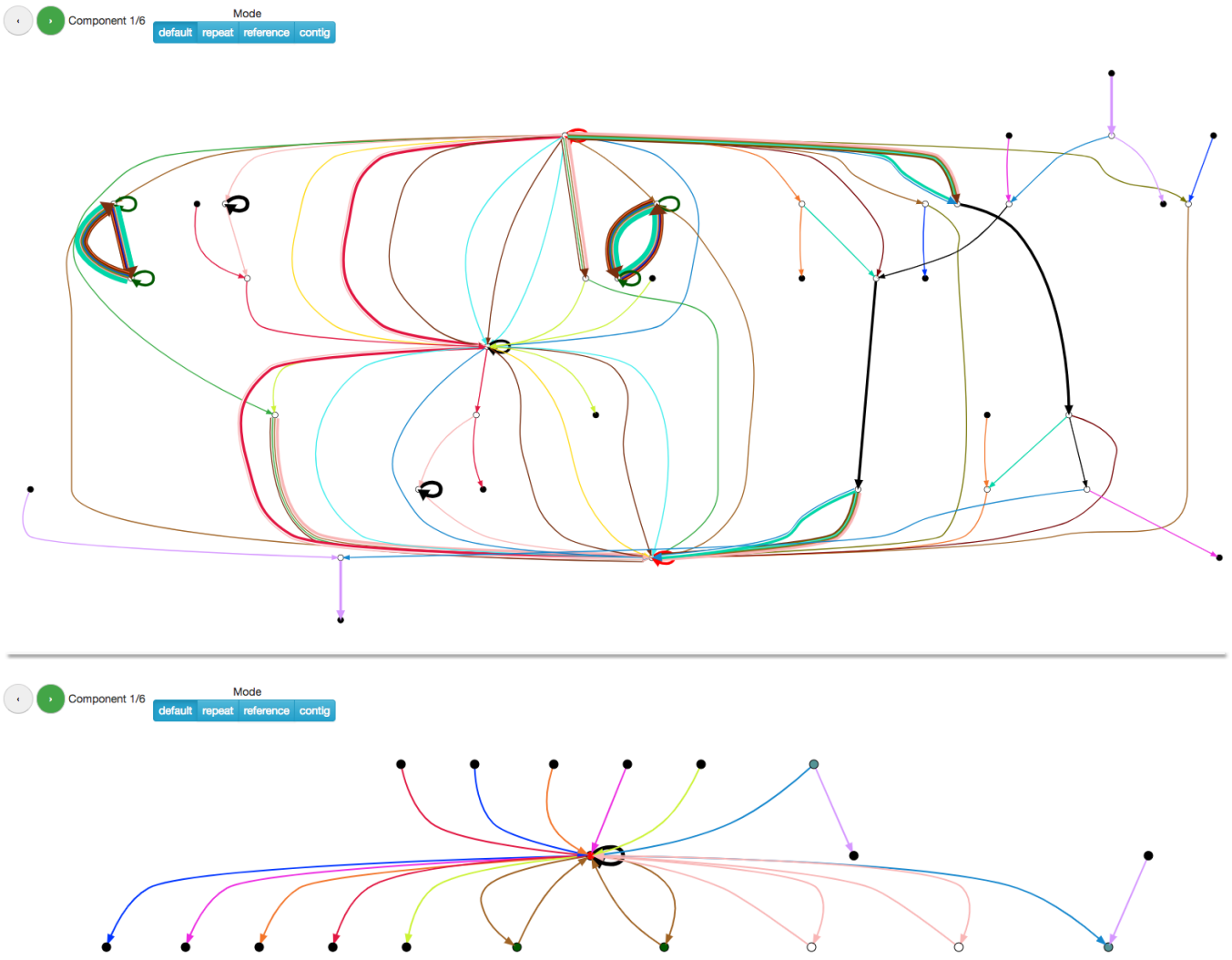
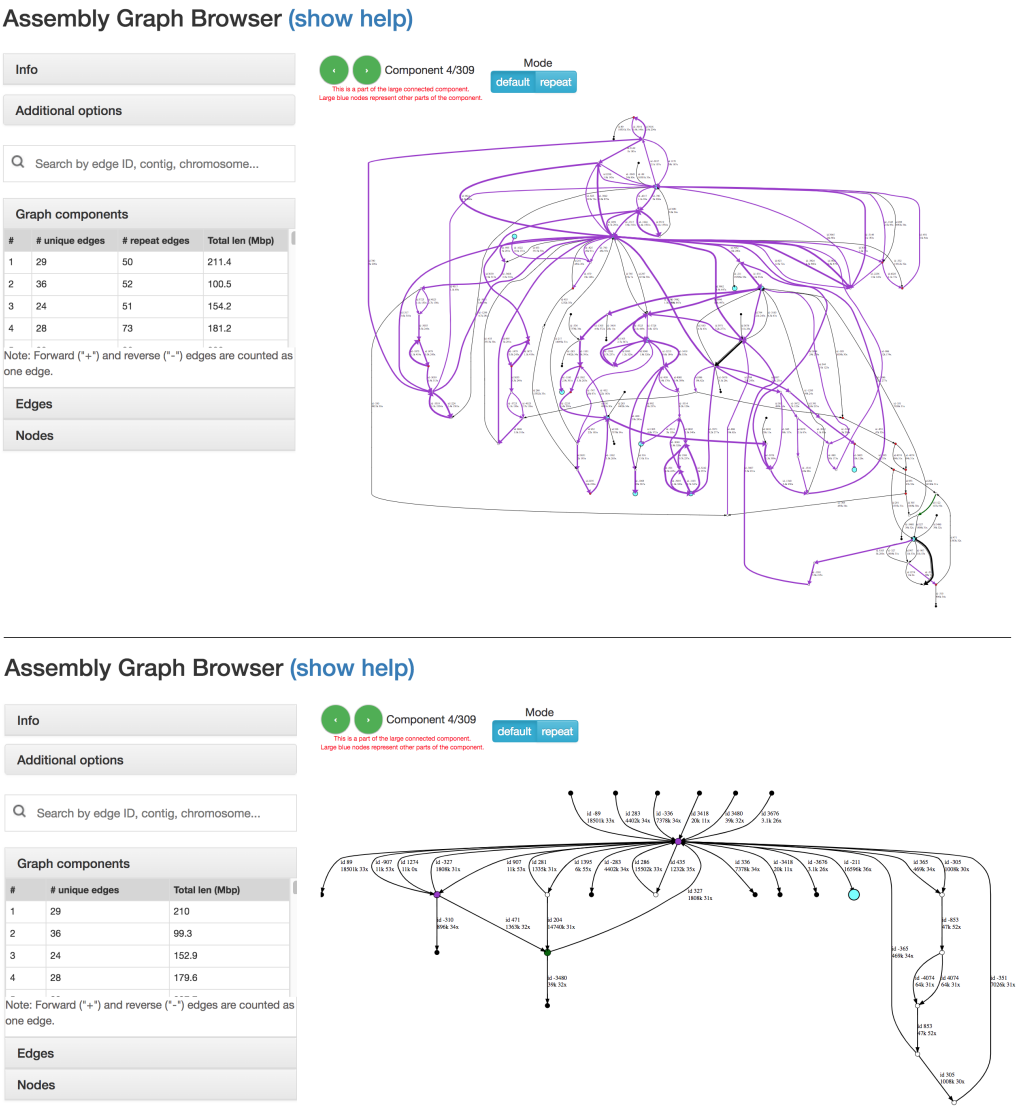
Supplementary Material for "Assembly Graph Browser: interactive visualization of assembly graphs"

