Heredity (2013) 111, 106–113

www.nature.com/hdy

ORIGINAL ARTICLE

Lineage-specific mapping of quantitative trait loci

C Chen[1] and K Ritland

We present an approach for quantitative trait locus (QTL) mapping, termed as 'lineage-specific QTL mapping', for inferring allelic changes of QTL evolution along with branches in a phylogeny. We describe and analyze the simplest case: by adding a third taxon into the normal procedure of QTL mapping between pairs of taxa, such inferences can be made along lineages to a presumed common ancestor. Although comparisons of QTL maps among species can identify homology of QTLs by apparent co-location, lineage-specific mapping of QTL can classify homology into (1) orthology (shared origin of QTL) versus (2) paralogy (independent origin of QTL within resolution of map distance). In this light, we present a graphical method that identifies six modes of QTL evolution in a three taxon comparison. We then apply our model to map lineage-specific QTLs for inbreeding among three taxa of yellow monkey-flower: *Mimulus guttatus* and two inbreeders *M. platycalyx* and *M. micranthus*, but critically assuming outcrossing was the ancestral state. The two most common modes of homology across traits were orthologous (shared ancestry of mutation for QTL alleles). The outbreeder *M. guttatus* had the fewest lineage-specific QTL, in accordance with the presumed ancestry of outbreeding. Extensions of lineage-specific QTL mapping to other types of data and crosses, and to inference of ancestral QTL state, are discussed.
Heredity (2013) **111,** 106–113; doi:10.1038/hdy.2013.24; published online 24 April 2013

Keywords: Quantitative trait locus mapping; evolution of mating systems; *Mimulus*

INTRODUCTION

Knowledge of the genetic architecture of complex traits can offer insight into several facets of evolution and adaptation, including adaptive differentiation, signatures of selection, speciation genetics, epistasis, mating system evolution and sexual selection (Erickson *et al.*, 2004). In both natural and domesticated populations, this is classically obtained by inferring quantitative trait loci (QTLs) in segregating progenies of a cross between two differentiated populations (Slate, 2005). Although the biases and inaccuracies of QTL mapping are well known, QTL analysis can provide fundamental information about the size, location and effects of individual QTL that differ between the two parents of the cross (Broman, 2001; Price, 2006).

QTL mapping techniques have been used to dissect the genetic architecture of complex traits in many applications, such as for human diseases (Cardon and Bell, 2001), adaptation in natural populations (Slate, 2005) and breeding of animals (Mott *et al.*, 2000). In plants, a classic example of QTL mapping for adaptive traits has involved the comparison between bumblebee pollinated *M. lewsii* and hummingbird pollinated *M. cardinalis* (Bradshaw *et al.*, 1995; Schemske and Bradshaw, 1999). However, all these studies are based upon pairwise comparisons, which allow estimates of those differences only along a single lineage.

In this paper, we propose a phylogenetic approach for QTL mapping, in which the genetic effect of QTL along phylogenetic lineages is inferred, using several pairwise crosses among the taxa examined. At the simplest, by bringing in a third taxon, one can infer the QTL changes that have occurred along the two lineages that lead to the two most closely related taxa. The third lineage traces from the common ancestor of these two taxa, back to the common ancestor of all three taxa and forwards again to the third taxon. After mapping QTL onto lineages, we can determine if changes of QTL effects at the same apparent map position are orthologous (arising singly in an ancestral lineage leading to derived taxa) or paralogous (arising twice independently in derived lineages, due to different mutations). Also, akin to Orr's test (Orr, 1998), the distribution of QTL in a species network can test for drift-mutation balance versus directional selection for quantitative traits.

To illustrate our approach, we use three species from the yellow monkey-flower species complex (Vickery, 1978). In the yellow monkey-flower (*M. guttatus*) species complex, several daughter species with varying degrees of inbreeding and narrow distributions occur, derived from the outcrossing and widespread *M. guttatus* (Vickery, 1978). We expect QTL changes in the two inbreeding lineages should mostly be promoting inbreeding, and that in the outbreeding lineage should have fewer or no QTL changes, as outbreeding is ancestral. In exploring this method, we find several challenges that should be addressed in future studies.

MATERIALS AND METHODS

Model species and traits

M. guttatus is extensively distributed in western North America, with a relatively low inbreeding coefficient of 0.38 (Ritland and Ritland, 1989). *M. platycalyx* occurs in the coast ranges north of San Francisco (Dole, 1992) and is a moderately inbred species with an inbreeding coefficient of 0.54 (Ritland and Ritland, 1989). *M. micranthus* is endemic to the Coast Range foothills of California, and is a highly selfing species with an inbreeding coefficient of 0.73 (Ritland and Ritland, 1989). The shape and size of the

Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, British Columbia, Canada
Correspondence: Professor K Ritland, Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, British Columbia, Canada V6T1Z4.
E-mail: kermit.ritland@ubc.ca
[1]Current address: NBB-B27, International Maize and Wheat Improvement Center (CIMMYT) Km. 45, Carretera Mexico-Veracruz, Texcoco, Mexico.
Received 18 April 2012; revised 6 March 2013; accepted 8 March 2013; published online 24 April 2013

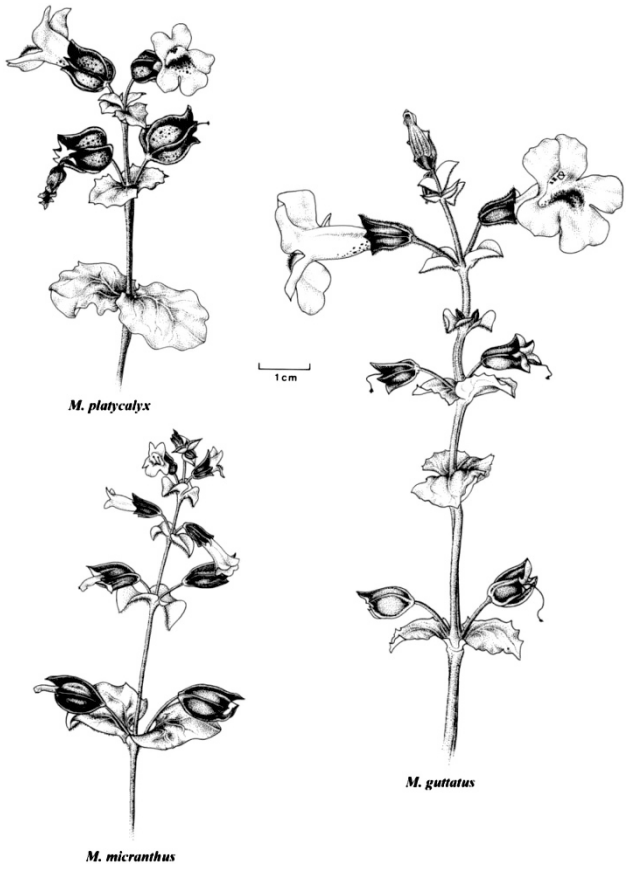

Figure 1 The three intercrossable *Mimulus* taxa used in this study. *M. guttatus* is predominately outcrossing, whereas *M. platycalyx* and *M. micranthus* are inbreeders. Although *M. platycalyx* has a relatively large flower, it is highly autofertile.

flowers of these species is illustrated by Figure 1. The evolution of inbreeding is accompanied by changes of an entire syndrome of floral traits, including male allocation, reduced size of floral characters, reduced attraction to pollinators, and reduction of the spatial and temporal separation of male and female reproductive organs within the flower (Jain, 1976; Ritland and Ritland, 1989). Figure 2 gives the five metric floral characters we measured as representative of inbreeding for *Mimulus*.

