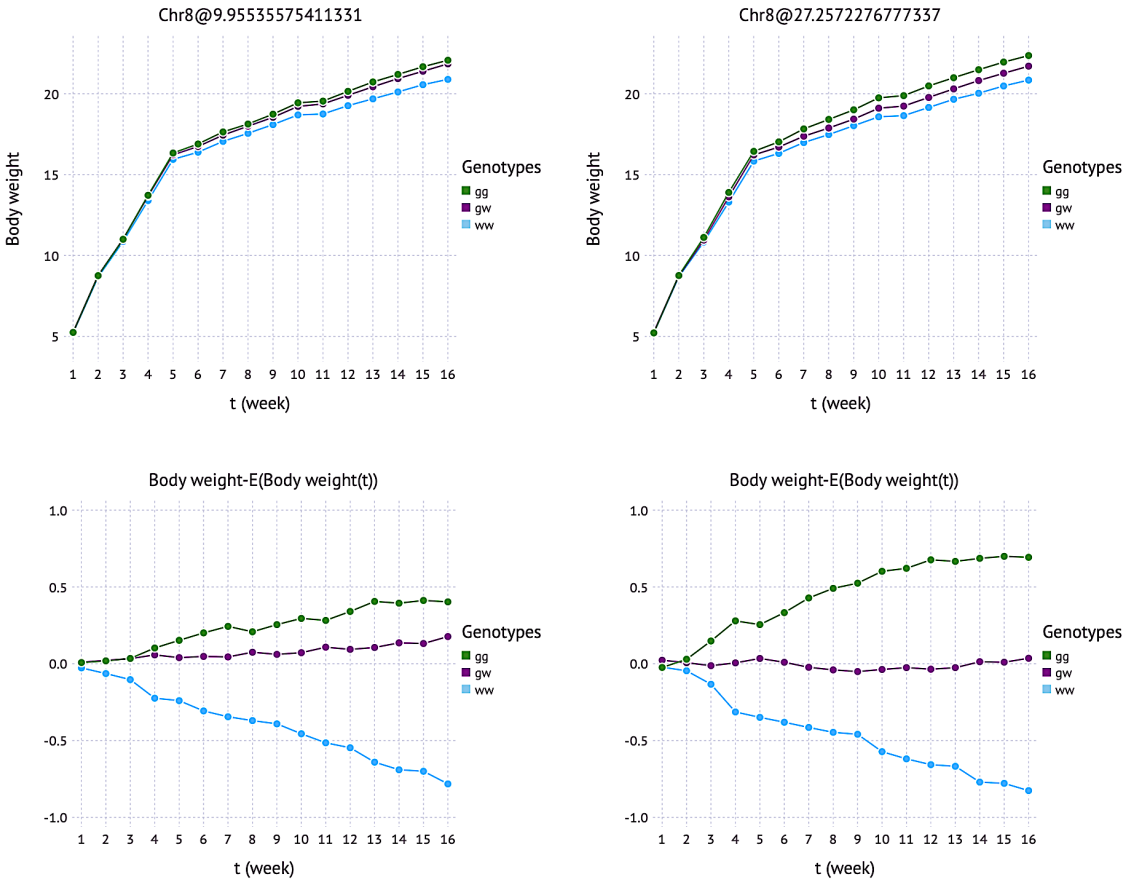


Figure 20: Application to the longitudinal trait data (F2 intercrosses between Gough Island mice and WSB/EiJ): Body weight effects by genotypes and corresponding deviations from weekly overall mean in Chromosome 8.

