SI Appendix

Phenotypic effect

Two different estimations of phenotypic effect were used. First, we estimated phenotypic effects as a function of the phenotypic difference between the CSS Donor and Host strains:

$$ES_i = 100*[CSS_i - Host]/[Donor - Host]$$

Second, we noted that the range of mean trait values among the strains in a CSS panel as well as the two parental strains defines the total phenotypic variance, which includes both the genetic and error variance. We therefore computed the phenotypic effect (ES_i) for the *i*th CSS as:

$$ES_i = 100*[CSS_i - Host]/[Hi - Low]$$

where Host is the phenotypic value of the strain that was used as the background onto which chromosomes from the donor strain were substituted, Hi is the highest mean phenotypic value among the progenitor strains and the CSS panel, and Low is the strain with the lowest mean phenotypic value.

Standard error of measurement (SEM) for the cumulative effect

Next we consider methods for estimating SEM of individual and cumulative effects variable *w* from other variables.

In general, if $w = f(\mathrm{x, y}, \ldots)$, then the SEM of w, $S(w)$ is $g(\mathrm{x, y}, \ldots)$. For simplicity, we only use two variables in the following equation:

$$S(w) = \sqrt{\left(\frac{\partial w}{\partial x} \cdot S(x)\right)^2 + \left(\frac{\partial w}{\partial y} \cdot S(y)\right)^2 + 2r_{xy}\left(\frac{\partial w}{\partial x} \cdot S(x)\right) \cdot \left(\frac{\partial w}{\partial y} \cdot S(y)\right)}$$

If the correlation coefficient of relationship between two variables *x* and *y*, r_{xy}, is 0, then the estimation of *S(w)* can be simplified as the following:

$$S(w) = \sqrt{\left(\frac{\partial w}{\partial x} \cdot S(x)\right)^2 + \left(\frac{\partial w}{\partial y} \cdot S(y)\right)^2}$$

We used this equation to calculate the SEM for the individual and cumulative phenotypic effects. Let X be the mean trait value of CSS_i, Y be the trait value for C57BL/6J, R the value for A/J, P the value for Hi, and Q the value for Low. Let S_i be the standard error of the phenotypic effect of CSS_i, w_i be the effect of *i*th CSS compared to C57BL/6J. The following calculations were made:

- For each trait, estimate phenotypic effect in their measured units for CSS_i, the cumulative phenotypic effect, the SEM of the phenotypic effect, and finally the SEM of cumulative effect:

  The phenotypic effect for CSS_i is $w_i = (X_i - Y)/(R - Y)$.

  The cumulative phenotypic effect is $\sum w_i = \sum (X_i - Y)/(R - Y)$.

  The SEM for the phenotypic effect of CSS_i is

  $$S_{w_i}^2 = C_0 S_i^2 + C_1 (X_i - R)^2 + C_2 (X_i - Y)^2$$

  where $C_0 = 1/(R - Y)^2$, $C_1 = C_0^2 S_Y^2$, and $C_2 = C_0^2 \cdot S_R^2$.

  The SEM for the cumulative phenotypic effect is $S_{cum}^2 = \sum S_{w_i}^2$.

  Let W be the difference between the cumulative phenotypic effect and that of parental effect:

  $$W = \sum_i (X_i - Y)/(R - Y) - 100\%$$

  The SEM of W is the same as that for the cumulative phenotypic effect.

- For each trait, estimate the phenotypic effect for CSS_i with the Hi-Low method. In general, SEM of the cumulative effect is estimated as follows:

  after the derivative,

  $$S_{w_i}^2 = \frac{1}{(P-Q)^2}(S_{X_i}^2 + S_Y^2) + \frac{(X_i - Y)^2}{(P-Q)^4}(S_P^2 + S_Q^2)$$

  where P, Q, X and Y are defined above.

  Let $C_1 = 1/(P-Q)^2$, $C_0 = C_1 \cdot S_Y^2$, and $C_2 = C_1^2 \cdot (S_P^2 + S_Q^2)$

  Then for each CSS effect, SEM square is:

  $$S_{w_i}^2 = C_0 + C_1 \cdot S_i^2 + C_2 \cdot (X_i - Y)^2 = C_0 + C_1 \cdot S_i^2 + C_2 \cdot (X_i - Y)^2 .$$

  The cumulative effect is estimated as follows:

  $$S_{cum}^2 = \sum S_{w_i}^2 .$$

  For the difference between cumulative effect and 100%, S is the same as S^2_{cum}.

  But there are exceptions where the SEM of the cumulative effect is estimated differently.

  a. When CSS_i is Hi or Low, the estimation of SEM is as follows:

  $$W_i = \frac{X_i - Y}{X_i - Q} = 1 + \frac{Q - Y}{X_i - Q}$$

  $$S_{w_i}^2 = C_0 + C_1^2 \cdot [(X_i - Q)^2 \cdot S_P^2 + (X_i - P)^2 \cdot S_Q^2]$$

  where *i* is the *i*th CSS with highest or lowest mean trait value, C_0, and C_1 are defined as before.

  b. When C57BL/6J or A/J is Hi or Low, the estimation of SEM is as follows:

  without loss of generality, assume that B6 is Hi, then P = Y

$$W_i = \frac{X_i - Y}{Y - Q}$$

$$S_{w_i}^2 = C_1 \cdot S_{X_i}^2 + C_1^2 \cdot [(X_i - Q)^2 \cdot S_P^2 + (X_i - P)^2 \cdot S_Q^2]$$

When C57BL/6J is Hi or Low and one of the CSSs is Low or Hi, it is straightforward that ES = 100% and the standard error is estimated as zero.

Clustering of QTL direction effects

For any given trait, assume that N bins, the maximum trait value is Max, the minimum trait value is Min and Interval is (Max-Min)/2N. Given any trait value x, the estimated bin number, B, that x will be assigned to is:

B= [((x – Min-Interval) * N /(Max - Min))],

where [x] represents the nearest integer to x, and the residual is

res = (x – Min-Interval) * N/(Max-Min) – B

The number of trait that falls into B is updated as following:

Hist(B) = Hist(B)+ (1 –abs(res/(Max-Min)));

Hist(B+sign(res)) = Hist(B+sign(res)) + abs(res/(Max-Min)) such that the residual contributes to the closer bin of the two adjacent bins. A Gaussian filter was then used to smooth the resulting histogram for each trait. A vector was constructed to represent each trait, each of which consists of the number of peaks that appear before and after r equals 0 (see above) and their corresponding locations.

K-means was used to cluster the histograms of all 41 traits with CSS entries that differed significantly from C57BL/6J. The maximum cluster number was set to four, with squared Euclidean distance as the metric and a maximum iteration of 100. Empty clusters were not considered in subsequent analyses.