Supporting Information

Algorithm to merge adjacent regions sharing 10Mb haplotypes: Within a chromosome, compare a pair of regions identified by ≥ 10Mb IBD sharing; if they overlap and at least 3 individuals are common in the two regions, combine them into one region that spans the full length of both. Two lists of individuals are kept for the new combined region, one with those sharing the entire region and another with the union of individuals who share the original 10M haplotypes being combined.

Germline/Dash Analysis of Palau Data. The Palau data were analyzed using the Germline and Dash programs. Analysis was done per chromosome. The steps:

1. Use phased haplotypes from Beagle to create phased input files for Germline. Only cases and controls were used.
2. Run Germline using the settings recommended in the Dash documentation.
3. Use output of Germline (.match files) to run Dash. Settings were for sliding window size of 10M (note that resulting segments can be much longer), and a minimum of 7 haplotypes to form a cluster. Default values were used for the minimum cluster density (.6) and the r^2 (.95).
4. Collected Dash cluster files (.hcl) into a single file, with columns added for counts of cases and controls for each cluster.
5. Compare to our simpler algorithm.

Estimating kinship between pairs of individuals.

In expectation, the relatedness of two individuals is defined as the total length of all IBD segments divided by twice the length of the genome. To estimate IBD segments from pairs of subjects, we used a technique called IBD_{half}. After implementing this algorithm for all pairs of subjects and across their genomes, estimated relatedness values were collected into a relatedness matrix L' with elements L_{ij}' for subjects i and j. Due to potential errors in identifying IBD segments, as well as Mendelian sampling, L' is often noisy, over- and underestimating relationships. To eliminate some of this noise we applied treelet covariance smoothing [58,59] to L' to produce a refined L.

Number of meioses separating subjects in a pedigree. Counting the number of meioses separating the affected subjects from a pedigree is straightforward. The complex inter-relationships among Palauans, however, made simple counting impossible in many instances and thus this number was approximated. To predict the number of meiosis (M) we relied on the following three measures of the density of the pedigree among the cases; 1) number of cases to consider (N), 2) average relationship among the N cases × 100 (aveRel) and 3) the average of the maximum relationship of individual i with all other individuals in N × 100 (aveMax). To determine the prediction equation we used the 6 pedigrees of Figure 1 in Knight et al., 2012, augmented with two pedigrees with 11 cases from the Palau dataset. The resulting regression was $M = 10.63 + 2.09 \times N - 0.65 \times aveRel - 0.13 \times maxRel$, with an R^2 for the model of 99.8%. All effects were significant: N and aveRel at 0.001; and maxRel at 0.01. Increasing the number of cases (N) will

increase the number of meiosis while for individuals who are more closely related than average it will decrease. When adding a case to a set, the number of meiosis is expected to increase by 2. When the average relationship among the group of cases increases by 10% one expects the number of meiosis to decrease by 6.5. Similarly, when the average closest relative is 10% more closely related the number of meiosis decreases by 1.3.

Details of the fit for the 8 families are in the following table.

Supplementary Table 2. Prediction of the number of meioses separating subjects in a complex Palauan pedigree.

Pedigree	True M	N	aveRel×100	maxRel×100	predM
Fig 1 - meiosis 15	15	8	10.1	42.2	15.4
Fig 1 - meiosis 16	16	7	8.4	32.4	15.6
Fig 1 - meiosis 21	21	6	2.3	5.2	21.0
Fig 1 - meiosis 22	22	9	6.5	26.7	21.8
Fig 1 - meiosis 25	25	8	2.5	4.5	25.2
Fig 1 - meiosis 32	32	13	4.4	23.1	32.0
Palau - fam A	22	11	11.0	34.4	22.0
Palau - fam B	25	11	9.2	21.2	25.0

Reassuringly, not only is the fit of the model excellent, the estimated M is highly correlated with the number of affected subjects per haplotype cluster ($r = 0.70$) in Supplementary Table 1, as expected.

Inference. Because variants separated by substantial genomic distances can be in substantial LD and interrelationships amongst subjects are complex, it is challenging to interpret the significance of N affected individuals sharing a long haplotype in the Palau

population. Counting the estimated number of recombinants amongst all affected subjects should be a piece of information, as described by refs [27,28], but a simple threshold for significance seems unlikely given some of the complex LD patterns we report here. Our approach was to calculate standard LOD scores for clear-cut pedigrees and use that well established framework for inference. We do not assign a p-value for other clusters of affected individuals, from whom pedigree relationships are less apparent.