Lecture 04: Trait-augmented marker-based approaches

UNE course:

The search for selection

3 -- 7 Feb 2020

Bruce Walsh (University of Arizona)

jbwalsh@email.arizona.edu

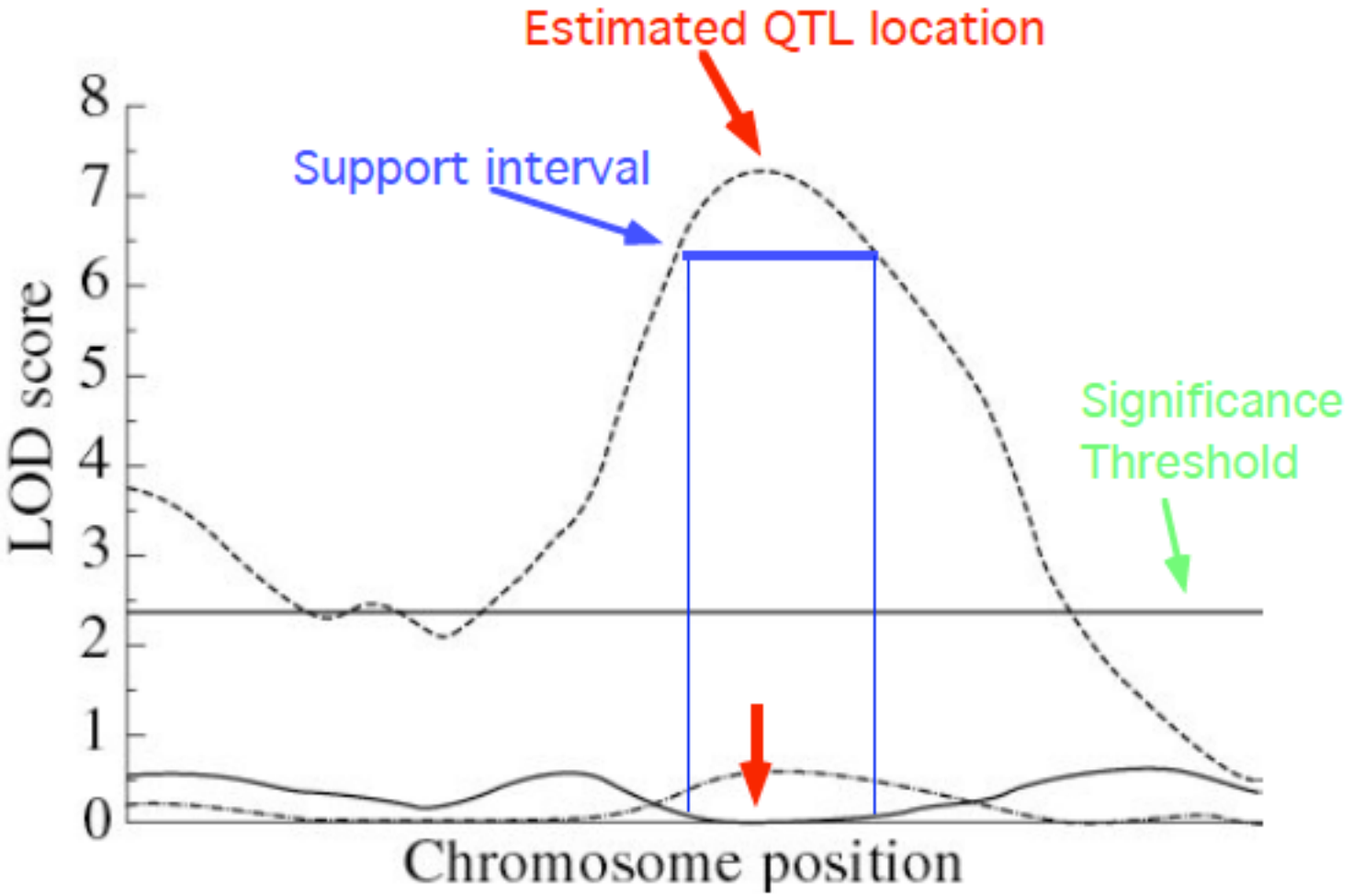
Trait-specific marker approaches

- The idea is to use trait-associated markers, such as QTLs or GWAS hits
  - Test for a nonrandom distribution of QTL/GWAS effects or frequencies
- Hence, although such tests use markers, they are *trait-specific* (unlike the tests to be covered shortly that are entirely based on markers and hence are *trait-independent*)
- Data are now categorical (as opposed to the continuous data on divergence of means)
  - Use signature over an *entire collection* of markers

Outline

- Orr's QTL sign tests
  - QTLST-EE (assumes all loci have equal effects)
  - QTLST
- GWAS- based tests
  - tSDS scores
  - Berg-Coop Q_x test
- Applications to genomic data, esp. levels of gene expression

A typical QTL map from a likelihood analysis



Allen Orr's QTL sign tests

Is there an excess of "plus" QTL alleles in the larger line?