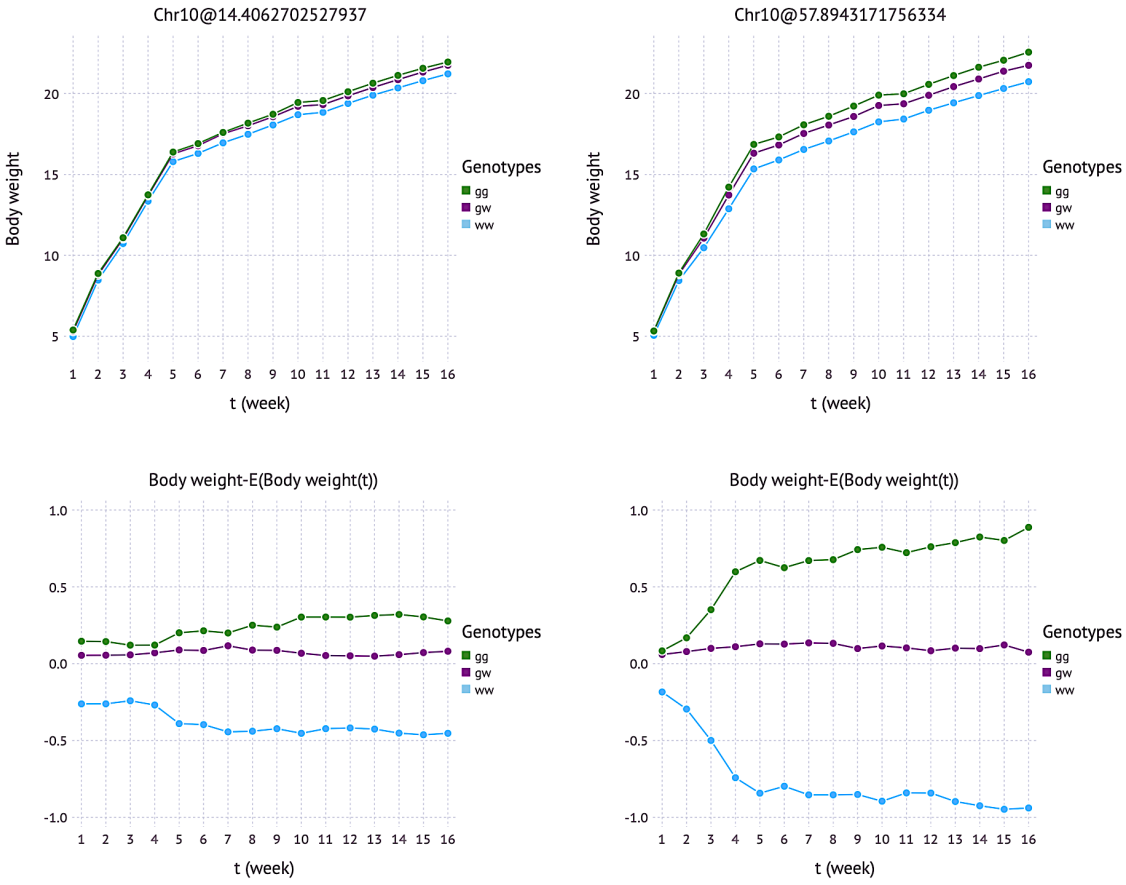


Figure 21: Application to the longitudinal trait data (F2 intercrosses between Gough Island mice and WSB/EiJ): Body weight effects by genotypes and corresponding deviations from weekly overall mean in Chromosome 10.

2 Supplementary Tables

Adding 1 or 2 QTL	Δ pLOD	pLOD	Deleting 1 or 2 QTL	Δ 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 1: QTL selection for Arabidopsis thaliana (multiple QTL analysis) as in Supplementary Figure 15. 11 significant QTL were finalized since the largest decrease (Δ pLOD = -2.872) in the backward elimination step when dropping one QTL from 11 QTL in the forward selection step occurred.

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 2: Significant QTL for Arabidopsis thaliana as described in Supplementary Figure 15 and Supplementary Table 1.

3 Supplementary Note: Mulitivariate Linear Mixed Model for FlxQTL

The *Flex*ible multivariate linear mixed model for *QTL* mapping (FlxQTL) extended from a standard multivariate linear mixed model (MLMM) [28] is defined as

