Supplementary Material:
Comparison of alternative local kinship estimators

At any locus, among 31 individuals there are 496 (local) kinship values, $\{\phi_{ij}, 1 \leq i \leq j \leq 31\}$ including the self-kinship values ($i = j$). All IBD estimates were based on the same data on 10188 SNP markers observed on 31 individuals. We computed linkage statistics at a subset of 334 marker positions, and so compare IBD estimates at these same 334 positions across the 200 cM chromosome.

From our analyses we have five sets of 496 local kinships:

1. The true values from the simulated IBD.
2. The average value over 200 realizations from *ibd_stitch*
3. These values constrained to the allowed values, 0, 1/4, 1/2 and 1.
4. The values estimated from allelic identity estimated over windows of approximated 100kb, and then smoothed over windows of 1 to 2 cM.
5. These values constrained to the allowed values, 0, 1/4, 1/2 and 1.

For clarity we refer to the estimators (2) and (3) deriving from *ibd_stitch* as IBD-based, and estimators (4) and (5) based on allelic similarity as IBS-based.

The correlation of each set of the estimated values (2)-(5) with the true value (1) is shown in the following figure:

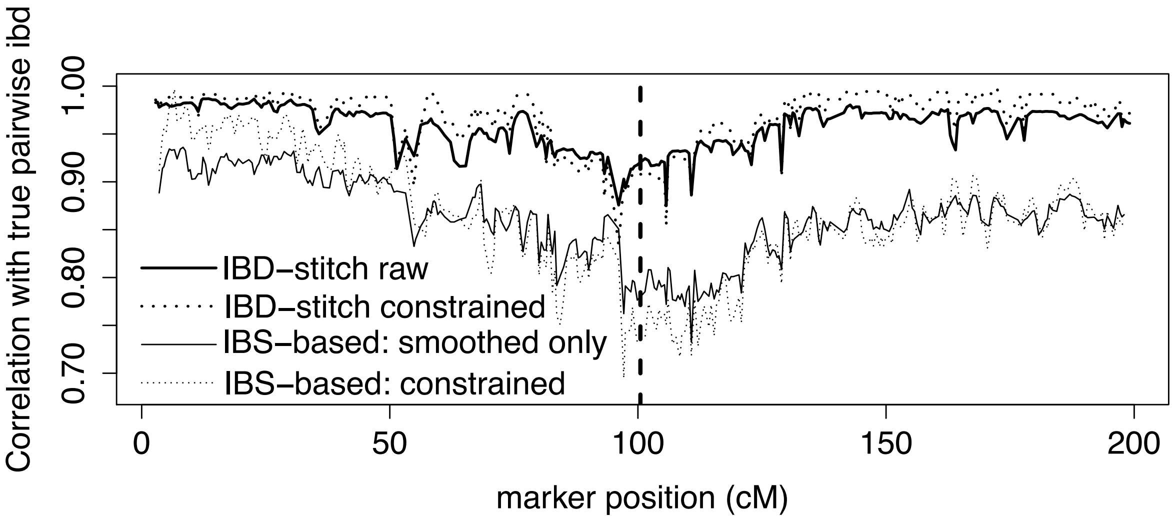

Figure S1: Correlation of pairwise IBD estimates with true values

We considered also the bias and root-mean-square-error of these four estimators across the 334 test marker positions with the same qualitative conclusions (Data not shown).

Overall, the correlations deriving from *ibd_stitch* are substantially higher than from the IBS-based estimates. At the left end of the chromosome there is little remote IBD and the constrained IBS estimator does well. At the right-hand end there is more remote IBD, but unrelated to the trait: *ibd_stitch* finds these small segments of remote IBD more accurately. In the region of the trait locus, there are high levels of IBD among multiple individuals including a total of 401 pairwise kinship values of $\phi = 1$ (State-1 of Table 1). All methods perform worse in this area, but the IBS methods fail to detect these high IBD values, and constraining values to $H = \{0, 1/4, 1/2, 1\}$ actually slightly decreases the correlation with the simulation truth. Generally, constraining the estimates to $H = (0, 1/4, 1/2, 1)$ shows little impact and not always improvement, although this might be different in an approach such as that of [9] in which smoothing and constraining are accomplished in a single optimization.

Among 31 individuals there are 496 kinship values, ϕ_{ij} including the self-kinship values ($i = j$). Over the 334 markers at which values are scored there are thus 165,664 values. Under the true IBD configuration, there are, 123,801, 27,949, 13,513 and 401 values of 0, 0.25, 0.5 and 1, respectively.

The following boxplots show the distribution of the non-constrained estimators: IBD-based (2) and IBS-based (4):

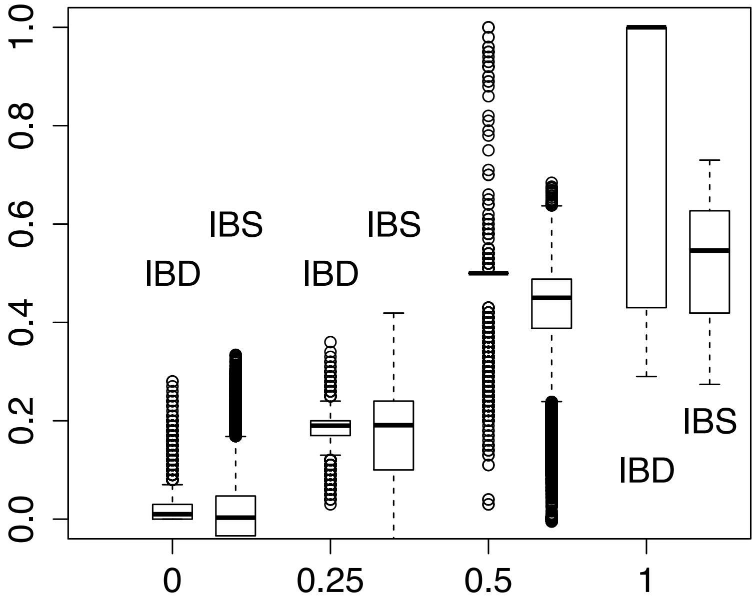

Figure S2: Comparison of estimators at each true local kinship value

In the figure, IBD indicates the IBD-based estimate (2) from *ibd_stitch*, while IBS indicates the estimator (4) based on allelic similarities.

At each of the four true values the estimates based on IBD graphs show a more compact distribution, although for the true value 0.25 there is a wide range of outliers. The distributions based on the pairwise IBS approach are biased downward at 0.5, while this approach fails to find any of the 401 true values of $\phi = 1.0$.