Orr's *QTLST* and *QTLST-EE* Sign Tests

Assume that n detected QTL differences (alternative fixed alleles at n loci) are found via a standard QTL mapping experiment involving a cross between two lines (LW Chapter 15). Under neutrality, there should be no systematic directionality as to whether a line is fixed for increasing (plus) alleles over decreasing (minus) alleles at any particular QTL. This simple idea forms the basis of sign tests, but it requires modifications to account for the actual biology. For example, when the line means differ, the high (larger trait value) line is expected to contain more plus alleles (assuming equal effects; with a distribution of allelic effects, this need not be the case, as is discussed below). Orr noted that by choosing the larger line, we have introduced an ascertainment bias, as this line is *expected* to contain an excess of plus alleles. To proceed, we need some appropriate conditioning on this fact to obtain an unbiased statistic representing the value that constitutes an excess of plus alleles. The simplest approach is Orr's **equal-effects model**, where all n QTLs have close to equal effects. Here, the large line must contain at least $[n/2]$ high (plus) QTLs, where

$$[n/2] = \begin{cases} (n/2)+1 & \text{for } n \text{ even} \\ (n+1)/2 & \text{for } n \text{ odd} \end{cases}$$

In other words, the high line must contain *at least* one more high allele than the low line

In other words, the high line must contain *at least* one more high allele than the low line (because all have equal effects). Determining whether an observed number, n_{high}, of plus alleles in the high line constitutes an excess now becomes a simple combinatorial problem. The probability of k high alleles in one line (under neutrality) follows from the binomial, where there is an equal chance that a random line gets a plus or a minus allele at any particular QTL, yielding

$$\Pr(n_+ = k) = \binom{n}{k}(1/2)^k(1/2)^{n-k} = \binom{n}{k}(1/2)^n$$

Note that all values of k contain a $(1/2)^n$ term. We now condition this probability of k alleles in the high line on the fact that this line *must* contain at least $[n/2]$ plus alleles, yielding

$$\Pr(n_+ \geq n_{high} \,|\, n_+ \geq [n/2]) = \frac{\Pr(n_+ \geq n_{high})}{\Pr(n_+ \geq [n/2])} = \sum_{i \geq n_{high}}^{n} \binom{n}{i} \Big/ \sum_{j \geq [n/2]}^{n} \binom{n}{j} \qquad (12.31a)$$

where the common term of $(1/2)^n$ in both the numerator and denominator cancels. This is Orr's **QTL sign test for equal effects**, or ***QTLST-EE***. Orr noted that a *minimum* of $n = 6$ detected QTLs is required for this test to be applied. To see this, note for $n = 6$ that $[n/2] = 4$, and the most extreme value, $n_{high} = 6$, gives a p value of $1/22 \sim 0.05$, while for $n = n_{high} = 5$, the smallest p is $1/16 \sim 0.0625$. For large values of n, Orr noted that Equation 12.31a can be approximated by a normal, with

$$\Pr\left(n_+ \geq n_{high} \,|\, n_+ \geq [n/2]\right) \simeq 2\left[1 - \Phi\left(\frac{n_{high} - [n/2]}{\sqrt{n/4}}\right)\right] \tag{12.31b}$$

where $\Phi(x) = \Pr(U \leq x)$ for $U \sim N(0,1)$.

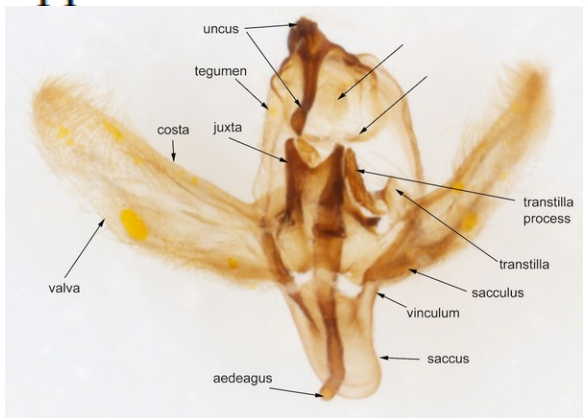
Example 12.8. True et al. (1997) found that none of the eight detected QTLs for the posterior lobe in the male genitalia in a *Drosophila* cross were **antagonistic**, i.e., with effects in the opposite direction of the line value, such as low (minus) alleles in the high line or high (plus) alleles in the low line. Orr suggested that the equal-effects model may be reasonable for this trait. Assuming that this model holds, we have $n = 8$, $[n/2] = 5$, $n_{high} = 8$, and Equation 12.31a yields the probability of having all eight detected QTLs from the high line being plus alleles as

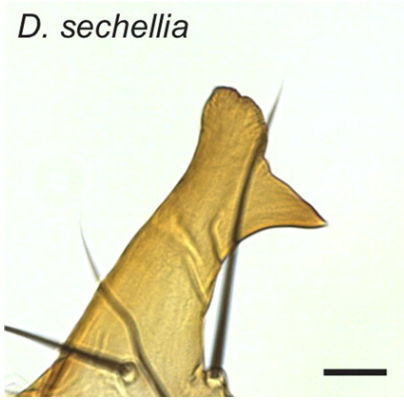
$$\sum_{i=8}^{8} \binom{8}{i} \Big/ \sum_{j\geq 5}^{8} \binom{8}{j} = \frac{\binom{8}{8}}{\binom{8}{5}+\binom{8}{6}+\binom{8}{7}+\binom{8}{8}} = \frac{1}{56+28+8+1} = 0.011$$

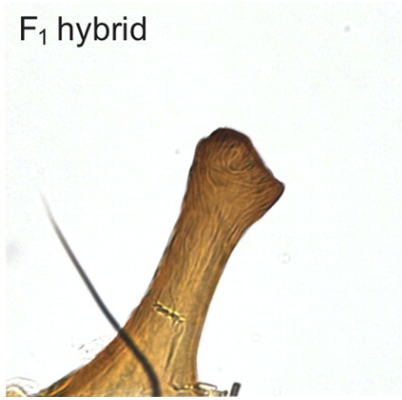
showing that this is a highly significant excess. Orr's normal approximation (Equation 12.31b) yields

$$p \simeq 2\left[1-\Phi\left(\frac{8-5}{\sqrt{8/4}}\right)\right] = 2\left[1-\Phi\left(2.1213\right)\right] = 0.034$$

The latter approximation is rather conservative, but not surprising, as this is a large-sample approximation and the number of detected QTLs here is very modest.