Lineage-specific QTL mapping for three taxa: models

QTL mapping has traditionally involved a single cross between differentiated taxa. In a two taxa QTL map, an inferred QTL has a lineage-specific effect 'U_1' relative to the common ancestor, and another lineage-specific effect 'U_2' relative to the common ancestor. The lineage-specific values are U_1 and U_2; only the difference $U_1 - U_2$ can be estimated.

Between three taxa, there are three QTL effects. With regard to the three *Mimulus* species, these are denoted U_G for the *M. guttatus* lineage, U_P for the *M. platycalyx* lineage and U_M for the *M. micranthus* lineage. As in the single-cross data, the three possible pairwise crosses yield estimates of $(U_G - U_P)$, $(U_G - U_M)$ and $(U_P - U_M)$. The solution for the individual branch effects U_G, U_P and U_M involves the insolvable relation $\begin{bmatrix} U_G - U_P \\ U_G - U_M \\ U_P - U_M \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} U_G \\ U_P \\ U_M \end{bmatrix}$, as the matrix is singular. Individual branch effects are not estimable without some kind of constraint. If we introduce the constraint $U_G + U_P + U_M = 0$, for example, the sum of the QTL effects across lineages is zero, then there are effectively two unknowns instead of three. This introduced constraint makes the lineage-specific estimates relative, in that changes are relative to an arbitrary mean of zero. The resulting relation $\begin{bmatrix} U_G - U_P \\ U_G - U_M \\ U_P - U_M \end{bmatrix} = \begin{bmatrix} 2 & 0 & 1 \\ 2 & 1 & 0 \\ 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} U_G \\ U_P \\ U_M \end{bmatrix}$ can now be solved for lineage-specific effects U_G, U_P and U_M, as the matrix is non-singular (determinant = 3).

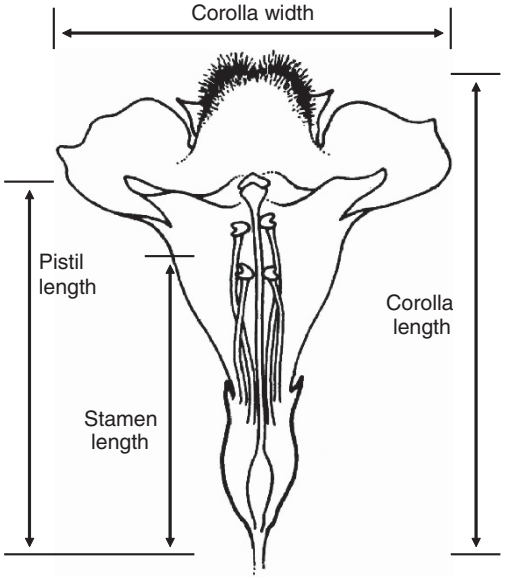

Figure 2 The four quantitative traits measured in this study; the fifth, stigma-anther separation, is the difference between pistil length and stamen length.

An alternative approach is likelihood, as commonly used for phylogenies of DNA sequences. As in the three-sequence model, the likelihood would be $L = \Pr(U_0)\Pr(U_1|U_0)\ \Pr(U_2|U_0)\ \Pr(U_3|U_0)$, where $\Pr(U_0)$ is the probability of ancestor state U_0 and $\Pr(U_i|U_0)$ are the probabilities of mean U_i given ancestral state U_0, for $i = 1, 2, 3$ (Ritland and Clegg, 1987). For DNA sequences, there are base substitution models that provide explicit Pr() functions; for quantitative traits, Pr() would involve a Gaussian function. However, the major problem is that informative DNA phylogenies require data across many sites, while a QTL is equivalent to one nucleotide site (for example, one observation with one degree of freedom). Informative likelihood lineage-specific QTL mapping would thus require assuming QTL sites all evolve identically, preventing inferences about locus-specific patterns of evolution.

Lineage-specific QTL mapping for three taxa: parameter space

For this three taxon comparison, Figure 3 depicts the 'space' of QTL effects. If we constrain the joint values of U_G, U_P and U_M to sum to zero, we can plot it in a manner similar to a ternary plot (de Finetti diagram), except that each axis ranges from -1 to $+1$ (QTL effects conceivably might lie outside this interval, but in our study we normalized traits by dividing by the variance, and indeed our results found no QTL of effect larger than |0.3|). From a given point within the triangle, the intersection of a line perpendicular to an axis gives the QTL effect observed for that axis. Figure 3a shows how a given triplet of estimates plots in this space; in this example, the *M. guttatus* lineage has a QTL of $+0.23$, the *M. platycalyx* lineage has a QTL of $+0.08$ and *M. micranthus* has a QTL of -0.31.

Figure 3b shows six fundamental divisions of this space, and for each, the most parsimonious phylogenies of QTL evolution, when the ancestral QTL effect is positive. Ancestry is most likely positive as a large flowered outbreeding species is most likely ancestral (Vickery, 1978). These partitions are indicated by (a), (b), (c), (d), (e) and (f), corresponding joint QTL effects of $++-$, $+--$, $+-+$, $--+$, $-++$ and $-+-$ for the *M. micranthus, M. guttatus* and *M. platycalyx* lineages, respectively. Within each of the six partitions is a three-taxon phylogeny. The slashes indicate the

branch that most parsimoniously explains the lineage-specific QTL change observed among the current taxa.

Homology is the common ancestry of a trait or gene, as opposed to functional similarity. It is mainly applied to gene sequences and gene products (Fitch, 2000; Petsko, 2001). Homology of QTL has been classically defined as QTL that map to the same position in different populations or species (for example, see Freeman *et al.*, 2009). In lineage-specific QTL mapping, we can subdivided homology into 'orthology of QTL', as QTLs that share common ancestry (share the same mutational event of an ancestor), and 'paralogy of QTL', as QTLs that have different mutations in the same region of the genetic map, but which arise independently. Note that paralogous QTL are not true homologues but 'apparent' homologues in the classical QTL definition as mapping to the same position. In Figure 3b, for positive ancestral QTL, cases (b) and (f) illustrate paralogous QTL, whereas the other four cases illustrate orthologous QTL. We expect orthologous QTL to be more common than paralogous QTL, as two changes are needed for paralogy but just one for orthology. If the ancestral state is negative (for example, inbreeding state), the space of paralogy is rotated 180°, with cases (c) and (e) indicating paralogy.

