

Online Resource 7 FlexQTL parameter settings and justifications for values chosen

FlexQTL™ parameters	Parameter description	Value	Value description
indiN	Indicator for how nuisance should be used	-1 or 0	-1 = Exclusion 0 = use as co-factor with uniform distribution
The column defining the co-factor values for each individual relative column is part of the data input file. This column was activated in co-factor analyses with <i>Ma-Q₂</i> as a co-factor. Co-factor values were 0–3, determined by the <i>B</i> -allele dose of SNP marker ss475881815 augmented by 1; by default, missing values are indicated by value “0”.			
indicQTL	Indicator for QTL or IBD analysis	1	1 = QTL analysis 2 = IBD analysis
FlexQTL™ software conducts either a QTL or an IBD (Identity-By-Descent) analysis. We performed a QTL analysis.			
maximQTL	Maximum number of QTL to be included in the model	5	Up to 5 QTLs allowed in each run
This is a random value based on the presumed complexity of the trait under analysis. Based on literature, we did not expect more than 2 major QTLs influencing acidity. We typically choose to identify and further analyze QTLs only when they have a certain level of statistical evidence ($2\ln BF > 5$), whereas FlexQTL™ considers regions with a lower thresholds ($2\ln BF > 2$). For this reason, the maximum number is usually doubled. When difficulties are met in reaching convergence, higher maximum numbers can be explored.			
priorQTL	Prior distribution for the number of QTL, the variable being the mean of a (truncated) Poisson distribution	2	
FlexQTL™ is based on Bayesian statistics that require priors. The minimum value is 1. Based on previous information (literature), we chose 2 as prior.			
addGENE	Indicator for additive genetic effects for individual QTLs	5	Additive genetic effects with normal prior distribution
FlexQTL™ is capable of simulating additive and dominance genetic models. For each of these two genetic models there were six options with unique identifiers from 0–5, where 0 = model excluded from analysis. We chose to run additive genetic models with normal prior distributions and random (Co) variance matrix diagonals (5) because this is the standard in previous literature (Bink et al. 2012, 2014; Roach et al. 2016).			
domGENE	Indicator for dominant genetic effects for individual QTLs	0 or 5	0 = Excluded 5 = Dominant genetic effects with normal prior distribution
Only used when testing for the presence of a dominance component, and always in combination with addGENE 5			
length	Length of Markov chain	200,000	200,000 QTL simulations
FlexQTL™ simulates Markov chain Monte Carlo (MCMC) iterations. For statistical inference, the sampling of 1000 models is common in the literature (e.g., Bink et al. 2012, 2014). Applying a thinning of 200 (see below), a chain length of 200,000 was needed to obtain 1000 independent samples.			
Thinning	Thinning of Markov chain with respect to writing samples to file	200	With a chain length of 200,000 this writes 200,000/200 or 1000 models (samples) to an output file
Thinning is needed to allow the sampling of simulations that can be considered as independent from each other. The size depends on the complexity of the trait and data. When results show convergence, the			

thinning was adequate from a statistical point of view. Subsequently, lower values might be explored to reduce chain sizes and thereby reduce computation time.			
markerblock	Number of marker loci within a block that are jointly updated with respect to linkage phase among marker (and QTL) alleles	5	5 markers per block
FlexQTL™ guidelines advise a minimum block size of 2 (especially in case of IBD estimation) and a maximum block size of 5 (may be increased further for high density maps but sampling will require more computational time). Five was appropriate for the high level of linkage disequilibrium in our germplasm and the density of our map.			
skipSampleMarkers	Indicator variable to skip sampling of marker information in Markov chain	5000	
Indicator variable that defines how frequently the linkage phase of markers is re-estimated. With dense, high quality SNP data, this value could be 5% of the chain length for a QTL analysis. We were conservative, using a value equal to 2.5% of the chain length. When the aim is to accurately define SNP haplotypes rather than perform a QTL analyses, this variable should be zero.			
distanceQTLs	Minimal distance between any pair of two QTLs	0.0	Did not restrict distance
We preferred 0.0 because we wanted to explore all possibilities. Higher values might not have allowed us to detect QTLs in close proximity to one another.			
MSegDelta	Neutralizer for extreme segregation distortion for marker loci during the marker consistency check preceding the actual QTL analysis	1	0 = Extreme segregation distortion is not accepted 1 = Extreme distortion is accepted
This indicator variable defines whether extreme segregation distortion (e.g., lack of a genotype class in a segregating progeny) should be rejected or accepted. Such a distortion could be due to genotyping errors or have a biological basis, such as gametic incompatibility or selection against disease susceptibility. When distortion is not accepted, FlexQTL™ will adjust the parental genotype calls according to the genotype calls of offspring.			
DeleteDR	Indicator variable for deleting ALL observed singletons	1	0 = All marker calls that cause singletons are maintained 1 = Calls are deleted
Double recombination around a single marker should be extremely rare with high-density SNP data. When present, these are mostly due to calling errors or marker ordering rather than to true double recombination. We chose to delete marker calls causing double recombinants, replacing the call with the symbol for missing values.			
Seed	A random integer value to start the simulation process and to also define the number of QTL to be included in the initial model of the MCMC simulation	variable	variable
Large integer value that is used in random number generating algorithms of the MCMC simulation process. With FlexQTL™ the last digit of the seed determines the number of QTL in the initial model of the Markov chain: with 0–4 the initial model has no QTL and with 5–9 the initial model has the maximum number indicated by “maximQTL”. Replicated runs with different seeds were performed to examine whether results were robust across seeds or whether outcomes varied considerably. Although this number is supposed to be randomly chosen, it is common practice to edit the last digit such that both possible starts are used.			