Under this strong assumption of equal effects, *QTLST-EE* is a nonparametric test, making no other assumptions, and not using any information on the actual difference between the high and low lines. Anderson and Slatkin (2003) noted that this test can be highly biased by trait choice (whereby the investigator, often unconsciously, chooses traits showing excessive divergence). While Orr's approach corrects for ascertainment bias *within* any specific trait comparison, it assumes that the traits were chosen at random. To examine the impact from nonrandom trait sampling (which, as previously discussed, also biases Q_{ST} tests), Anderson and Slatkin simulated T identical and independently distributed traits, each with 10 QTLs of equal effects, and then chose the most divergent single trait from this set for the subject of a *QTLST-EE* test. They found that this process of trait ascertainment introduces a significant bias. For $n = 10$ QTLs, Equation 12.31a gives the probability of 9 or more plus alleles in the high line as 0.0285. However, when the trait was not randomly chosen, but rather the high line from the *most divergent* trait in a set of 25 traits was used, then over 50% of the time it contained at least 9 high alleles. This lack of robustness with respect to the trait ascertainment scheme means that significant *QTLST-EE* results must be interpreted with caution.

A second class of tests proposed by Orr avoids this problem, and indeed, simulations show that it is conservative (Anderson and Slatkin 2003; Rice and Townsend 2012). For these tests, let D be the difference between the high and low lines. This may be either the actual observed difference, or the difference based on summing the effects over all detected QTLs. With a distribution of QTL effects in hand, one can then condition on the number of plus alleles, *given* the observed difference, D. This is Orr's **QTL sign test** (***QTLST***). The seemingly problematic issue of the distribution of QTL effects can be easily handled via a bootstrap approach in one of two ways. First, one could use the observed distribution of absolute QTL effects ($|a|$), and then fit this using some standard distribution. Orr used the gamma distribution (Equation A2.25a; Figure A2.2) because of its flexibility and the fact that the exponential, a commonly assumed distribution of effect sizes (Chapter 27), is a special case. Note that estimating the distribution parameters that give the best fit is done using a truncated distribution, as QTL effects below a critical absolute size would not be detected. One can then generate the p value for the observed number of plus alleles through parametric bootstrapping. To do so, we generate n draws from this distribution, randomly assign each a sign, and only keep those samples for which the total (G) of signed QTL effects equals or exceeds D. The resulting distribution of plus alleles in the high line is now conditioned on neutrality (signs drawn at random), the assumed distribution of QTL effects, and the actual divergence D, yielding

$$p = \Pr(n_+ \geq n_{high} \,|\, G \geq D) = \sum_{i \geq n_{high}}^{n} \Pr(n_+ = i \,|\, G \geq D) \qquad (12.32)$$

where $\Pr(n_+ = i \,|\, G \geq D)$ is simply the fraction of times that exactly i plus alleles were found in the retained bootstrap samples (i.e., those showing a divergence of at least D). Alternatively, instead of sampling from the fitted distribution, one could use standard bootstrapping and directly sample (with replacement, and with draws randomized with respect to sign) from the observed distribution of QTL effects (e.g., Rice and Townsend 2012).

While the *QTLST* adjusts for ascertainment bias, it does so the expense of power. As noted by Rice and Townsend (2012), the difference (D) provides some information on the amount of any previous selection, and by conditioning on its value, we are removing this evidence. Consider the extreme case where a line fixes plus alleles at all ten QTLs, and all have equal effect. In order to obtain the observed value of D, we must condition on only those cases where all ten are fixed, giving this test zero power (Griswold and Whitlock [2003] also noted the low power of this test). Rice and Townsend found that both the power and the false-positive rate increase with the variance of QTL effects.

Applications of QTL Sign Tests

Using *QTLST-EE*, Rieseberg et al. (2002) performed a meta-analysis of over 2600 QTL effects from 572 traits in 86 studies. Their summary statistic was the QTL ratio: the fraction of antagonistic QTLs for the comparison of interest (Table 12.3). Roughly half of the studies involve wild × domesticated crosses, where strong directional selection is suspected for domestication traits. Upon restricting analysis to those examples with six or more QTLs per trait (Orr's condition for such tests to have any power), 35 of the 54 qualifying traits (65%) believed to be involved in domestication showed significant departures from neutrality (i.e., too few antagonistic QTLs). By contrast, only 14 of 84 nondomestication traits (15.6%) in crosses involving domesticated species showed significant departures. Treating this latter class of traits as a control demonstrates that *QTLST-EE* behaved in the direction predicted for these crosses (revealing signatures for domestication traits and a lack of signatures for nondomestication traits).