1 Assembly Graph Browser for the *S.cerevisiae* dataset

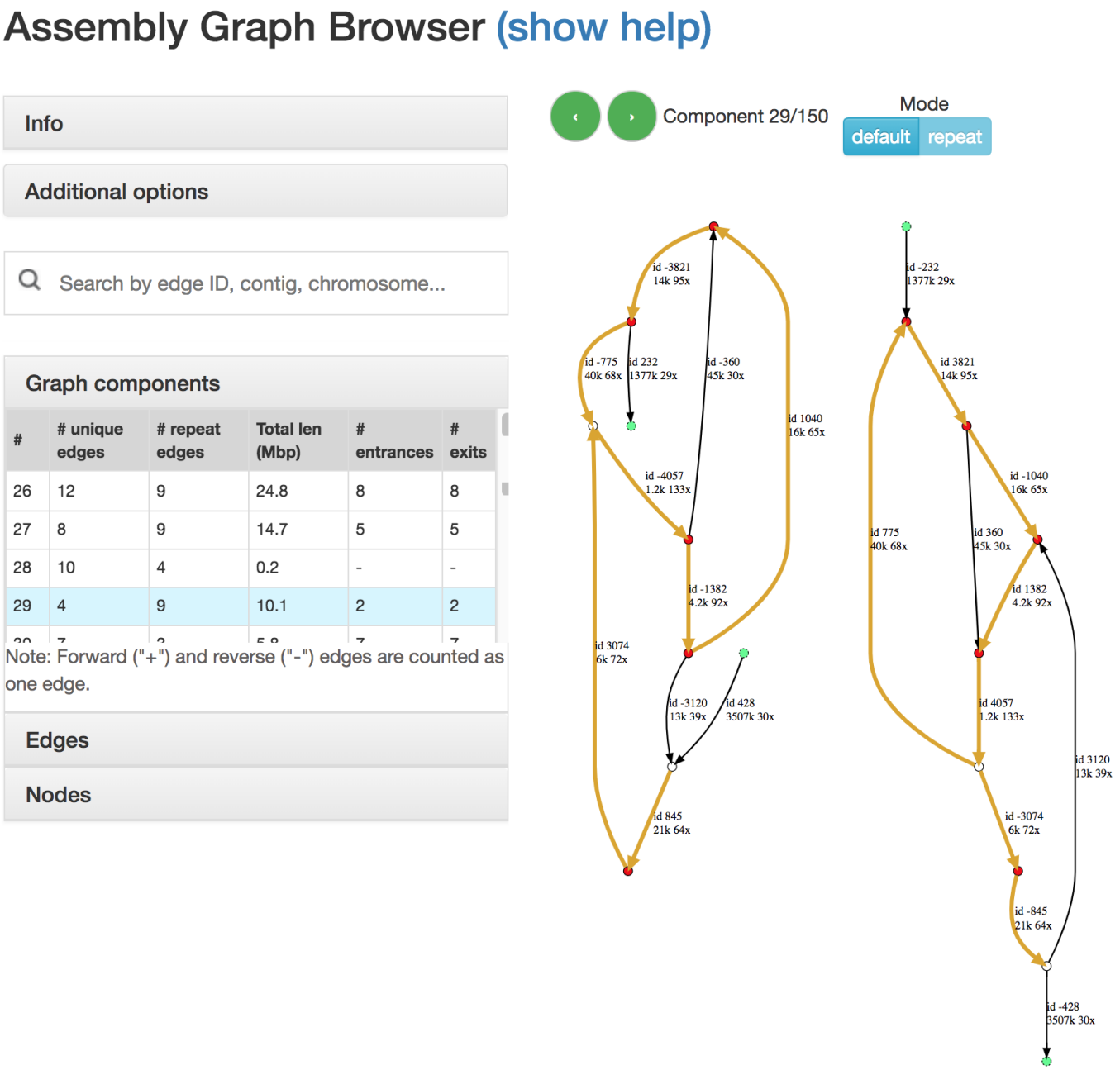


Supplementary Fig. S1: An example of the Assembly Graph Browser for the assembly of the *S.cerevisiae* dataset generated with the Flye assembler. (Top) Default view with the edges colored according to threir chromosome mappings (in case of ambigious mappings edges have multiple colors). Unique edges are shown by thin lines, and repetitive edges are shown by thick lines. The edge labels were hidden to simplify the overview. (Bottom) The contracted assembly graph produced by collapsing all repetitive edges. The contracted graph is less complex and reveals the chromosome structure. Each mosaic repeat component was collapsed into a single node that can be expanded by double-clicking.

