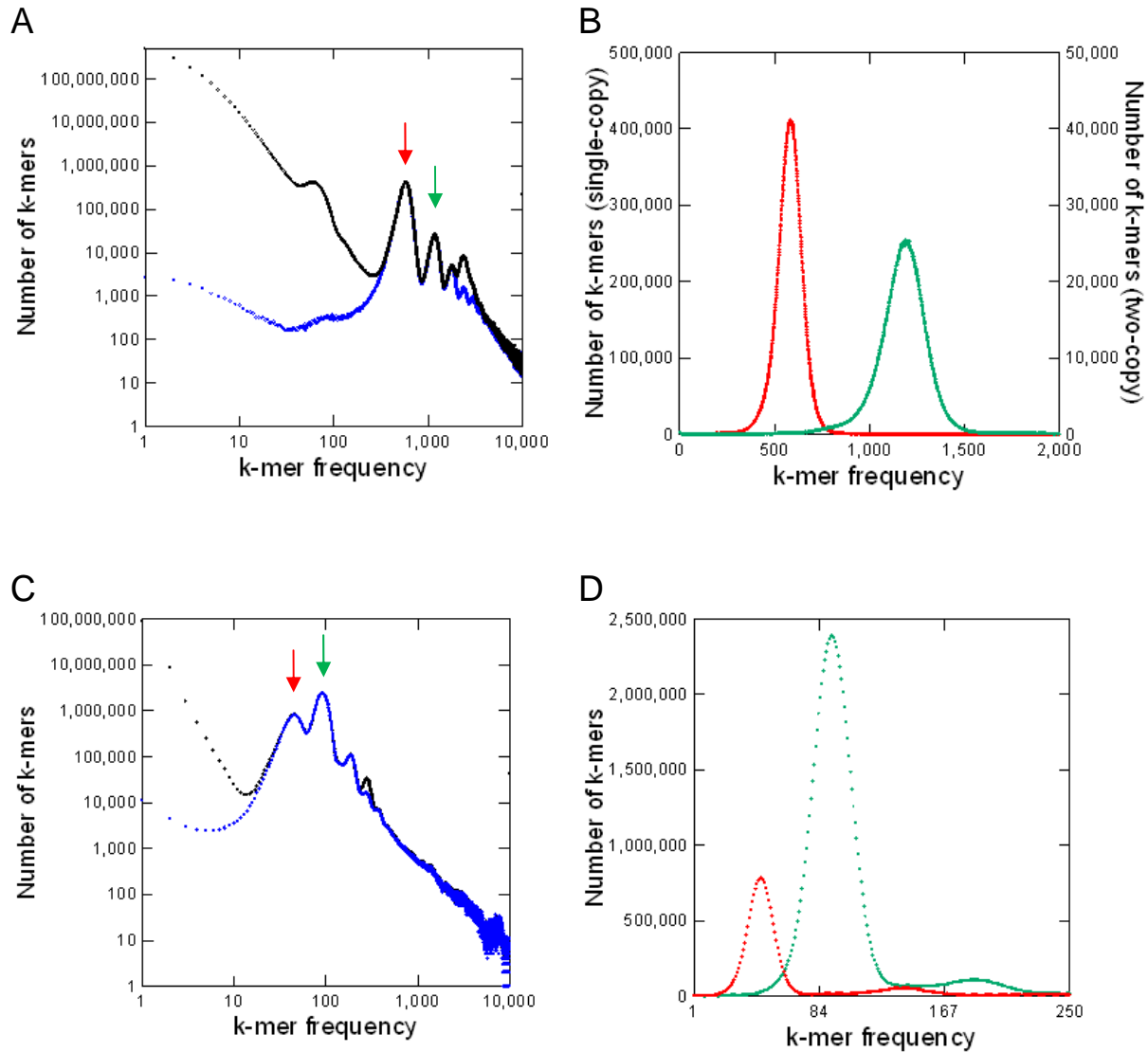


Supplemental Fig. S5



Supplemental Figure S5. Fine separation of k -mers with different copy numbers with Illumina sequencing. Panel A: frequency distributions of all k -mers (black) and all valid k -mers (blue) from *C. elegans* Illumina reads. The two arrows mark the peaks shown in panel B, which shows the distribution of valid single-copy k -mers (red) and valid k -mers with two copies in the genome (green). Panel C: frequency distribution of all k -mers (black) and all valid k -mers (blue) from *D. melanogaster* Illumina reads. The two arrows mark the peaks shown in panel D, which shows the distribution of X or Y-linked k -mers (red; peak at 46x) and autosomal k -mers (green; peak at 93x). Note that both in *C. elegans* and *D. melanogaster* datasets the two distributions are well separated. The lists of valid, single-copy, two-copy, sex-linked and autosomal k -mers were obtained from the reference genomes. The *D. melanogaster* reference genome still have gaps, but the data shown should be very similar to the true distributions because most of the gaps correspond to highly repetitive sequences, which contribute relatively little to the number of distinct k -mers.