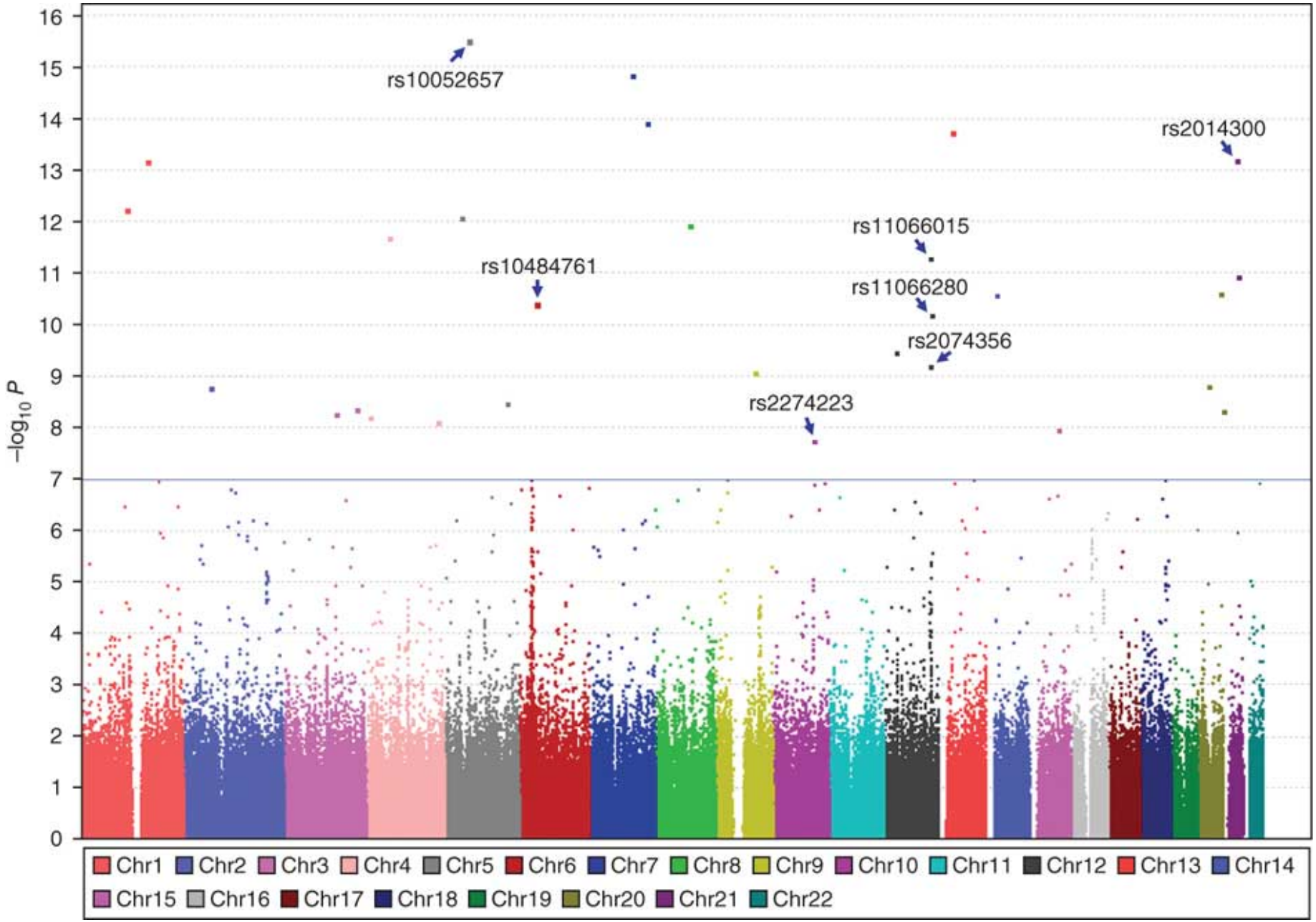
Loren Rieseberg

Trait Category	Antagonistic QTLs	Total QTLs	QTL ratio	LS Means
Animals	73	312	0.234**	0.185 ± 0.039†
Plants	128	439	0.292**	0.202 ± 0.025†
Interspecific	47	245	0.192**	0.137 ± 0.154
Intraspecific	154	506	0.304**	0.250 ± 0.243
Outcross	98	425	0.231**	0.170 ± 0.174
Self	103	326	0.316**	0.217 ± 0.262
Life history	111	540	0.206**	0.139 ± 0.175
Morphology	138	508	0.272**	0.266 ± 0.255
Physiology	8	40	0.200*	0.176 ± 0.125
Phenology	37	124	0.298**	0.236 ± 0.219
Total	201	751	0.268	

We conclude by briefly highlighting the utility of two applications of sign tests to specific biological problems (as compared to the broad generalizations explored above). Albertson et al. (2003) examined traits in the massive species radiation occurring among cichlid fishes in the East African rift lakes. One striking feature of this radiation is extensive convergent evolution across lakes in feeding morphology, suggesting parallel directional selection. The authors used *QTLST* (with effect sizes drawn from a gamma distribution) to examine the genetic basis of feeding morphology through crossing two wild species from Lake Malawi. Because most individual traits had less than six QTLs, they grouped the traits, finding that only 4 of the 46 QTLs were antagonistic for jaw and teeth features. The highly significant p value supports directional selection on these feeding traits. Muir et al. (2014) examined QTLs in tomatoes (*Lycopersicon*) to explore leaf-related traits in wild species thought to be associated with adaptation to precipitation. They found no significant departure from neutrality for two leaf and two trichome (leaf hair) traits, but a significant departure from neutrality for two stomatal traits. They computed p values using both *QTLST* and *QTLST-EE*, and they found (in agreement with Anderson and Slatkin 2003) that *QTLST* was more conservative, yielding p values about twice as large as those obtained from *QTLST-EE*.