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

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 $n \times m$ matrix of multivariate responses (multiple traits or a trait in multiple environments), where n is the number of individuals (lines), m is the number of environments (or time points). X_0 is a $n \times p$ matrix of genotypes including the intercept (or genotype probabilities) for a marker to be tested and optionally contains individual level covariates such as sex and age. Z_0 is a $m \times q$ matrix of q trait covariates such as (site) contrasts, (orthonomal) basis functions and so on. $B_{p \times q}$ is then fixed effects to be estimated. R_0 is a matrix of genetic random effects, and E_0 is a matrix of residual errors. K_G is a $n \times n$ genetic relatedness (kinship) matrix, and K_C is a $m \times m$ trait kernel computed by information associated with the traits. In multi-environment trials (MET), the trait kernel can be a climatic relatedness matrix using high-dimensional environment information such as daily minimum or maximum temperature, precipitation, etc. τ^2 is a scalar by parameterizing a $m \times m$ convariance matrix proportional to K_C for dimension reduction. Note that we confine K_C

and K_G to full-rank and symmetric positive definite matrices. If they are semi-positive definite, one can take advantage of a Shrinkage approach [29,30] to force it to be positive definite. The eigen-decomposition to the matrices yields $K_G = U_G \Lambda_G U_G'$ and $K_C = U_C \Lambda_C U_C'$. Multiplying $U_C' \otimes U_G'$ to the model gives

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

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

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 $E_{n\times m} = [\mathbf{e}_1', \mathbf{e}_2', \cdots, \mathbf{e}_n']'$.

Or, this is simplified as a vectored form,

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

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

Verification of multivariate random interactions Assume that a multivariate random interaction term has a form of $R_0 = GB_{rand}C'$, where $C_{m\times q_1}$ is q_1 background (or high-dimensional) trait (or environment) information, $G_{n\times p_1}$ is p_1 background genetic markers such as genotypes and genotype probabilities, and $B_{rand}^v \sim MVN(0, \tau^2 I_{q_1} \otimes I_{p_1})$. Then, the distribution of the multivariate random interactions in (6) is obtained as follows.

$$\begin{aligned} E(R_0^v R_0^{v\prime}) &= E((C \otimes G) B_{rand}^v {B_{rand}^v}' (C' \otimes G')) \\ &= (C \otimes G) E(B_{rand}^v {B_{rand}^v}') (C' \otimes G') \\ &= (C \otimes G)(\tau^2 I_{q_1} \otimes I_{p_1})(C' \otimes G') \end{aligned}$$

$$
\begin{aligned}
&= \tau^2(CC' \otimes GG') \\
&= \tau^2 K_C \otimes K_G,
\end{aligned}
$$

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

4 Supplementary Note: Expectation Conditional Maximization (ECM)

Since R is unobservable, the joint log-likelihood function for the ECM [14,21] is written as follows.

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

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 [28]

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

where

$$
\begin{aligned}
V_i^{(t)} &= \tau^{2(t)}\lambda_i\Lambda_C + \Sigma^{(t)}, \quad \hat{\mathbf{r}}_i = \tau^{2(t)}\lambda_i\Lambda_C(V_i^{(t)})^{-1}(\mathbf{y}_i - ZB^{(t)'}\mathbf{x}_i), \\
\hat{\Theta}_i^{(t)} &= \tau^{2(t)}\lambda_i\Lambda_C - \tau^{2(t)}\lambda_i\Lambda_C(V_i^{(t)})^{-1}\tau^{2(t)}\lambda_i\Lambda_C. \qquad (12)
\end{aligned}
$$

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

$$
E\left(l(Y, R|B, \tau^2, \Sigma)|R|Y, B^{(t)}, \tau^{2(t)}, \Sigma^{(t)}\right)
$$

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

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 $\Sigma^{(t)}$ [14].

$$
\begin{aligned}
B^{(t+1)} &= (X'X)^{-1}X'(Y-\hat{R}^{(t)})\Sigma^{-1}Z(Z'\Sigma^{-1}Z)^{-1}, && (14)\\
\tau^{2(t+1)} &= \frac{1}{nm}\sum_{i=1}^{n}tr(\lambda_i^{-1}\Lambda_C^{-1}(\hat{\mathbf{r}}_i^{(t)}\hat{\mathbf{r}}_i^{(t)'}+\hat{\Theta}_i^{(t)})), && (15)\\
\Sigma^{(t+1)} &= \frac{1}{n}\sum_{i=1}^{n}(\hat{\mathbf{e}}_i^{(t)}\hat{\mathbf{e}}_i^{(t)'}+\hat{\Theta}_i^{(t)}), && (16)\\
\text{where}\quad \hat{\mathbf{e}}_i^{(t)} &= \mathbf{y}_i-ZB^{(t+1)'}\mathbf{x}_i-\hat{\mathbf{r}}_i^{(t)}. && (17)
\end{aligned}
$$