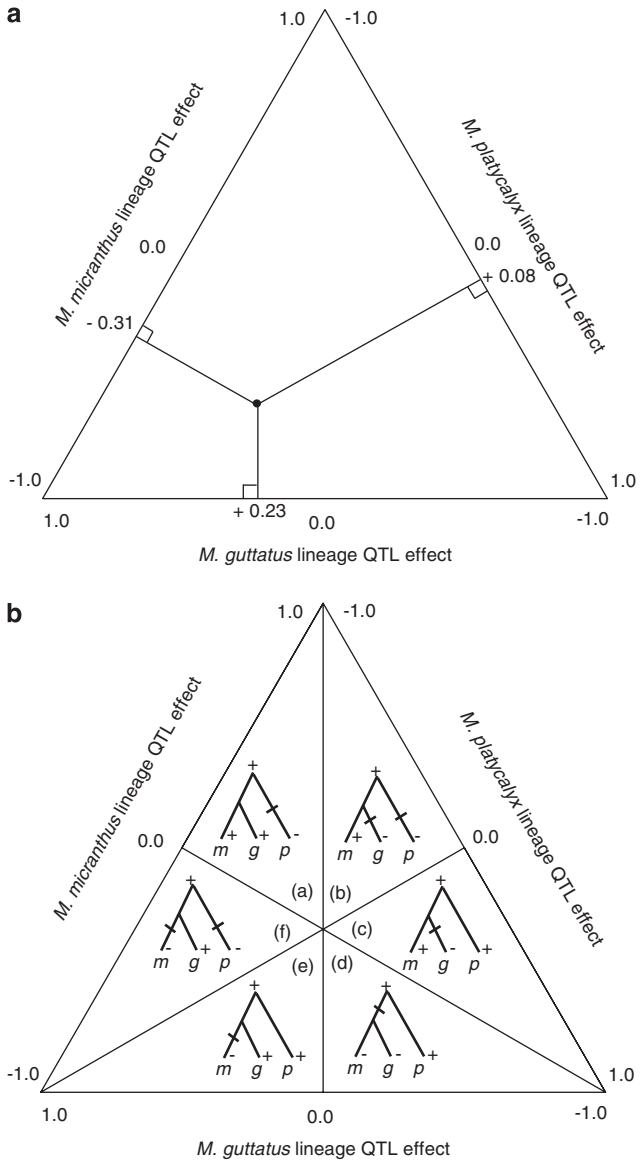

Figure 3 The space of lineage-specific QTL effects in a three taxa comparison. (**a**) Lines emanating from a given point, representing a triplet of estimates, perpendicularly intercept axes at their respective lineage-specific QTL estimates (see text). (**b**) There are six fundamental divisions of this space, and for each, the most parsimonious phylogenies of QTL evolution is shown (hashes denote changes of QTL; −denotes a negative QTL effect, +denotes a positive QTL effect; *m* = *M. micranthus*, *g* = *M. guttatus*, *p* = *M. platycalyx*).

Lineage-specific QTL mapping for three taxa: estimation

To statistically identify QTL from genetic data in segregating crosses, we initially assume that species are fixed for QTL, but that QTL effects differ between species (this is a common assumption for species comparisons). Hence, the F_1 can be heterozygous for a QTL, whereas the parent species is homozygous. Under this assumption, the expected means of progeny, conditioned upon the marker genotype, are given in Table 1 where the F_1 derived from taxon *i* and *j* is backcrossed to taxon *i*. In this table, there are two crossing configurations, depending upon whether markers are in a backcross configuration ($Aa \times aa$) or an intercross configuration ($Aa \times \underline{A}a$); this includes coupling versus repulsion linkage phases. The third column gives the expected means of progeny genotypes, conditioned upon parent genotypes and linkage phase.

The joint likelihood of quantitative traits and markers is

$$L = \prod_{\text{crosses } c} \max_{\text{phases } p(c)} \left(\prod_{\text{progeny } k} \exp\left(-\left(Q_{ck} - E_{t(c),p(c),g(c,k)}(i,j)\right)^2 \right) \right) \qquad (1)$$

where Q_{ck} is the observed quantitative trait value (normalized to zero mean and unit variance), and $E_{t(c),p(c),g(c,k)}(i,j)$ is the expected mean (given in rightmost column of Table 1) of a progeny of cross *c* involving an F_1 between taxa *i* and *j*, backcrossed to taxa *i*, with parent cross type *t(c)* at the marker locus (1 = backcross, 2 = intercross), linkage phase *p*(*c*) (1 = coupling, 2 = repulsion) and progeny genotype *g*(*c*,*k*) (1 = dominant, 2 = recessive). This assumes a normal distribution of quantitative traits. The most likely phase is chosen ('max'). If a given F_1 is a parent of more than one family, then the expression is more complicated; the most likely phase is chosen over all involved crosses. In principle, just six crosses, three to generate F_1's between all three pairs of taxa, and three backcrosses to just one of the two parents, provide enough information to jointly estimate u_G, u_P and $u_{M.}$ (the *u*'s in Table 1, across the tree lineages). Below we used a more intensive crossing design that involved all six possible backcrosses.

Mimulus crosses and genotyping

All three inter-taxon crosses were performed (*M. guttatus* × *M. platycalyx*, *M. guttatus* × *M. micranthus* and *M. platycalyx* × *M. micranthus*). The F_1 progeny were then backcrossed to each parent species (six reciprocal backcrosses). For the purpose of data analysis, this design can be written as a loop of six crosses: PP × PG × GG × GM × MM × MP × PP (the last 'PP' is the same as the first),

Table 1 Expected mean phenotype for a quantitative trait, conditioned upon progeny genotype and crossing configuration of parents

Crossing configuration	*Progeny*	*Expected mean*	*Symbol*
$Aa \times aa$			
$(Q_1A)/(Q_2a) \times (Q_1a)/(Q_1a)$	*Aa*	u_i	$E_{1,1,1}$
	aa	$(u_i + u_j)/2$	$E_{1,1,2}$
$(Q_1a)/(Q_2A) \times (Q_1\ a)/(Q_1\ a)$	*Aa*	$(u_i + u_j)/2$	$E_{1,2,1}$
	aa	u_i	$E_{1,2,2}$
$Aa \times Aa$			
$(Q_1A)/(Q_2a) \times (Q_1A)/(Q_1a)$	*A_*	$(5u_i + u_j)/6$	$E_{2,1,1}$
	aa	$(u_i + u_j)/2$	$E_{2,1,2}$
$(Q_1a)/(Q_2A) \times (Q_1A)/(Q_1a)$	*A_*	$(2u_i + u_j)/3$	$E_{2,2,1}$
	aa	u_i	$E_{2,2,2}$

First parent is F_1 between taxa *i* and *j*; second parent is pure taxa *i*. *A_* denotes *AA* or *Aa* (dominant, banded phenotype).

where PP, GG and MM are the parent species, and PG, GM and MP are the F_1s. Note that to map a QTL, at least two of the three F_1 (PG, GM, MP) must be heterozygous for a marker (this is where more polymorphic markers, such as microsatellites, provide more power for lineage-specific mapping). One individual plant was used for each species, to ensure that there is just one possible parental genotypic configuration in these crosses. These species are quite distinct such that one individual captures most QTL differences between species.

