A seed $k$-mer is sorted in msBWT to create a de Bruijn graph, which is then simplified to create linear contigs.

B Phase-informative reads are used to determine the correct phase of the contigs.
Figure S1. Targeted de novo assembly using the multi-string Burrows-Wheeler Transform (msBWT). (A) The msBWT and its associated FM-index implicitly represent a suffix array of sequencing reads, such that read suffixes sharing a k-mer prefix are adjacent in the data structure. This allows rapid construction of a local de Bruijn graph starting from a k-mer seed (dark blue) and extending by successive k-mers (light blue) containing the (k-1)-length suffix of the previous k-mer. A (k-1)-length prefix with more than one possible suffix (red and orange) creates a branch point. Adjacent nodes in the graph with in-degree and out-degree one can be collapsed into a single node, yielding a simplified graph, which can then be traversed to obtain linear contig(s). (B) Paralogs of R2d can be disentangled using the local de Bruijn graph by exploiting differences in copy number. Edges in the graph are weighted by read count, and linear contigs for the R2d1 and R2d2 paralogs obtained by traversing the graph in a manner that minimizes the variance in edge weights along possible paths. Phase-informative reads (those overlapping multiple paralogous variants) provide a second source of evidence.