**IBD segment analysis and Total Information Potential**

The pedigree of 2,906 Kosraen individuals was divided into three groups without replacement: two parents and a single child (trio), a single parent and a single child (duo), and single samples (unrelated). Using the BEAGLE framework (BROWNING and BROWNING 2009), the individuals were phased and missing data inferred taking into consideration their respective group structure. The phased genotype data was processed with GERMLINE under default parameters and with genetic distance annotation data corresponding to the Affymetrix 500k chip to generate the genotype-based IBD shared segments. The same data was additionally processed with GERMLINE under the phase-specific haplotype-extension parameters, which explicitly treats each homolog separately in generating matches.

The INFOSTIP analysis was performed on both genotype and haplotype oriented IBD segments. For haplotype data, INFOSTIP executed upon each homolog as if it were an independent set of shared segments, but in choosing a sample for the sequence panel excluded all of the matches originating from that individual on either homolog. As such, a site must be either autozygous, or contained within an IBD segment of two differing sequenced individuals to be fully inferred. For genotype data, INFOSTIP ran with no modification and hence, a site is considered fully inferred if either homolog is in IBD with a sequenced individual. Because the imputed regions were SNP-chip oriented, the total cohort genome length was calculated as the individual end-to-end length of the genome that contained SNPs, multiplied by the number of samples (for genotype data) or twice the number of samples (for haplotype data). FIGURE S1 shows a comparison of the two inference techniques (haplotype, genotype) as well as the two selection methodologies (greedy, random). Due to the non-negligible presence of some autozygosity within the cohort, these two distributions represent an upper and lower bound on the imputation capacity.