

Supplementary data file S2:

Influence of haplotype diversity on GIC

One of the factors that may influence the genotypic information coefficient (GIC) is the composition of haplotypes within the population. We performed a small simulation study to determine the effects (if any) that the haplotypic make-up of a mapping population can have on GIC and, ultimately, on the power of QTL detection in such populations.

Simulation description

We compared the mean GIC across populations derived from a founder pool of 5, 10 or 20 haplotypes (H5, H10 and H20). Each founder haplotype was a randomly-simulated string of SNP positions (0/1) such that the minor allele frequency (frequency of “1” across all haplotypes) did not exceed 0.5. With these haplotype pools we simulated ten founder tetraploid individuals by randomly sampling four haplotypes (repeats allowed). Each individual carried a single chromosome of length 100 cM (centromere at 50 cM) with 101 marker positions (0 cM - 100 cM) at 1 cM spacing.

The founder individuals were then randomly mated using the simulation software PedigreeSim (Voorrips and Maliepaard, 2012) to give a starting population of 200 individuals which underwent 10 generations of random mating (each generation had 200 individuals). Finally, five F_1 mapping populations for each of the sets (H5, H10 and H20) were generated by randomly choosing pairs of individuals from the final generation and generating an F_1 population of 200 individuals from them. The resulting marker dosages were used as input for TetraOrigin (Zheng et al, 2016), which was run using the default settings (also described in the main text) and a bivalent pairing model only (no quadrivalents contributed to the simulated meioses in PedigreeSim). Only segregating marker data was used as input for TetraOrigin *i.e.* all non-segregating markers (0x0, 0x4, 4x0 or 4x4) were filtered from the datasets before generating the TetraOrigin input files.

The resulting IBD probabilities were used to calculate the genotypic information coefficient (GIC) along each homologue per population as described in the main text. Finally, an overall mean GIC per simulation set (H5, H10 and H20) was calculated.

Simulation results

We visualised the mean GIC per simulation set, which showed a slight increase in the GIC as the number of founder haplotypes increases (Figure S2.1.a). However, although we simulated all datasets to have equal numbers of markers, we found that a large proportion of markers were non-segregating in the populations derived from a small number of founder haplotypes (H5), which we screened out before running TetraOrigin. This was because in a normal situation, a linkage map is provided to TetraOrigin, for which only segregating markers can be used. In any case, there is no inheritance information that TetraOrigin could have used from these “dud” markers. The proportion of non-segregating (N.S.) markers was not equal across the different population types (Figure S2.1.b), although the marker segregation type breakdown among the remaining marker types (which are presented in their fundamental / converted form) was relatively consistent otherwise.

Therefore, it appears that the number of founder haplotypes does influence the GIC (and therefore also the power of a QTL detection study ultimately), but that this probably arises due to the reduction in the number of segregating markers available in such material.