All plants, including parental, F_1 and backcrosses, were grown in Pro-Mix soil in growth chambers at 14 °C/8 °C day/night and 18 h of daylight. To avoid effects caused by the heterogeneous environment among growth chambers, trays of seedlings were periodically rotated among growth chambers. To avoid pollen contamination, flowers were bagged immediately after crossing. Five floral traits were measured (Figure 3): corolla width, corolla length, pistil length, stamen length (there are two sets of anthers that differ in length, the average length was measured) and stigma-anther separation (the difference between the previous two traits). In a combined-cross analysis, crosses with greater variability in the phenotype will have a greater influence (Li *et al.*, 2005). Hence, before QTL analysis, we standardized traits by division by standard deviations across all crosses; this was done regardless of cross as such variance stabilizing within crosses can introduce bias into our lineage-specific analysis.

Genomic DNA was isolated from leaf tissue with the CTAB method (Doyle and Doyle, 1990). Assays for amplified fragment length polymorphisms (AFLP) were performed following Vos *et al.* (1995) and Remington *et al.* (1999) with modifications for the LiCor 4200 DNA sequencer (LI-COR, Inc., Lincoln, NE, USA). Five hundred nanogram of DNA was digested with *Eco*RI and *Mse*I. Preamplification was carried out using standard AFLP *Eco*RI and *Mse*I primers containing the selective nucleotides *Eco* + C and *Mse* + CC. Selective final amplifications were conducted by eight combinations of *Eco* primers with three nucleotides and *Mse* primers with three nucleotides (Table 2). After running the LiCor gels, RFLPscan Version 3.0 (Scanalytics Inc., Fairfax, VA, USA) scored segregating loci. Markers showing significant segregation distortion ($P<0.05$) were excluded. A total of 368 polymorphic loci were obtained, and 675 individuals from the six backcrosses were genotyped (PP × PG = 112, PG × GG = 135, GG × GM = 124, GM × MM = 121, MM × MP = 135 and MP × PP = 68).

Inference of parent genotypes and linkage

As the dominance of AFLP markers prevents detection of heterozygous parents, we inferred parent genotypes using the progeny of all six backcrosses, as follows. Table 3 lists the segregation probabilities for bandless versus banded progeny, conditioned on parent genotypes. Let these probabilities be $P_u(i,j)$ for the bandless phenotype and $P_b(i,j)$ for the banded phenotype. Across all six parents, there $3^6 = 729$ possible configurations of parent genotypes. For any particular genotypic configuration, the likelihood of the observed data across the six crosses is

$$\prod_{k=1}^{6} p_u(k_{1k}, k_{2k})^{N_{u,k}} p_b(k_{1k}, k_{2k})^{N_{b,k}} \qquad (2)$$

where $N_{u,k}$ is the number of bandless progeny and $N_{b,k}$ is the number of banded progeny in cross k; and $k_{1k}$and k_{2k} are the putative male and female parent genotypes of cross k. A computer program was written that enumerated all $3^6 = 729$ possible parental genotype configurations (values of $k_{1k}$and k_{2k}) and chose the most likely configuration of the six parents for each AFLP locus. The likelihood of the second most likely parental genotypes for each locus is also given in the program, which incorporates uncertainty of inferred parental genotypes. To account for genotyping error, at each cross, the less frequent phenotype was truncated to zero and the likelihood of the data evaluated against the original data. If the increase in likelihood was greater than expected by a 5% genotyping error, the less frequent phenotype was truncated to zero. Parentage inference was quite reliable via Equation (2), as relative to the most likely set of six parent genotypes, we found the next most likely set of six-parent genotypes were 1000 times less likely 90% of the time, and 10 times less likely 97% of the time.

Using the most likely parental genotypes, Equation (1) was applied to each of the 368 markers. To avoid problems with numerical estimation, we simply evaluated the likelihood surface across all possible values of U_G, U_P and U_M, each ranging from -1 to $+1$ in increments of 0.02. The joint estimate was chosen as that combination of the three *U's* that gave the highest likelihood. Variance explained was calculated as $(V_0 - V_E)/V_0$ where $V_E = \sum_c \sum_k (Q_{ck} - E_{ck})^2$ and $V_0 = \sum_c \sum_k Q_{ck}^2$. This formula assumes that the magnitude of QTL effect between any two species is the sum of the lineage-specific QTLs, for example, QTLs evolve additively. It also assumes that QTL effects are fixed between taxa, and none are segregating within taxa.

Statistical significance of QTL was ascertained in two ways. First, we permuted quantitative traits and markers 1000 times (traits randomized among genotypes). The likelihood of these permutated data were compared with the original un-permuted data; if 50 or less of the permuted data were more likely than the original data, the estimates are deemed significant. In the second way, we use the bootstrap to estimate standard errors of individual branches, wherein progeny were re-sampled with replacement within crosses 1000 times.

Also, we estimated recombination rates using a joint likelihood function (Hu *et al.*, 2004), which combines information across the six crosses. For those pairs of markers, which were linked closer than 10 map units, one marker was omitted. This and the above calculations were implemented in a FORTRAN computer program written for this data (code available upon request from KR).

Table 2 AFLP primers and numbers of polymorphic fragments (*N*)

Pre-amplification step	*Name*	*Final amplification step*	N
*Eco*RI + AC/*Mse*I + CC	B2	*Eco* + ACA/*Mse* + CCT	25
*Eco*RI + AC/*Mse*I + CC	B3	*Eco* + ACA/*Mse* + CGC	54
*Eco*RI + AC/*Mse*I + CC	B4	*Eco* + ACA/*Mse* + CCA	59
*Eco*RI + AC/*Mse*I + CC	B5	*Eco* + ACT/*Mse* + + CGG	38
*Eco*RI + C/*Mse*I + CC	C1	*Eco* + CA//*Mse* + CCG	61
*Eco*RI + C/*Mse*I + CC	C3	*Eco* + CCG/*Mse* + CCG	31
*Eco*RI + C/*Mse*I + CC	C4	*Eco* + CCA/*Mse* + CCT	54
*Eco*RI + C/*Mse*I + CC	C7	*Eco* + CTC/*Mse* + CCT	46

Abbreviation: AFLP, amplified fragment length polymorphism.
Primer sequence of *Eco*RI is 5′-GACTGCGTACCAATTC-3′; primer sequence of *Mse*I is 5′-GATGAGTCCTGAGTAA-3′.

Table 3 Probabilities of bandless (first number) versus banded progeny (second number), conditioned on the genotypes of the two parents of a cross (bandless is the recessive condition)

	Parent 2		
Parent 1	*AA*	*Aa*	*aa*
AA	(0, 1)	(0, 1)	(0, 1)
Aa	(0, 1)	($\frac{1}{4}$, $\frac{3}{4}$)	(½, ½)
aa	(0, 1)	(½, ½)	(1, 0)

RESULTS

Among the eight primer pairs (Table 2), 368 polymorphic AFLP loci were genotyped and scored (markers with distorted segregation ratios were excluded). Of these, 190 were informative about lineage-specific QTL (where all three pairwise crosses were heterozygous F_1s). Nine markers were excluded from further analysis due to close linkage to another marker. The estimated genetic distances between *M. guttatus* and *M. micranthus* was 0.08 (s.e. = 0.01), between *M. guttatus* and *M. platycalyx* was 0.19 (s.e. = 0.02) and between *M. platycalyx* and *M. micranthus* was 0.20 (s.e. = 0.02). A dendrogram based upon these distances is shown in Figure 4, with the 'A'

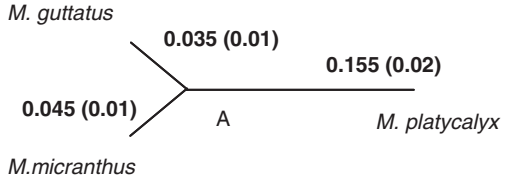

Figure 4 Dendrogram of genetic distances among *M. guttatus*, *M. platycalyx* and *M. micranthus* based upon 368 AFLP loci (standard errors in parenthesis); the true location of the most common in on the branch leading out to *M. platycalyx*.

indicating the approximate location of the common ancestor of these three taxa (roughly the point within the tree that is equidistant to all three taxa). This dendrogram shows that *M. micranthus* is the more recently derived inbreeder. Although *M. platycalyx* has the largest branch length, it is likely that *M. guttatus* was the common ancestor. A band-sharing index (Nei and Li, 1979) gave the same result as this dendrogram.