GWAS-based approaches

- Usually a lot more GWAS "hits" than QTL hits for a given trait.
- Again, the key idea is to perform a test over a collection of markers for a given trait and contrast this with some neutral expectation.


16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
−log10 P
rs10052657
rs2014300
rs11066015
rs10484761
rs11066280
rs2074356
rs2274223
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12 Chr13 Chr14
Chr15 Chr16 Chr17 Chr18 Chr19 Chr20 Chr21 Chr22

Approaches Based on Combining Signals

The basic idea of combining signals over a set of GWAS markers has been exploited in several different ways. The initial suggestion was **gene set enrichment analysis (GSEA)** from genomics (Subramanian et al. 2005), wherein one considers clusters of genes on the basis of membership in some functional group (i.e., the same gene ontology, GO, class). This tactic was used by Daub et al. (2013), who computed the average F_{ST} value over a set of pathway-connected genes and contrasted this with the average F_{ST} value over a same-sized set of putatively neutral markers. Using this approach, they found evidence for selection on several human pathways, many connected with pathogen response. They also noted that long-distance LD was detected, which they attributed to epistatic interactions. While this could be correct, a confounding factor is that selection is also expected to generate such long-distance (i.e., between loosely-linked sites) LD with strictly additive genes (Chapters 16 and 24).

In the Daub et al. analysis, the "trait" was a specific pathway, while other analyses have consider more classical human traits, in particular, height. A simple, but robust, approach was used by Turchin et al. (2012). They examined allele frequency differences for 139 GWAS markers for height between Northern- and Southern-European populations. Under neutrality, allele-frequency increases in plus alleles should be randomly distributed between two populations (i.e., the sign test introduced above). Instead, what they found was that 85 of the 139 markers (sign test $p = 0.01$) showed an increase for high alleles in the Northern-European population. Note that one advantage of GWAS data is that typically a reasonable number of hits (marker-trait associations) are found, while QTL-based studies often fail to have more that five detected QTLs for a focal trait (and hence Orr's test is not applicable).

Using GWAS information

Tests Based on *tSDS* Scores

Recall Field et al.'s (2016) singleton density score, *SDS* (Equation 9.42), for detecting very recent selection on a given single site. In that paper, they also showed how to extend this approach to search for polygenic selection on a given candidate trait, given a set of associated GWAS marker scores. This requires that both the *SDS* values and GWAS test statistics (such as a z value under a normality test) for a set of markers were generated using the same population. The *SDS* score for a given marker is first translated into a ***tSDS*** (**trait-*SDS***) score, where the sign of the *SDS* score is changed so that trait-increasing markers receive positive scores. Their simplest approach was to combine the *tSDS* scores associated with all the significant GWAS markers for a target trait, using this mean as the test statistic.

Field et al. noted that most of the trait variance is usually explained by markers whose GWAS test statistics do not pass the genome-wide significance threshold (Chapter 24), and hence are not included in the test set. To incorporate information from these nonsignificant (but potentially biologically important) markers, they used a regression-based approach. Data points for the regression were generated by first binning SNPs with very similar GWAS scores, taking the bin average GWAS score as the predictor variable and bin average *tSDS* score as the associated response variable. A significant regression (or correlation) is expected under selection, but not under drift. Both the average *tSDS* score and regression approaches detected clear signals of selection for increased height in Britain over the past 2000 to 3000 years. Several other traits (infant head size, body mass index [BMI], and female hip size, to name a few) also showed evidence of recent polygenic selection.

The Berg and Coop Q_x Test

The final approach leveraging GWAS-estimated marker effects for a target trait is due to Berg and Coop (2014), building on previous work by them (Coop et al. 2010; Günther and Coop 2013) as well as by Ovaskainen et al. (2011). Let $\mathbf{p}_i^T = (p_{i,1}, p_{i,2}, \cdots, p_{i,m})$ denote the vector of allele frequencies for the ith marker over m subpopulations, where $p_{i,j}$ denotes the allele frequency in population j. Example 9.5 showed that the expected distribution of $\mathbf{p}_i$ under neutrality is approximately given by

$$\mathbf{p}_i \sim \text{MVN}_m \left[p_{i,0}\mathbf{1}, p_{i,0}(1-p_{i,0})\boldsymbol{\Omega} \right] \tag{12.33a}$$

where $\boldsymbol{\Omega}$ is a (marker-estimated) matrix of expected covariances in allele frequencies over the subpopulations and $p_{i,0}$ is the allele frequency in the ancestral population. As detailed in Chapter 9, this formed the basis of Coop's *Bayenv* test for excessive divergence at a *specific locus*. Berg and Coop (2014) extended this result to a *trait-based* test as follows. Suppose n GWAS hits are discovered for the focal trait, where the trait-increasing allele for the ith marker changes the trait by a value of g_i, with $p_{i,j}$ denoting the frequency of this allele in population j. The GWAS-predicted mean genetic value for the trait in population j thus becomes