2 Assembly Graph Browser for the *H.sapiens* dataset

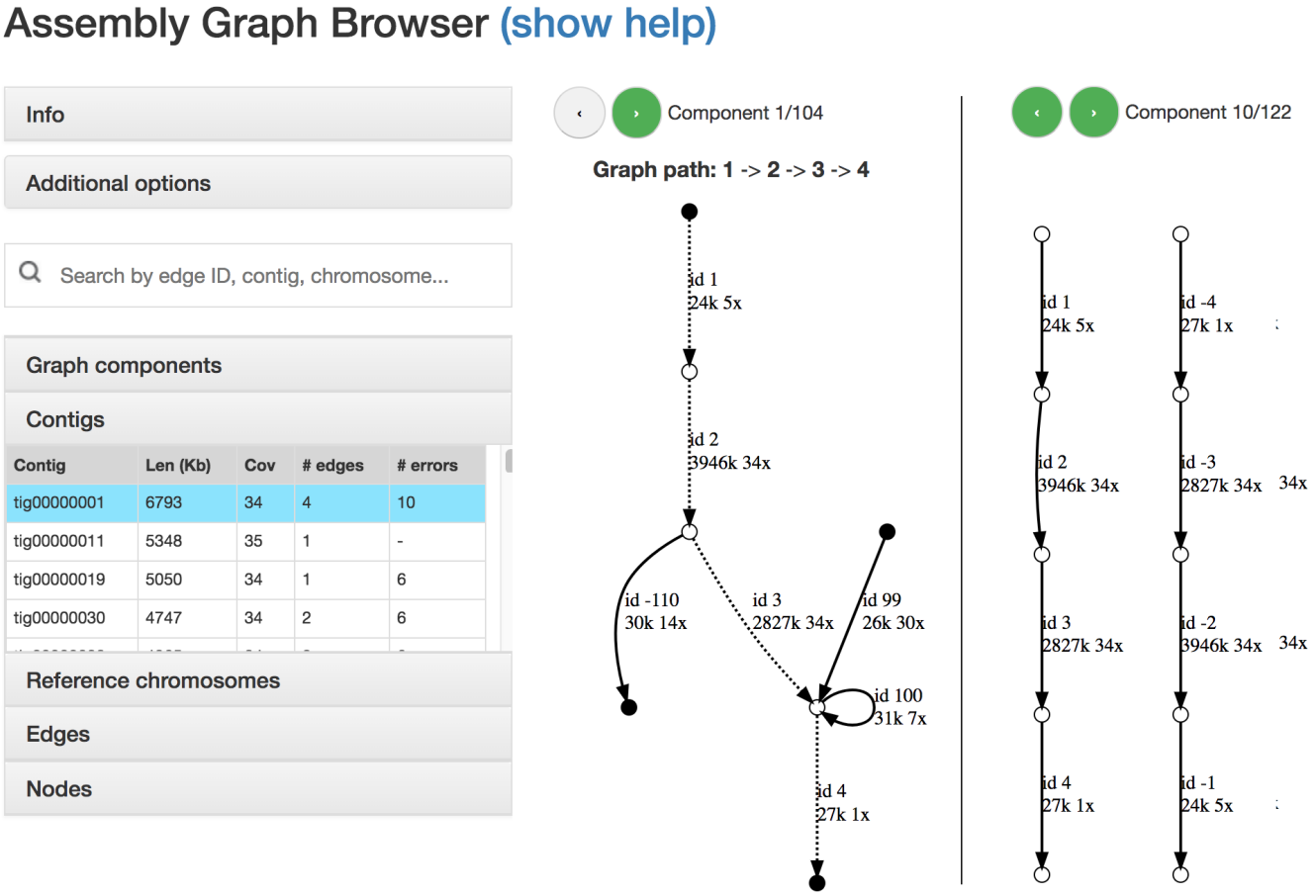


Supplementary Fig. S2: A subgraph of the assembly graph of the *H. sapiens* assembly generated with the Flye assembler. The full graph was partitioned into 309 components. (Top) The table of connected components in the left panel shows the number of edges and the total length of each component, which allows a user to select the component of interest. (Bottom) A subgraph produced by collapsing repeat edges. The table of nodes in the left panel shows indegree and outdegree for each node and the number of hidden repetitive edges.

