Statistical tests for selective neutrality and hitchhiking of 10 segments with highest CHE scores

For each segment, we denote its two alleles as $A$ and $a$. We consider inducers developed from crosses of type 1 ($I \times N$), type 2 ($(I \times N) \times (I \times N)$) or type 3 ($(I \times N) \times I$), where $I$ represents an inducer genotype homozygous for presence of the $A$ allele and $N$ is a non-inducer sampled at random from the set of non-inducers which has frequency $p_A^*$ for allele $A$ and frequency $(1 - p_A^*)$ for allele $a$. According to the description in Supplementary Table 1, we have 11 crosses of type 1 and 2, and 8 crosses of type 3.

The probability of recovering genotype $AA$ in a progeny inducer $I$

The probability of recovering genotype $AA$ in a progeny inducer $I$ descending from one of the three type of crosses described above is given by:

$$P[I = AA] = p_A^* \times 1 + (1 - p_A^*) \left( \frac{1}{2} + \Delta \right)$$

or

$$P[I = AA] = p_A^* \times 1 + (1 - p_A^*) \left( \frac{3}{4} + \Delta \right)$$

where $\Delta$ corresponds to the change in the frequency of allele $A$ due to directional selection for HI in the development of progeny inducers for $I$ descending from a cross of type 1 or 2 (Eqn. (1)) and a cross of type 3 (Eqn.(2)).

Null hypothesis and alternative hypothesis

The biological hypothesis that allele $A$ is selectively neutral, corresponds to the null hypothesis $H_0$: $\Delta = 0$, whereas the alternative hypothesis $H_1$: $\Delta > 0$ corresponds to the statement that allele $A$ was selected for and, as a result, its frequency increased.

Test for selection of allele $A$ at a specific locus

For a specific segment detected in the inducers, the frequency $p_A^*$, which corresponds to the probability that a randomly chosen non-inducer carries this haplotype, and can be directly obtained from Table 1. Thus, using Eqns. (1) and (2), the probability of observing genotype $AA$ in a newly developed inducer at the locus under investigation is given by the expression

$$f(\Delta) = \left[ p_A^* \times 1 + (1 - p_A^*) \left( \frac{1}{2} + \Delta \right) \right]^{11} \left[ p_A^* \times 1 + (1 - p_A^*) \left( \frac{3}{4} + \Delta \right) \right]^{8}$$

By solving the equation $f(\Delta) = \alpha$, we obtain the lower limit $\Delta_\alpha$ of the $(1-\alpha)$% Clopper-Pearson confidence
interval (Clopper and Pearson 1934), corresponding to a statistical test of $H_0$ at the significance level $\alpha$. If $\Delta_u > 0$, we reject the null hypothesis $H_0$ based on our experimental data, indicating there is a positive selection at this locus; otherwise, we accept the null hypothesis, indicating that allele $A$ is selectively neutral.

In this study, we used the significance level $\alpha=0.01$ and 0.001 and the Bonferroni adjusted multiple testing significance level $\alpha=0.001$ and 0.0001 for the top 10 segments with the highest CHE score (Table 1).