$$a_j = 2\sum_{i=1}^{m} g_i\, p_{ij} \tag{12.33b}$$

To proceed, Berg and Coop expressed all of the a_j values as deviations from the grand mean, yielding $a_j^* = a_j - \bar{a}$. This uses one degree of freedom, and returns the vector $(\mathbf{a}^*)^T = (a_1^*, a_2^*, \cdots, a_{m-1}^*)$, where one population is dropped. As Berg and Coop note, information from the dropped population is fully retained by the vector $\mathbf{a}^*$, so that the choice of which population to drop has no impact on the resulting analysis. The resulting vector is now distributed as

$$(\mathbf{a}^*)^T \sim \text{MVN}_{m-1}\left[\mathbf{0}, 2V_A\boldsymbol{\Omega}\right] \tag{13.33e}$$

As discussed in Appendix 5, a standard trick with a vector of correlated variables is to use a transformation to return a vector of uncorrelated variables of unit variance. Berg and Coop did this by using the Cholskey decomposition (Appendix 5) of $\boldsymbol{\Omega} = \mathbf{C}\mathbf{C}^T$, using the transformation

$$\mathbf{x} = \frac{1}{\sqrt{2V_A}}\mathbf{C}^{-1}\mathbf{a}^* \tag{13.33f}$$

which returns

$$\mathbf{x} \sim \text{MVN}_{m-1}(\mathbf{0}, \mathbf{I}) \tag{13.33g}$$

This is the basis for the **Berg-Coop Q_x test**, whose statistic is given by

$$Q_x = \mathbf{x}^T\mathbf{x} = \frac{(\mathbf{a}^*)^T\boldsymbol{\Omega}^{-1}\mathbf{a}^*}{2V_A} \tag{12.33h}$$

Under neutrality, $Q_x \sim \chi^2_{m-1}$, as $\mathbf{x}^T\mathbf{x}$ is the sum of $(m-1)$ squared unit-normal random variables. Note by comparing this result to Equation 9.13c, that the Q_x test is very similar in form to the Günther-Coop (2013) $\mathbf{x^T x}$ test for selection on a single site, but with estimated trait genetic values replacing allele frequencies.

Robinson et al. (2015) applied this test to height and BMI based on ~9400 individuals from 14 European countries, finding evidence that selection favored increased height and reduced BMI. Mathieson et al. (2015) also applied this test to Europeans, but used ancient DNA from 230 individuals (who lived between 6400 to 300BC), and reported evidence for two independent episodes of selection for height.

Divergence in gene expression levels

- The amount of mRNA for a candidate gene is a quantitative trait, and hence one can test for whether an observed amount of divergence is excessive
- Less of an ascertainment issue, as one can score the expression levels of all genes within a given genome
- *Cis* and *trans* versus local and distant
- ASE (Allele-specific expression)

Applying Lande's test

With this concern in mind, the version of Lande's test used by these investigators starts by noting that $E[V_A] = 2N_e\sigma_m^2$ for an additive trait at mutation-drift equilibrium (Equation 11.20c). Assuming expression values are drawn from a normal distribution under the null model, then when L lineages are used to estimate the among-group variance and k individuals per line were used to estimate V_A, Equation 12.5a shows that both estimators approximately follow χ^2 distributions, with

$$V_B \sim (2\,t\,\sigma_m^2)\cdot\chi^2_{L-1}/(L-1) \qquad \text{and} \qquad V_A \sim (2N_e\,\sigma_m^2)\cdot\chi^2_{k-1}/(k-1) \tag{12.34a}$$

where t is the time of divergence since the common ancestor. These expressions suggest a modified version of Lande's F_{MDE} test statistic,

$$F^*_{MDE} = \frac{V_B/(2t\sigma_m^2)}{V_A/(2N_e\sigma_m^2)} = \frac{V_B}{V_A}\cdot\left(\frac{N_e}{t}\right) \sim \frac{\chi^2_{L-1}/(L-1)}{\chi^2_{k-1}/(k-1)} \tag{12.34b}$$

where this statistic follows an F distribution, with $F^*_{MDE} \sim F_{L-1,k-1}$ (as it is the ratio of two χ^2 random variables, scaled by their degrees of freedom; see LW Appendix 5). A scaled ratio less than a critical value of $F_{\alpha/2}$ is suggestive (at level α) of too little divergence, and hence suggestive of stabilizing selection, while a scaled ratio in excess of $F_{1-\alpha/2}$ implies too much divergence, suggestive of directional selection. These critical values are given by

$$\frac{V_B}{V_A} \leq F_{\alpha/2,L-1,k-1}\left(\frac{t}{N_e}\right) \qquad \text{and} \qquad \frac{V_B}{V_A} \geq F_{1-\alpha/2,L-1,k-1}\left(\frac{t}{N_e}\right) \tag{12.34c}$$

where $F_{\alpha,M,N}$ denotes critical values for an F distribution and satisfies

$$\Pr\left(F_{M,N} \leq F_{\alpha,M,N}\right) = \alpha$$

In the case where just two populations ($L = 2$) are compared by using their squared difference, d^2, then recalling that $V_B = d^2/2$ (Equation 12.8c), the conditions given by Equation 12.34c become

$$\frac{d^2}{V_A} \leq F_{\alpha/2,1,k-1}\left(\frac{2t}{N_e}\right) \qquad \text{or} \qquad \frac{d^2}{V_A} \geq F_{1-\alpha/2,1,k-1}\left(\frac{2t}{N_e}\right) \tag{12.34d}$$