Lineage-specific QTL effects

Table 4 lists the markers that gave significant lineage-specific QTL effects for at least one of three lineages. Among the 190 informative markers, 40 QTLs were detected across the five traits, for an average of eight QTLs per trait. The same markers often had QTLs for several of these five traits, indicating high pleiotropy. For example, for corolla width and length in *M. guttatus*, QTL b2_109 is pleiotropic, generally controlling floral size. Asterisks denote estimates that are either above zero, or below zero, in 95% or more of the bootstraps. Note that estimates deemed significant by this measure may not have permutation probabilities of less than 5%. This is due to the complex structure of data, wherein one lineage may not have significant QTL but another lineage will. In addition, the percentage variance explained by markers was very low, on the order of 1%, despite the estimates showing statistical significance. This is extremely low for classical pairwise QTL experiments.

QTL of positive effect increases the size of the trait (promoting outcrossing), whereas negative QTL effect decreases trait size (promoting selfing). We expect the *M. guttatus* lineage to show positive QTL effects, and the *M. platycalyx* and *M. micranthus* lineages to show negative effects. By and large this was true; for corolla width 9 of 12 estimates were positive for the *M. guttatus* lineage, but only one significantly so. The *M. platycalyx* also had largely positive corolla width/length QTL effect, except some were significantly negative. *M. micranthus* had the most negative and most significantly negative QTL effects for corolla width and length, in accord with its smallest flower size. Stamen length showed the least number of QTL ($n=4$) followed by stigma-anther separation ($n=6$). *M. platycalyx* gave the greatest number of significant QTLs ($n=21$), followed by *M. micranthus* ($n=12$) then *M. guttatus* ($n=5$), which accords with the greater length of the *M. platycalyx* lineage (Figure 4).

To obtain a broader picture of the pattern of lineage-specific QTLs, Figures 5 and 6 show the 'space' of QTL effects. To increase the sample size at the expense of adding some false positives, we included all markers with QTL permutation probabilities of 0.10 or less (thus the numbers of markers is greater than in Table 4), with those of probability of 0.05–0.10 indicated by the smaller dots.

Comparing Figure 5 against the hypothetical evolutionary scenarios given in Figure 3b, corolla width showed predominately scenarios (a), (d) and (e). Scenario (a) and (e) corresponds to QTL change in an

Table 4 Estimates of lineage-specific QTL genetic effects (standard errors in parenthesis)

AFLP marker	*Lineage-specific QTL effect (s.e.)*			V	P
	M. guttatus	M. platycalyx	M. micranthus		
Corolla width					
b2_109	0.16* (0.09)	−0.18* (0.10)	0.02 (0.08)	0.96	0.002
b3_444	0.12 (0.10)	−0.20* (0.10)	0.08 (0.08)	1.06	0.097
b3_343	0.06 (0.07)	−0.18* (0.08)	0.12 (0.10)	0.95	0.086
c3_366	0.02 (0.09)	0.16 (0.10)	−0.18* (0.09)	1.07	0.055
c3_328	0.04 (0.12)	0.14 (0.10)	−0.18* (0.10)	0.71	0.060
c7_480	0.04 (0.07)	0.12* (0.08)	−0.16* (0.07)	0.82	0.055
c7_461	0.04 (0.07)	0.08 (0.08)	−0.12* (0.07)	0.35	0.091
c7_390	−0.04 (0.09)	0.16* (0.11)	−0.12 (0.10)	0.60	0.076
c7_385	0.06 (0.10)	0.08 (0.11)	−0.14* (0.08)	0.67	0.006
c7_323	−0.12 (0.10)	0.24* (0.12)	−0.12 (0.10)	0.99	0.054
c7_248	0.02 (0.08)	0.14* (0.08)	−0.16* (0.08)	0.78	0.055
c7_171	−0.08 (0.10)	0.16* (0.12)	−0.08 (0.09)	0.50	0.083
Corolla length					
b2_109	0.18* (0.10)	−0.22* (0.09)	0.04 (0.11)	1.51	0.001
c3_366	0.02 (0.08)	0.20* (0.08)	−0.22* (0.07)	1.40	0.050
c3_328	0.04 (0.12)	0.14 (0.11)	−0.18* (0.10)	0.83	0.059
c7_480	0.00 (0.07)	0.14* (0.08)	−0.14* (0.06)	0.79	0.039
c7_385	0.08 (0.09)	0.08 (0.09)	−0.16* (0.09)	0.87	0.011
c7_323	−0.16 (0.10)	0.24* (0.11)	−0.08 (0.09)	1.19	0.021
c7_248	0.02 (0.08)	0.14* (0.08)	−0.16* (0.07)	0.65	0.040
c7_224	0.04 (0.08)	0.08 (0.08)	−0.12* (0.08)	0.49	0.074
Pistil length					
b2_249	0.10 (0.08)	−0.16* (0.09)	0.06 (0.08)	0.75	0.013
b2_109	0.16 (0.09)	−0.20* (0.08)	0.04 (0.08)	1.17	0.003
c3_366	0.06 (0.12)	−0.18* (0.10)	0.12 (0.09)	1.02	0.035
c3_328	0.04 (0.11)	−0.16* (0.08)	0.12 (0.09)	0.89	0.037
c3_168	−0.16 (0.11)	0.42* (0.13)	−0.26* (0.09)	2.31	0.034
c3_080	−0.02 (0.07)	−0.14 (0.09)	0.16* (0.08)	0.62	0.088
c7_480	0.02 (0.08)	0.08 (0.09)	−0.10* (0.07)	0.42	0.027
c7_390	−0.10 (0.11)	0.18* (0.11)	−0.08 (0.09)	0.68	0.018
c7_385	0.10 (0.09)	0.00 (0.09)	−0.10* (0.08)	0.52	0.012
c7_323	−0.14 (0.11)*	0.20* (0.11)	−0.06 (0.11)	0.85	0.005
Stamen length					
b2_249	0.06 (0.08)	−0.18* (0.10)	0.12* (0.09)	0.95	0.009
b2_109	0.18* (0.09)	−0.20* (0.10)	0.02 (0.09)	1.38	0.004
c7_390	−0.12 (0.09)	0.20* (0.11)	−0.08 (0.09)	0.78	0.087
c7_323	−0.16* (0.09)	0.24* (0.10)	−0.08 (0.09)	1.19	0.053
Stigma-anther separation					
b2_249	0.08 (0.07)	−0.22* (0.10)	0.14* (0.09)	1.32	0.073
b2_239	0.26* (0.10)	−0.28* (0.10)	0.02 (0.09)	1.96	0.025
b5_052	0.10* (0.08)	−0.20* (0.08)	0.10 (0.07)	1.06	0.095
c3_372	0.08 (0.09)	−0.20* (0.09)	0.12 (0.08)	1.08	0.033
c3_366	0.14 (0.11)	−0.16* (0.09)	0.02 (0.08)	0.73	0.046
c3_283	0.12 (0.08)	−0.16* (0.10)	0.04 (0.09)	0.79	0.053

Abbreviations: AFLP, amplified fragment length polymorphism; QTL, quantitative trait locus.
V is percentage variance explained by the markers in the set of crosses, *P* is the permutation probability.
Estimates that are significantly different from zero (95% level) as determined by bootstraps are indicated by asterisks.