Log-likelihood function The joint log-likelihood function for the ECM is simplified in the following way.

$$
\begin{aligned}
& l(Y,R|B,\tau^2,\Sigma)=-nm\log(2\pi) \\
& -\frac{1}{2}\log|(U_C\otimes U_G)(\tau^2\Lambda_C\otimes\Lambda_G)(U_C'\otimes U_G')|-\frac{1}{2}\log|(U_C\otimes U_G)(\Sigma\otimes I_n)(U_C'\otimes U_G')| \\
& -\frac{1}{2}(Y_0-X_0BZ_0'-R_0)^{v'}((U_C\otimes U_G)(\Sigma\otimes I_n)(U_C'\otimes U_G'))^{-1}(Y_0-X_0BZ_0'-R_0)^{v} \\
& -\frac{1}{2}R_0^{v'}((U_C\otimes U_G)(\tau^2\Lambda_C\otimes\Lambda_G)(U_C'\otimes U_G'))^{-1}R_0^{v} \\
= & -nm\log(2\pi)-\frac{nm}{2}\log(\tau^2)-\frac{m}{2}\log|\Lambda_G|-\frac{n}{2}\log|\Lambda_C|-\frac{1}{2}\log|\Sigma\otimes I_n| \\
& +2\log|(U_C'\otimes U_G')|-\frac{1}{2}(Y-XBZ'-R)^{v'}(\Sigma\otimes I_n)^{-1}(Y-XBZ'-R)^{v} \\
& -\frac{1}{2}R^{v'}(\tau^2\Lambda_C\otimes\Lambda_G)^{-1}R^{v}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}\{-m\log(2\pi) - \frac{m}{2}\log\tau^2 - \frac{m}{2}\log(\lambda_i) - \frac{1}{2}\log|\Lambda_C| - \frac{1}{2}\log|\Sigma| \\
&\quad -\frac{1}{2}\mathbf{e}_i'\Sigma^{-1}\mathbf{e}_i - \frac{1}{2}\mathbf{r}_i'(\tau^2\lambda_i\Lambda_C)^{-1}\mathbf{r}_i\}. \qquad (18)
\end{aligned}
$$

Derivation of conditional expected log-likelihood and the closed form solutions To obtain the 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:

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

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

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

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

where C is a constant.

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

$$
\begin{aligned}
\sum_{i=1}^{n}(\mathbf{x}_i\mathbf{x}_i' \otimes Z'\Sigma^{-1}Z)B'^v &= \sum_{i=1}^{n}(\mathbf{x}_i \otimes Z'\Sigma^{-\frac{1}{2}})(\Sigma^{-\frac{1}{2}}\mathbf{y}_i - \Sigma^{-\frac{1}{2}}\hat{\mathbf{r}}_i^{(t)}) \\
B'^v &= ((X'X)^{-1}X' \otimes (Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-\frac{1}{2}})(\Sigma^{-\frac{1}{2}}(Y' - \hat{R}^{(t)'}))^v \\
B' &= (Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1}(Y' - \hat{R}^{(t)'})X(X'X)^{-1} \\
\therefore B &= (X'X)^{-1}X'(Y - \hat{R}^{(t)})\Sigma^{-1}Z(Z'\Sigma^{-1}Z)^{-1}
\end{aligned}
$$

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

418

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

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

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

419

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

420

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

421

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

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

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

423 5 Supplementary Note: Nesterov's Accelerated Gradient with Speed Restart

424 **Algorithm** : ECM embedded in the speed restarting Nesterov's scheme [20]

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

426 $t_{min} \in \mathbb{N}$

$j \leftarrow 1$

while $t = 1, \| B^{(t)} - B^{(t-1)} \| + \| \tau^{2(t)} - \tau^{2(t-1)} \| + \| \Sigma^{(t)} - \Sigma^{(t-1)} \| > \epsilon$

for small enough ϵ

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