Example 12.9. Rifkin et al. (2003) examined variation in gene expression at the start of metamorphosis in six inbred lines of *Drosophila*: four *melanogaster*, one *simulans*, and one *yakuba*. Of the roughly 12,900 genes whose transcripts were examined, 52% (~6700 genes) showed expression changes in at least one lineage (either between species or within the *melanogaster* lines). For ~4500 of these genes, the authors could not reject the hypothesis that all six lineage-specific samples came from the same distribution, and these were deemed to be evolutionarily stable and potentially under stabilizing selection. Of the remainder, ~1700 genes showed no significant variation across the sampled *melanogaster* lines, but a significant difference between *melanogaster* and one of the other species. These were deemed to be under lineage-specific selection.

The evolutionary forces acting on the remaining 527 genes were examined using Equation 12.34b. Divergence was scored separately between *melanogaster* and each of the other two species ($L = 2$) using d^2, with V_A estimated from the among-group variance in four fully inbred *D. melanogaster* lines ($k = 4$). Because d^2 is used, critical values are given by Equation 12.34d, with one correction. The expected among-group variance (for an additive trait) between a set of fully inbred lines is twice the additive variance (from Table 11.3, with $2f = 2$), so that $2N_e$ replaces N_e in the critical values. The resulting upper and lower 2.5% critical values follow

first by noting that $\Pr(F_{1,3} \geq 17.4) = 0.025$ and $\Pr(F_{1,3} \leq 0.001) = 0.025$. These authors used an estimated effective population size for *D. melanogaster* of $N_e = 3 \cdot 10^6$, while the total divergence times (twice the separation time, in generations) were estimated as $2t = 4.6 \cdot 10^7$ (*melanogaster-simulans*) and $2t = 10.2 \cdot 10^7$ (*melanogaster-yakuba*). Hence, the critical values for excessive divergence were

$$F_{c,mel\text{-}sim} = 17.4 \cdot \frac{4.6 \cdot 10^7}{6 \cdot 10^6} = 133.4 \quad \text{and} \quad F_{c,mel\text{-}yak} = 17.4 \cdot \frac{10.2 \cdot 10^7}{6 \cdot 10^6} = 195.8$$

Transcripts whose ratio of d^2/V_A exceeded these values are unusually divergent. Using this criterion (as well as the lower threshold for too little divergence), of these remaining 527 genes, 464 were consistent with drift, while 63 were consistent with excessive divergence between at least one species pair.

Brownian motion or Ornstein-Uhlebeck?

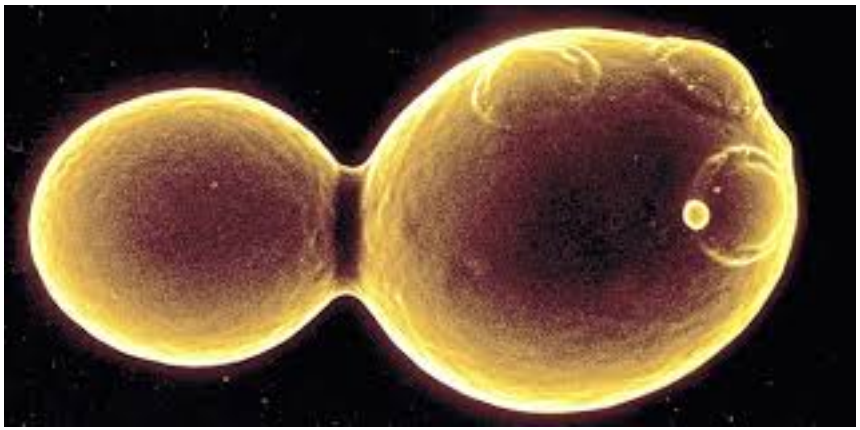
Under a Brownian motion model, the expected divergence (measured by the among-group variance) scales linearly with divergence time, t (Equation 12.10, under the infinite-alleles assumption). In contrast, under an Ornstein-Uhlenbeck (OU) process (drift countered by stabilizing selection), the total divergence approaches an asymptotic value (Equation 12.22c). Bedford and Hartl (2009) used an OU process to fit the pattern of expression divergence within a clade of seven species of *Drosophila*. In accordance with the OU model (and consistent with stabilizing selection), they found that the divergence variance does not increase linearly with time but, rather, quickly approaches an asymptotic value.

In contrast, Khaitovich et al. (2004, 2005) argued that gene expression can evolve in a mostly neutral fashion, based in large part on an observed linear increase in the divergence of among-species expression with time within the clade of great apes. They also noted the observation of Riftkin et al. (2005), namely, a positive correlation between levels of within- and among-species variation for the expression of different genes. Such a pattern is expected under neutrality, as both divergence and standing variation are functions of σ_m^2. However, this is not strong support for neutrality, as a number of other features can create such a correlation. For example, genes whose expression is strongly influenced by the environment may naturally exhibit higher levels of variation, both within and among samples. Likewise, linearity in divergence, by itself, is suggestive, but not sufficient. Unless the actual rate of divergence is consistent with the rate of polygenic mutation, linear patterns of evolutionary diversification need not imply neutrality.

