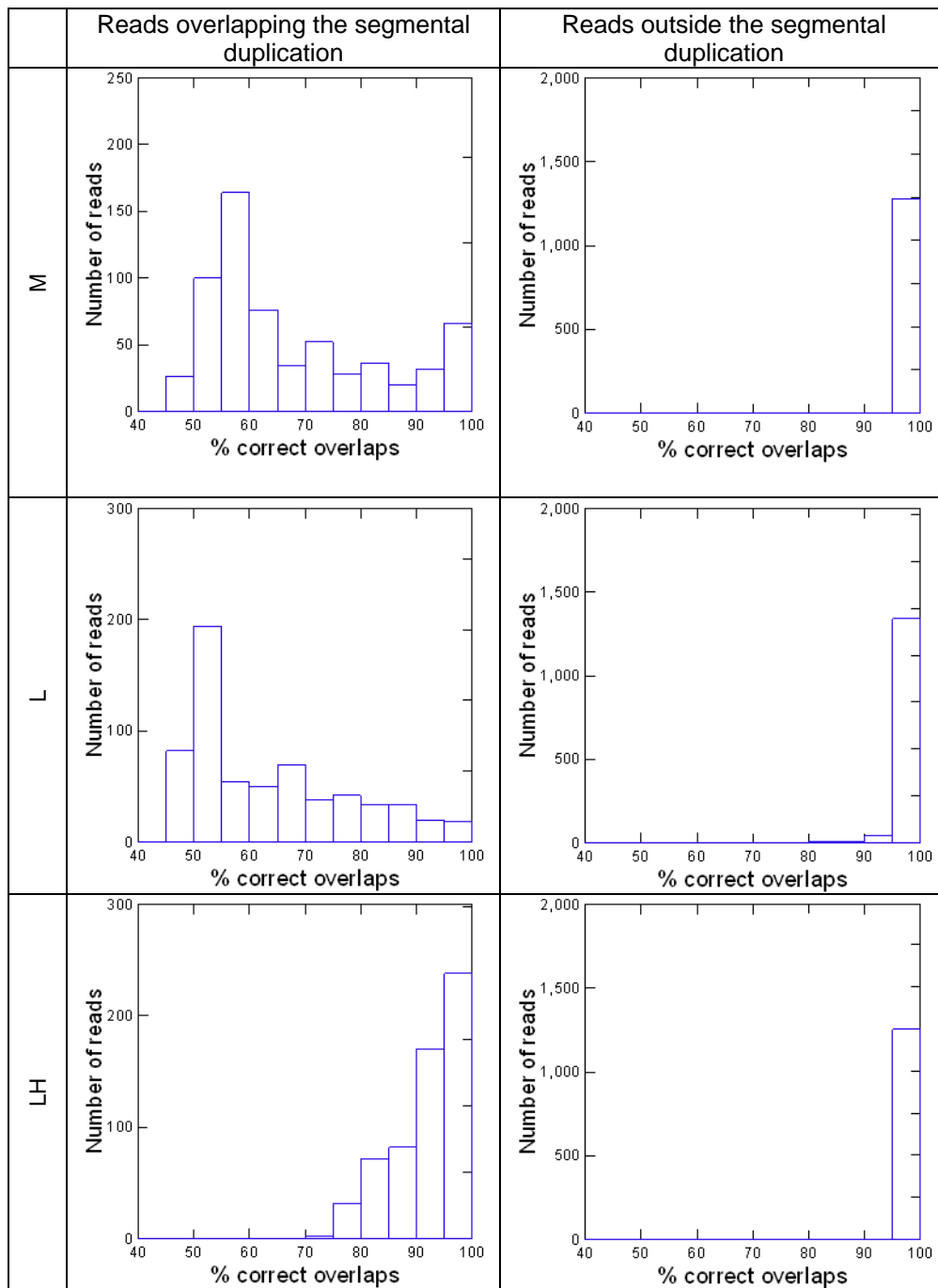


Supplemental Fig. S8



Supplemental Figure S8. Overlap detection accuracy in the "model genome" (a human segmental duplication). The MHAP overlacer was run alone with the uncorrected PacBio synthetic reads (and the same parameters used in the full assembly), and its output was saved as a text file. The figure shows data from three types of run: standard MHAP ("M"), MHAP with L-masking ("L"), and MHAP with LH-masking "LH"). For each read we scored the proportion of overlaps that are correct, *i.e.*, between two reads from the same contig ("left" or "right"; see Methods). Reads with at least one base-pair overlapping the segmental duplication region were conservatively labeled as "Reads overlapping the segmental duplication"; the remaining reads were labeled as "Reads outside the segmental duplication". Note that only LH-masking leads to correct overlaps in reads overlapping the segmental duplication; standard MHAP and L-masking basically do not distinguish the two copies of the segmental duplication (*i.e.*, frequently have ~50% correct overlaps). The grand means for reads matching the segmental duplication are 68% (MHAP), 64% (L), and 92% (LH). For reads not matching the segmental duplication it was above 99% in the three cases.