

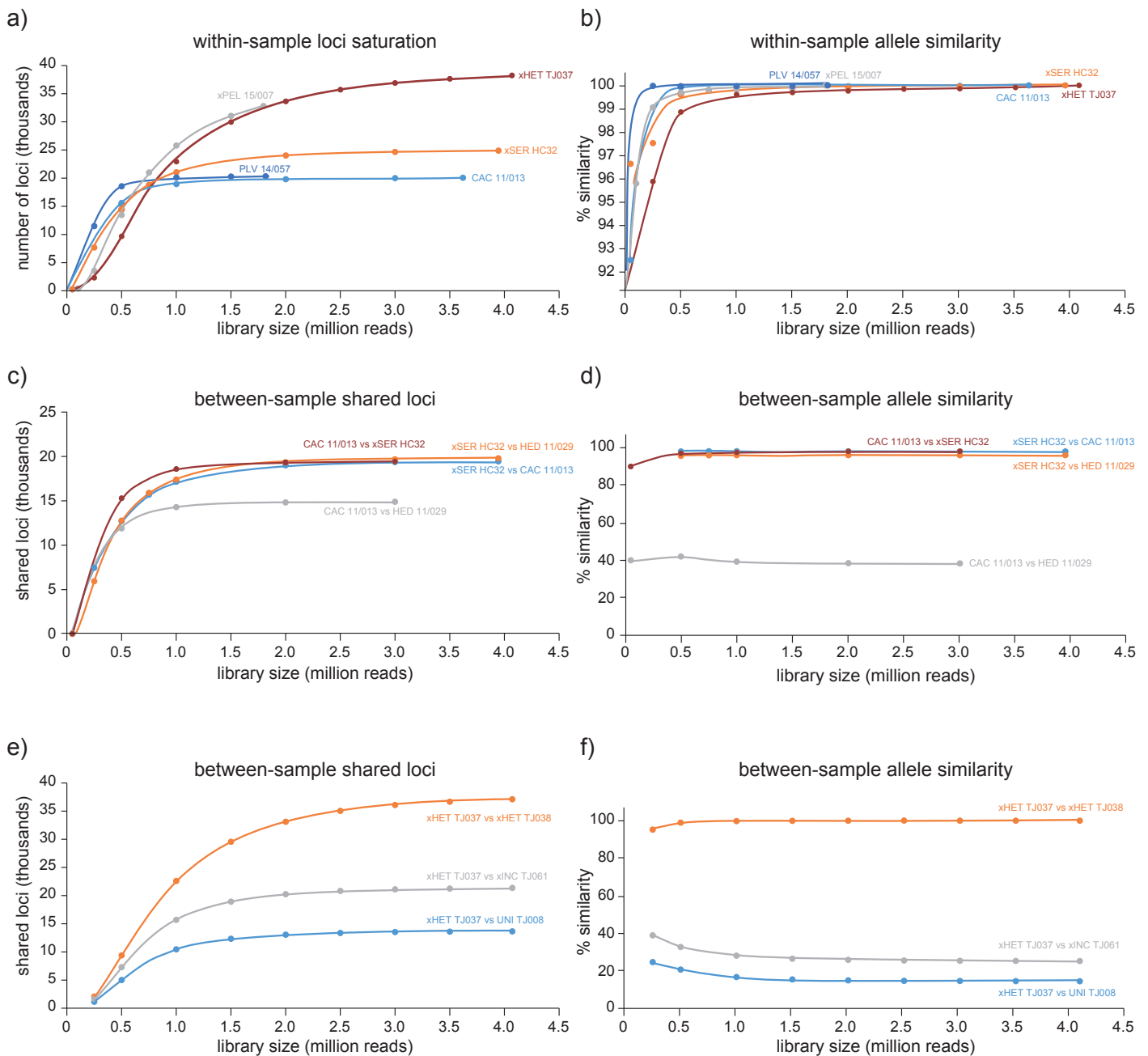


Figure S2

Saturation curve analysis of reads after preprocessing.

a) and b) saturation of the number of loci and the allele similarity for two diploid species (*P. cactorum* and *P. plurivora*), two diploid hybrids (*P. xserendipita* and *P. xpelgrandis*) and a polyploid hybrid (*P. xheterohybrida*), calculated by comparing the number of loci and/or allele similarity of subsampled datasets at increasing numbers of reads against all reads from that sample.

c) and d) saturation of the number of shared loci and the allele similarity between samples of the diploid progenitors *P. cactorum* and *P. hedraiaandra* and between them and their hybrid *P. xserendipita*, calculated by comparing the number of shared loci and/or allele similarity of subsampled datasets of the first sample at increasing numbers of reads against all reads of the second sample.

e) and f) saturation of the number of shared loci and the allele similarity between samples of species with a more complex genome, i.e. polyploid hybrids *P. xheterohybrida*, *P. xincrassata* and the diploid species with a large genome size (330 Mbp) *P. uniformis*, calculated by comparing the number of shared loci and/or allele similarity of subsampled datasets of the first sample at increasing numbers of reads against all reads of the second sample.