Applications of Sign-based Tests to Expression Data

While Orr's tests were framed in the increasingly dated technology of QTL mapping, their central underlying idea (effects are randomly distributed among lines under neutrality) fits very nicely with genomics-era data. We already mentioned a GWAS application of sign-based tests, and there is an increasing use of sign-based approaches to explore the nature of selection on gene expression. The standard QTL-based tests discussed above are not directly applicable, as most genes have very few detected eQTLs, and thus do not qualify for testing based on Orr's requirement of at least six QTLs per trait. However, as reviewed by Fraser (2011), two rather different approaches have been used to circumvent this limitation.

The first approach is simply to shift focus from the expression levels at *single genes* to the pattern of expression over a *set of genes*, pooling these to create a setting with more than six eQTLs for the trait. Bullard et al. (2010) used this approach in a cross of two closely related yeast species, *Saccharomyces cerevisiae* and *S. bayanus*. One key requirement in the statistical analysis is that each eQTL is independent, as a single eQTL that simultaneously influences k genes should be weighted as one change, not k changes, in the same direction. The use of *cis*-regulatory alleles ensures independence over a set of loosely linked genes. Bullard et al. accomplished this by only considering alleles showing ASE. An excessive number of up-regulated ASEs over a specific gene set from one species indicates the presence of lineage-specific selection, and a number of pathways were detected showing this feature. Fraser et al. (2011) used a similar approach in a cross of two subspecies of the mouse (*Mus musculus*). They chose gene sets defined by shared GO (Gene Ontology Consortium) membership, and found over 100 genes with evidence of lineage-specific selection. These studies are important, as (at least for these two crosses) they suggest that adaptation via gene-expression changes may be widespread, highly polygenic, and involves *cis*-regulatory sites.



A second modification of a sign-based test for expression data, which was offered by Fraser et al. (2010), applies to genes whose expression levels are influenced by both *cis* and *trans* eQTLs. The central premise of sign-based tests is that ***directionality is random under the null***, so that in a cross of lines A × B, if an eQTL from A is a *cis* up-regulator, this should provide no information as to whether a *trans*-acting factor from A (acting on the expression level at the same target gene) is an up- or down-regulatory allele. *Cis*- and *trans*-acting alleles whose influence is in the same direction (up and up, down and down) are called **reinforcing**, and those acting in opposite directions are called **opposing**. With a collection of genes whose expression is influenced by both *cis* and *trans* eQTLs, a simple 2×2 contingency table can be constructed and tested for departures from randomness. If a significant departure is seen, it is a straightforward process to estimate the amount of excess in a particular class (e.g., Example 10.1). Fraser et al. applied this approach in a cross of two yeast (*Saccharomyces cerevisiae*) strains that diverged roughly 10^7 generations ago and found an excess of roughly 242 genes showing reinforcing levels of *cis* and *trans*. While this approach suggests significant regulatory evolution over the genome, it does not indicate *which* specific transcripts are involved. This result is reminiscent of some of the approaches for detecting genome-wide signatures of selection examined in Chapter 10: evidence of a genome-wide pattern is seen, but no particular gene can be singled out with confidence as being a target of the selection process generating the observed pattern.

Example 12.10. Using *cis* and *trans* expression data, Emerson et al. (2010) suggested a test for neutral expression evolution that is related in spirit to sign-based tests. They combined their polymorphism data with divergence data from Tirosh et al. (2009) to examine the within- and among-species expression control in the yeasts *Saccharomyces cerevisiae* and *S. paradoxus*. They used a MacDonald-Kreitman approach (Chapter 10) by examining the fraction of *cis*- and *trans*-controlled transcripts measured within and between species. Their resulting contingency table,

	Polymorphism	Divergence
Cis	396	1270
Trans	412	541

was highly significant, with *trans* polymorphisms being slightly more common than *cis*, but over twice as many *cis* regions were fixed. Such a pattern could arise from either an excessive number of *cis* fixations between species, an excessive amount of *trans* polymorphism within a species, or a combination of both. Analogous to arguments with interpretations of MacDonald

Example 12.11. Fraser (2013) suggested yet another approach for detecting selection on the expression level of specific sets of transcripts. First, the genes comprising a specific functional set are chosen (e.g., UV protection), and then an expression score for a population is calculated from the mean frequency of all eSNPS that up-regulate members of this set. This was done over a series of roughly 60 human populations that have lived and evolved under different environmental conditions (such as summer UV flux), computing the correlation between the environment variable and expression score. The significance of this correlation was tested using a randomization approach, wherein the correlation between expression score and environmental variable is computed using a random set of genes. This was repeated $\sim 10^6$ times to generate a distribution for each gene-environment correlation under the null. This approach is very similar to methods examined in Chapter 9 to search for individual SNP frequency–environmental associations (Hancock et al. 2010a, 2010b, 2011; Fumagalli et al. 2011). However, the latter is performed on a SNP-by-SNP basis, whereas the focus here is on the *entire set* of regulatory actors in some network. Using Fraser's approach, significant signals were detected for transcript sets involved in UV response, immune cell proliferation, and diabetes. Further, using a catalogue of putative locally adaptive human SNPs, Hancock et al. (2011) found a roughly ten-fold enrichment of eSNPS and SNPs in *cis*-regulatory regions over amino-acid replacement SNPs in the same genes. This suggests a more important role in local adaptation for regulatory, as opposed to structural, changes.