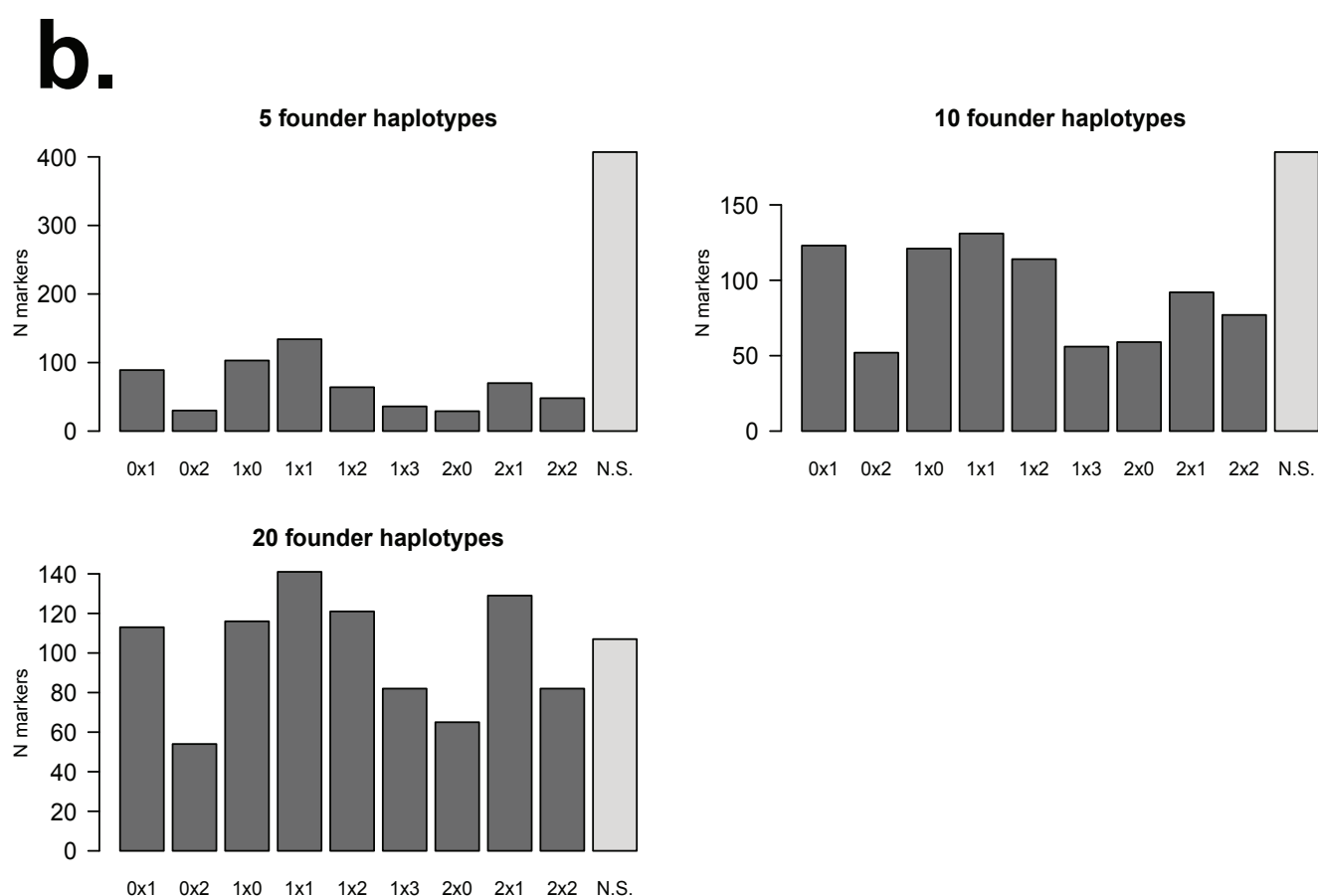
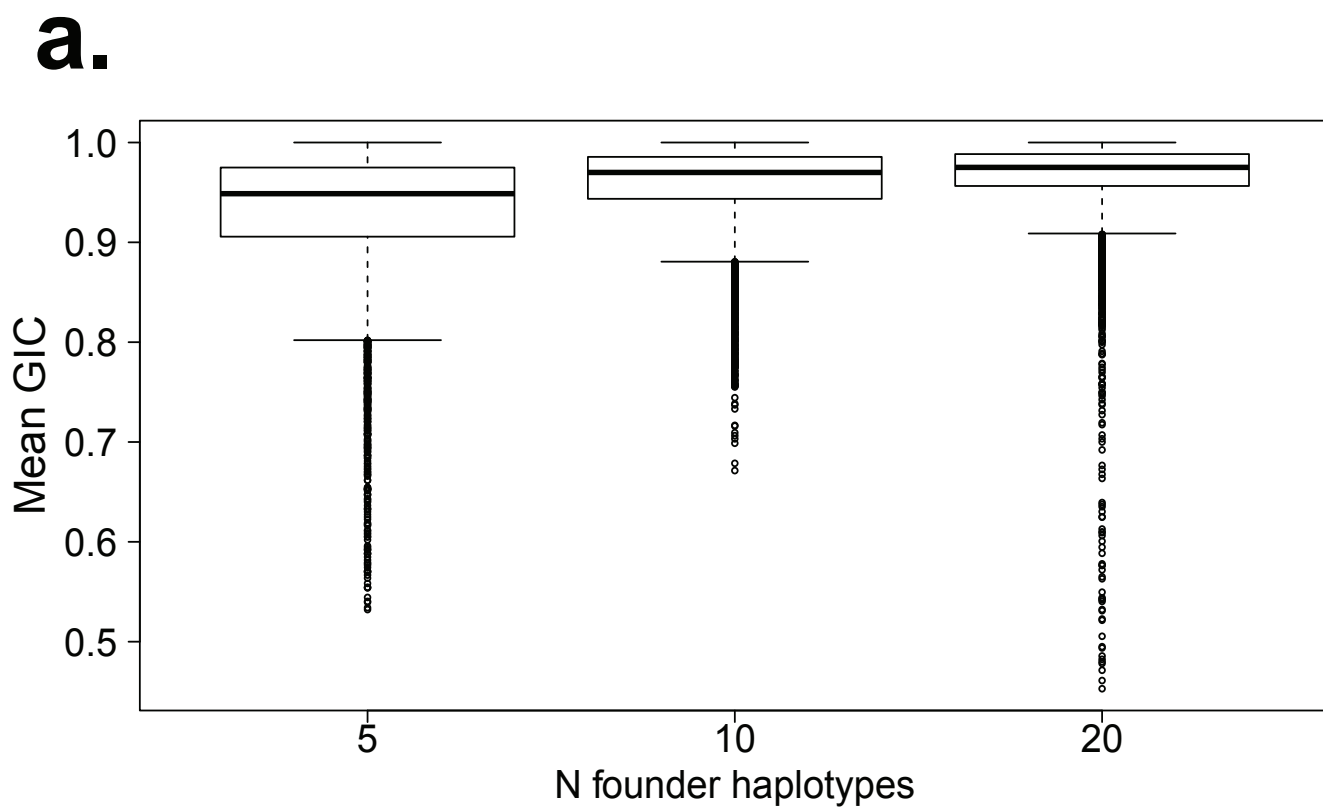


Figure S2.1 a. Boxplot showing mean GIC across all homologues of five populations simulated from different pools of founder haplotypes. On average, the GIC was found to be slightly higher from more haplotypically-diverse populations, although the influence of marker density (specifically, the numbers of non-segregating markers) was not corrected for.

b. Pooled marker segregation breakdown in each simulation set. N.S. refers to non-segregating markers (0x0, 0x4, 4x0 and 4x4) which were removed from the datasets before further processing with TetraOrigin. Populations derived from relatively few founder haplotypes were more likely to carry fully-homozygous markers, which is problematic for mapping if present in both parents.