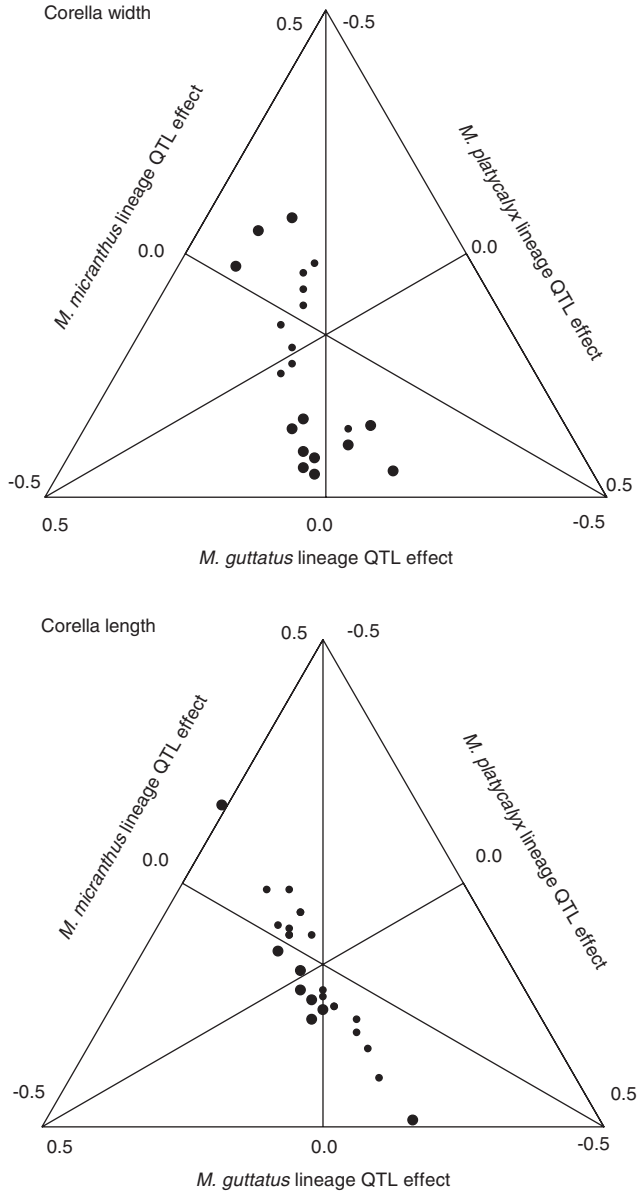

Figure 5 The space of lineage-specific QTL evolution among three monkeyflower species, for the floral size measures of corella width and correla length (larger dots are QTLs with less than 5% permutation probabilities and smaller dots are QTLs with 5–10% permutation probabilities).

inbreeder lineage. A notable exception are four QTL for corolla width in scenario (d), in which *M. platycalyx* lineage has a QTL effect for larger corollas. Also these three modes are all orthologous. Corolla length showed a similar pattern with the exception that two QTL showed significant paralogy (scenario f). Figure 6 shows the scenarios for floral organs. All six quadrants were occupied by pistil length QTL with roughly equal numbers of QTL in each quadrant, except for scenario (c). However, the more significant QTL mapped mainly to (a), (d) and (e), as in the floral size traits. Stamen length also mapped to these three quadrants. Finally, stigma-anther separation exclusively occupied quadrant (a), indicating that reduced stigma-anther separation is unique to the derived *M. platycalyx* lineage. Overall, very few QTL mapped to the paralogous modes (b, f).

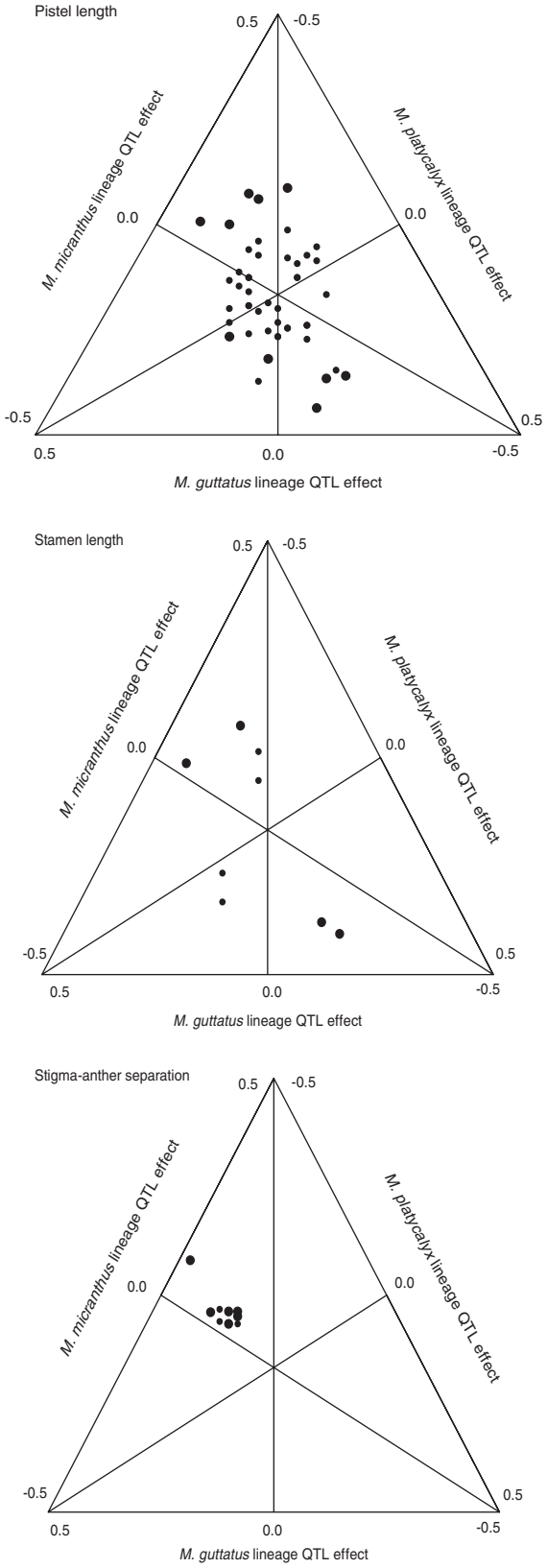

Figure 6 The space of lineage-specific QTL evolution among three monkeyflower species, for the floral outcrossing measures of anther length, stamen length and stigma-anther separation (larger dots are QTLs with less than 5% permutation probabilities and smaller dots are QTLs with 5–10% permutation probabilities).

