**Gene-level quantitative trait mapping in *C. elegans*.**

Luke M. Noble, Matthew V. Rockman, Henrique Teotónio

**Genetic material**

Outbred populations and recombinant inbred lines will be distributed by the authors on request.

**Genotype data**

All raw read data is available from the NCBI SRA under BioProject [PRJNA557613](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA557613). Processed genotype calls are available as supplementary data and from <http://lukemn.github.io/cemee>.

S1\_WS220\_CeMEEv2\_founder\_geno.csv.gz – founder GATK genotype calls filtered to 354,063 diallelic SNPs segregating in RILs, coded [0,1,2], against the WS220 N2 reference genome.

S2\_WS220\_CeMEEv2\_ RIL\_geno.csv.gz – RIL genotypes, as above. Heterozygous GATK calls are coded as 0.5, uncertain HMM imputations are >0 and <1.

**Simulation code**

S3\_additive\_qtl\_simulations.R – R functions to filter the raw genotype data, simulate additive polygenic architectures, run linear mixed model association tests, define quantitative trait loci, and make summary plots. In addition to the packages listed within, helper functions in util/CeMEE\_util.R, and supplementary data files S1, S2, S6, S7 and S8 are required. A wrapper for formatting and filtering raw data and running MultiTrans to generate significance thresholds is in util/runMultitrans.sh.

S4\_2D\_simulations.py – functions to filter the raw data, simulate phenotypes under epistatic genetic architectures (calling the helper function util/simAA.R), fit hierarchical linear models, and save test null and alternate likelihood ratio statistics. Empirical p-values are then calculated with S5\_2D\_simulation\_pvals\_MAFQ.py. Tested against Python 3.6.3, with limix 2.02, limix-core 1.0.2, limix-lmm 0.1.2.