

Editor Comments

Multi-manuscript Issues (identical across responses to reviewers)

Editor: Now that you have all three manuscripts potentially going forward for publication in G3, they can be more tightly linked.

We have updated the manuscripts such that each one refers to the other two in the introduction and discussion.

Editor: For instance, one of the other papers could provide the benchmark data that is required for [the software] manuscript. So the software paper can include those files and refer to the other manuscript for details.

We have incorporated this recommendation, using the dataset from Kumar et al. 2013 that was used in one of the case studies in the manuscript titled “Mean-Variance QTL Mapping Identifies Novel QTL for Circadian Activity and Exploratory Behavior in Mice”. The time required to analyze this dataset, according to number of computer cores available and number of permutations desired, is described in the “Performance Benchmarks” section of the software paper and the relevant files are included in its Zenodo repository.

Issues Specific to this Manuscript

Editor: As suggested by Reviewer 1 it will be very useful to have a benchmark study with times/memory requirements as a function of number of markers and sample size. This study will give those interested on using the package an understanding of how this package would scale. It will be important to include in the discussion also elements related to scalability: is the package suitable for very large-p/large-n?

We have added this information in the “Performance Benchmarks” section. See detailed response to Reviewer 1 comment 3.

Editor: There is a very rich literature in the study of genetic control in environmental variance in both plant and animal breeding as well as model organisms. This literature should be mentioned. This comment may also apply to some of the companion papers. [Citation recommendations provided by the editor.]

The editor suggested references that relate to the heritability of environmental variance. This literature was foundational toward the development of the work presented in this manuscript, so we agree with the editor that it should be cited in the introduction. We have added a paragraph to the introduction that describes how recognizing the heritability of environmental variance is a necessary step toward vQTL and mvQTL mapping.

Reviewer 1 Comments

Reviewer: The authors mention “in almost all traits of interest in human health and disease”. The detection of QTL has also been widely used in plant and animal populations.

The reviewer is quite correct and we have edited the introduction to reflect the fact that QTL mapping has been applied productively in plant and animal research.

Reviewer: The authors used a set of simulated data with 3 chromosomes and 11 equally spaced markers. I would like to see an example with a larger genotype data set.

To most clearly illustrate the standard workflow associated with package `vqt1`, the authors have demonstrated its usage on a simulated cross with an unrealistically-small genome. The authors are encouraged by the reviewer's enthusiasm to see the package applied to real datasets. This manuscript has been submitted as part of a three-paper companionship, where each of the other papers contains an analysis of real data using package `vqt1`. Thus, the authors contend it is appropriate to maintain this paper with only limited reference to real data — a real dataset is used in in the first performance benchmark, as described below.

Reviewer: The CPU time is indicated for the data set of the example. I would like to have the CPU time information in a more realistic data set. Perhaps authors can use a free data set from the web.

Thank you for this recommendation. The authors agree that including this information will make the paper more useful to geneticists who are considering using the package — we have added two sets of benchmarks.

First, we demonstrate the performance of the package according to number of computer cores used and number of permutations conducted. For this set of benchmarks, we used the intercross from Kumar et. al 2013 which we reanalyzed in [Corty et al 2018+](#). This intercross population has 244 mice and was genotyped at 93 markers. After applying the hidden Markov model in `R/qt1` with `step.size` of 2, the number of loci to be tested was 582. As recommended by the editor, this example helps tie together the companions, because the results of this analysis are presented in the manuscript titled “Mean-Variance QTL Mapping Identifies Novel QTL for Circadian Activity and Exploratory Behavior in Mice”. This addition can be found on page 4.

Second, we demonstrate the performance of the package according to the size of the mapping population and the number of genetic markers, holding constant the number of permutations at 1000 and the number of cores at 32. For this set of benchmarks, we used crosses simulated with `R/qt1` so we could control the population size and number of genetic markers. This addition can be found on page 4.

Reviewer: In the simulation, the authors should indicate the magnitude of the genetic variance explained by the QTL.

Thank you for this recommendation. We have added this information to the manuscript and it now appears on page 2.

Reviewer: I think the authors should also indicate if the software is capable in terms of computing power to analyze the data from SNP genotyping devices or even sequencing.

Thank you for this recommendation. In short, the workflow presented here, which relies on permutation to assess statistical significance, would not be practical with a dense genotyping array with > 10,000 markers, such as the MegaMUGA ($\approx 78,000$ markers) GigaMuga ($\approx 150,000$ markers), or the Mouse Diversity Genotyping array ($\approx 623,000$ markers). We have addressed this question in the “Performance Benchmarks” section, starting on page 4 and continuing onto page 5.