DISCUSSION

Patterns of lineage-specific QTL effects

'Lineage-specific QTL mapping' goes beyond the traditional inferences about number of QTLs and magnitudes of QTL effects. With pairwise crosses between several related taxa, one can now infer phylogenetic lineages along which QTL changes occur, but with some difficulties. These are discussed below in the prospects.

Assuming outbreeding being ancestral, the *M. guttatus* lineage showed the fewest (6) significant QTL across the five traits, as compared with 29 for *M. platycalyx* lineage and 17 for the *M. micranthus* lineage. The evolutionary distance of these two lineages were strongly associated with QTL number: that for *M. platycalyx* is considerably longer than that for *M. micranthus* (Figure 4). The average QTL effect was also greater in the *M. platycalyx* lineage (0.197) than in either the *M. micranthus* lineage (0.157) or the *M. guttatus* lineage (0.173).

Although we expected that the two inbreeding lineages would have QTL effects that cause smaller flowers and less stigma-anther separation, not all QTLs involving with the evolution of inbreeding were of such effect. Although almost all significant QTLs for the *M. micranthus* lineage were negative (the exception being a positive QTL for stigma-anther separation), the *M. platycalyx* lineage had a mixture of approximate equal proportions of positive and negative QTL effects for all traits except stigma-anther separation. These accords with the flower size of *M. platycalyx*, which is not much smaller than *M. guttatus*. It also may indicate a complex evolutionary history in the lineage of *M. platycalyx*, where QTL alleles for smaller flowers and QTL allele for larger flowers were either selectively favored at different times, or underwent considerable genetic drift. Interestingly, all QTL effects for stigma-anther separation were negative for the *M. platycalyx* lineage. This accords with the observation that *M. platycalyx* is highly autofertile in the greenhouse, and that the major feature distinguishing the evolution of this species is the fixation of several QTL for reduced stigma-anther separation.

Mapping QTL to the 'space' of QTL effects (Figure 3) allows one to distinguish alternative phylogenetic patterns of QTL evolution, including distinguishing orthology from paralogy, when QTL are deemed 'homologous'. The most parsimonious explanations of where QTL changes occur along with lineages are shown in Figure 3b, for the cases in positive ancestral QTL. When we compare this figure to the observed QTL space in *Mimulus* (Figures 5 and 6), we see some dramatic differences between traits for their pattern of evolution. Corolla width showed predominately scenario (a) and (e) in Figure 3b; these both correspond to QTL changes for smaller corollas appearing in an inbreeder lineages, as would be expected. A notable exception is four QTL for corolla width in scenario (d), in which *M. platycalyx* lineage has a QTL effect for larger corollas. Corolla length showed predominant scenario (a) in Figure 3b (smaller in the *M. platycalyx* lineage) but also some (e) and (d) scenarios, involving changes toward shorter corollas and paralogy in the *M. guttatus* lineage, respectively. However, (f) is a rare case, as it requires two mutational steps. Pistil length showed almost all six alternative modes of evolution in roughly equal proportion, with *M. guttatus* lineage showing a rare case of negative QTL (Table 4, Figure 6).

Holsinger (1991) and Uyenoyama and Waller (1991) worked with models for the evolution of selfing, where inbreeding depression must be purged before genes favoring self-fertilization can spread. They found that initial evolution toward selfing is more likely to occur with few loci of major effect, because associations easily develop between loci affecting inbreeding depression and loci controlling selfing. In a study of the genetic architecture of floral differences between *M. guttatus* and *M. micranthus*, Lin and Ritland (1997) suggested that genes with small to intermediate effects were considered responsible to the evolution of mating system. They speculated that the evolution of self-fertilization in *Mimulus* involves the initiation of selfing by a few genes with relatively larger effects and followed by subsequent minor changes of minor modifier loci. However, we cannot verify these two hypotheses with our current study, as we cannot infer where in our three lineages the QTL actually arise.

The large evolutionary distance for the *M. platycalyx* lineage occurs because the common ancestor of all three taxa lies a significant distance along this lineage (indicated by the 'A' in Figure 4). Unlike Figure 4, the isozyme phylogeny of yellow monkey-flowers presented in Ritland and Ritland (1989) is approximately star-like for these taxa (all branch lengths the same). This is likely due to the low resolution of isozyme data (the standard errors of branch lengths were up to half the total branch length).

The variance explained by QTL is very low, on the order of 1%. In normal pairwise comparisons, QTL explain at least 5% of the variance, if not up to 20%. Although low, our lineage-specific QTL were still statistically significant in most cases. Because we simultaneously map QTL in three lineages, it seems likely that the non-significant QTL in one lineage probably dilute the variance explained by a significant QTL in another lineage. In no case were QTL significant in all three lineages, supporting this dilution effect. Computer simulation of QTL evolution and lineage-specific QTL mapping are needed to verify this assertion.

Our model and analyses assumed that QTLs are fixed between taxa, and none are segregating within taxa. This is justified because we are examining QTL differences that distinguish phylogenetic lineages; therefore QTLs are effectively homozygous compared with QTLs segregating within populations. However, if there is heterozygosity for QTL, various problems arise (Slate, 2005). F_1s may not necessarily be heterozygous for QTL, thus QTL are not detected. Linkage phase between marker and QTL will also not necessarily consistent across families. These problems arise when outbred populations are surveyed for QTL.

Natural selection versus neutrality in the context of QTL lineage mapping

The sign of QTL changes can be used to indicate whether traits have been under selection, as opposed to neutrality. Under random genetic drift, there should be roughly equal numbers of '+' and '–' QTL between taxa (Orr, 1998). This was observed for the *M. platycalyx* lineage (roughly equal + and − changes for each trait). It was not observed in the *M. micranthus* lineage, as nearly all QTL in this lineage were negative, indicating that natural selection drove the evolution of floral traits in this lineage.

Lineage-specific QTL mapping allows for a second type of Orr-type neutrality test. If the genetic basis on the evolution of mating system was solely based on drift-mutation balance, the lineages with larger branch length (for example, *M. platycalyx*, Figure 4) should have more QTL. Indeed, this lineage does have more significant QTL. In contrast, natural selection would introduce a fewer, more major QTL, taking into account the distance of lineages. This further supports the role of drift in the *M. platycalyx* lineage.

Orthology versus paralogy of QTL

The term homology was introduced by Richard Owen as the similarity of characters due to shared ancestry (Owen, 1843). With regard to sequences, Walter Fitch first developed the distinction

between orthologs and paralogs (Fitch, 1970), although Fitch was referring to gene duplication and not mutation. Homology of QTL is normally defined as QTL that map to the same position in different populations or species (for example, see Freeman *et al.* (2009)). The two QTL may or may not be the same actual coding genes, but reside within a distance of several map units. In some cases, there may be other evidence of true orthology. For example, in the Poaceae, the orthology between rice (*Oryza longistaminata*) and sorghum (*Sorghum propinguum*) may be due to common key genetic regulators of morphological development in the Poaceae (Pereira and Lee, 1995; Hu *et al.*, 2003).

In lineage-specific QTL mapping, we can more strictly classify homologous QTL as orthologous (QTLs that share common ancestry due to same mutational event of an ancestor) versus paralogous (QTLs that have different mutations in the same region of DNA, which arise independently). Lineage-specific QTL mapping approach more directly infers homology, not by comparing separate maps, but by conducting a joint analysis of three taxa simultaneously. QTL orthology is identified by the presence of a QTL in a shared lineage between two taxa. Figures 3a, c, d and e represent true orthology of QTL. With our *Mimulus* data, we found most QTL to be orthologous, as opposed to paralogous (Figures 5 and 6).

Prospects

We regard our model and analyses as mainly conceptual in nature, showing that inferences about QTL can be made on specific lineages in phylogeny, which provides novel insights into the evolution of QTL. Several improvements or extensions should be possible.

Our inferences were limited the use of AFLP markers, which was the marker available at the time of this study. The major limitation is that AFLPs are dominant and diallelic, and parents often homozygous. Microsatellite markers would overcome these problems, and would make the analysis more straightforward and powerful. Specifically, parents can be genotyped and need not be inferred by their progeny, and loci will be more informative as parents are normally heterozygous. Investigations are needed into alternative crossing designs for lineage-specific mapping of QTL. As well, more elaborate models involving dominance and epistasis would be worthwhile to pursue, and it would be useful to develop interval mapping techniques.

We have assumed the sum of the three lineage-specific QTL effects to be zero. Thus, if one lineage is significantly positive, the other two lineages should normally be negative. But because the other two lineages divide this negative effect, there is a lack of power for statistical significance for the latter two lineages. It should also be noted that Figure 3b depicts single QTL changes along a lineage; other lineages may have incurred no QTL changes but because the sum of all three lineages is zero, a 'positive' mutation in one lineage would cause 'negative' changes to occur in the other lineages even though no actual mutations occurred. This emphasizes the relativity of our three taxon comparison.

In a related vein, we have assumed the ancestral state was positive (for example, large flowers, greater stigma-anther separation) as outcrossing is generally regarded as an ancestral state, whereas inbreeding is derived. To be more generally applicable, inference of ancestral state should be incorporated. Ancestral states might be inferred by including more species into the phylogeny. However, a graphical representation of the space of QTL effects (Figures 3, 5 and 6) would be difficult to devise with more than three taxa lineage-specific analysis.

DATA ARCHIVING

Data deposited in the Dryad repository: doi:10.5061/dryad.23r11.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the reviewers for their insightful, thorough and critical comments. This research was funded by NSERC of Canada.

Bradshaw HD, Wilbert SM, Otto KG, Schemske DW (1995). Genetic-mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature* **376**: 762–765.

Broman KW (2001). Review of statistical methods for QTL mapping in experimental crosses. *Lab Animal* **30**: 44–52.

Cardon LR, Bell JI (2001). Association study designs for complex diseases. *Nat Rev Genet* **2**: 91–99.

Dole JA (1992). Reproductive assurance mechanisms in 3 taxa of the *Mimulus guttatus* complex (Scrophulariaceae). *Am J Botany* **79**: 650–659.

Doyle JJ, Doyle JL (1990). Isolation of plant DNA from fresh tissue. *Focus* **12**: 13–15.

Erickson DL, Fenster CB, StenØlen HK, Price D (2004). Quantitative trait locus analyses and the study of evolutionary process. *Mol Ecol* **13**: 2505–2522.

Fitch WM (1970). Distinguishing homologous from analogous proteins. *Syst Zool* **19**: 99–106.

Fitch WM (2000). Homology: a personal view on some of the problems. *Trends Genet* **16**: 227–231.

Freeman J, Whittock S, Potts B, Vaillancourt R (2009). QTL influencing growth and wood properties in Eucalyptus globulus. *Tree Genet Genomes* **5**: 713–722.

Holsinger KE (1991). Inbreeding depression and the evolution of plant mating systems. *Trends Ecol Evol* **6**: 307–308.

Hu FY, Tao DY, Sacks E, Fu BY, Xu P, Li J *et al.* (2003). Convergent evolution of perenniality in rice and sorghum. *Proc Natl Acad Sci USA* **100**: 4050–4054.

Hu XS, Goodwillie C, Ritland KM (2004). Joining genetic linkage maps using a joint likelihood function. *Theor Appl Genet* **109**: 996–1004.

Jain SK (1976). Evolution of Inbreeding in Plants. *Annu Rev Ecol Syst* **7**: 469–495.

Li RH, Lyons MA, Wittenburg H, Paigen B, Churchill GA (2005). Combining data from multiple inbred line crosses improves the power and resolution of quantitative trait loci mapping. *Genetics* **169**: 1699–1709.

Lin JZ, Ritland K (1997). Quantitative trait loci differentiating the outbreeding *Mimulus guttatus* from the inbreeding *M. platycalyx*. *Genetics* **146**: 1115–1121.

Mott R, Talbot CJ, Turri MG, Collins AC, Flint J (2000). A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc Natl Acad Sci US A* **97**: 12649–12654.

Nei M, Li WH (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* **76**: 5269–5273.

Orr HA (1998). Testing natural selection vs genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* **149**: 2099–2014.

Owen R (1848). *On the Archetype and Homologies of the Vertebrate Skeleton*. John van Voorst: London.

Pereira MG, Lee M (1995). Identification of genomic regions affecting plant height in sorghum and maize. *Theor Appl Genet* **90**: 380–388.

Petsko G (2001). Homologuephobia. *Genome Biol* **2**: comment1002.

Price AH (2006). Believe it or not, QTLs are accurate! *Trends Plant Sci* **11**: 213–216.

Remington DL, Whetten RW, Liu BH, O'Malley DM (1999). Construction of an AFLP genetic map with nearly complete genome coverage in *Pinus taeda*. *Theor Appl Genet* **98**: 1279–1292.

Ritland C, Ritland K (1989). Variation of sex allocation among 8 taxa of the *Mimulus guttatus* species complex (Scrophulariaceae). *Am J Botany* **76**: 1731–1739.

Ritland K, Clegg MT (1987). Evolutionary Analysis of Plant DNA Sequences. *Am Nat* **130**: S74–S100.

Schemske DW, Bradshaw HD (1999). Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proc Natl Acad Sci USAm* **96**: 11910–11915.

Slate J (2005). Quantitative trait locus mapping in natural populations: progress, caveats and future directions. *Mol Ecol* **14**: 363–379.

Uyenoyama MK, Waller DM (1991). Coevolution of self-fertilization and inbreeding depression I. Mutation-selection balance at one and two loci. *Theor Popul Biol* **40**: 14–46.

Vickery R Jr (1978). Case studies in the evolution of species complexes in *Mimulus*. *Evol Biol* **11**: 405–507.

Vos P, Hogers R, Bleeker M, Reijans M, Vandelee T, Hornes M *et al.* (1995). AFLP—A new technique for DNA fingerprinting. *Nucleic Acids Res* **23**: 4407–4414.