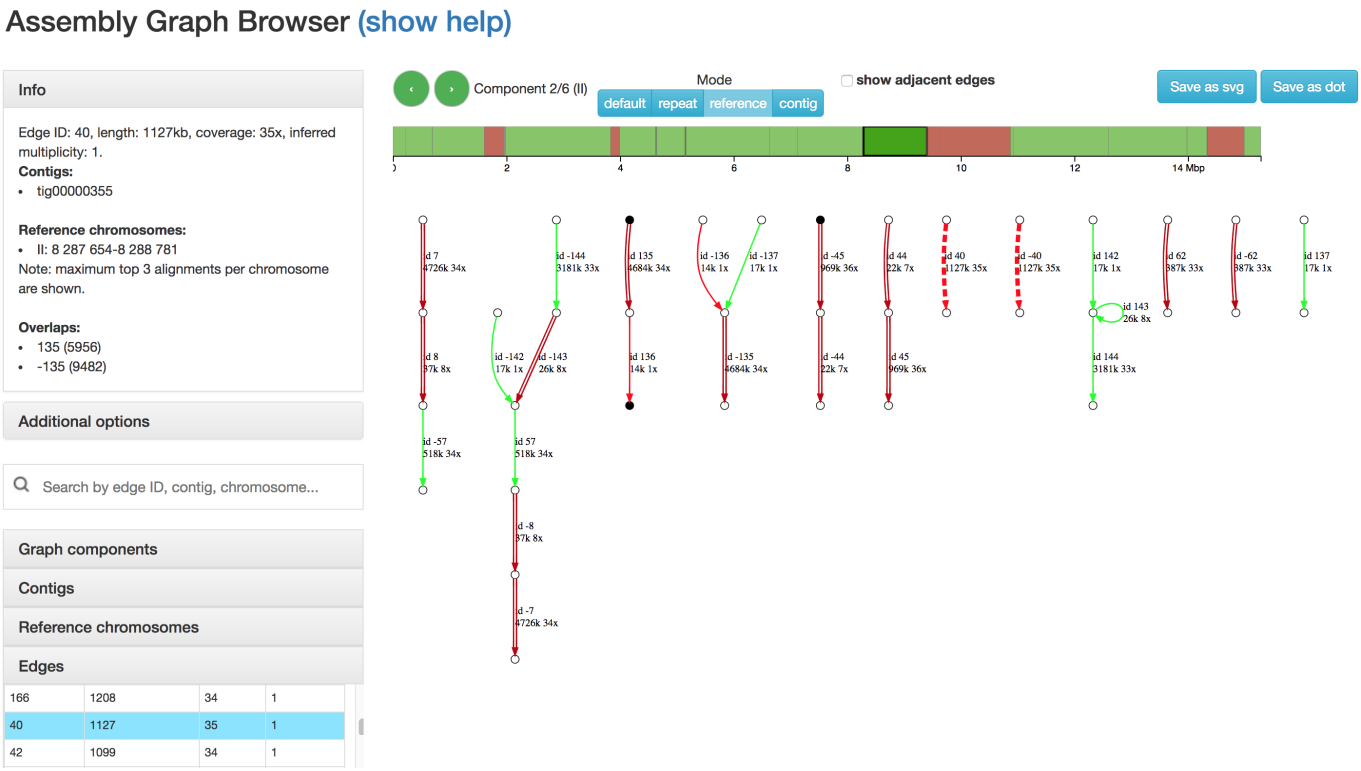


Supplementary Fig. S3: An example of a mosaic repeat cluster from the Flye assembly graph of the *H. sapiens*. It corresponds to a segmental duplication formed by 7 repetitive edges of length 1.2 kb - 40 kb. The unique edges adjacent to this mosaic repeat (shown in black) are included to preserve the graph structure. Light green nodes show the connections to the hidden parts of the graph. Four unique edges correspond to three entrances and three exits (two edges serve as both entrances and exits). Thus, this segmental duplication corresponds to three contiguous segments of human chromosomes.

3 Assembly Graph Browser for the *C.elegans* dataset



Supplementary Fig. S4: The assembly graph of the *C.elegans* generated with the Canu (Koren *et al.*, 2017). (Left) The connected component formed by a contig and adjacent edges, contig path is shown by dashed lines. (Right) Contig-based mode: only edges that belong to the tig00000001 contig are displayed. Total length, mean coverage and the number of edges for each contig are displayed in the table in the left panel.



Supplementary Fig. S5: An example of the reference-based mode with edges mapped to the chromosome II of the *C.elegans* genome (generated with the Canu assembler). Edges are colored according to the presence of assembly errors detected by QUAST-LG (Mikheenko *et al.*, 2018): green edges do not contain errors, red edges belong to the misassembled contigs (but correspond to correct genomic sequences), and dark red edges represented by parallel lines are erroneous. Edge with ID 40 is selected (highlighted with the dashed line). The information about edge length, coverage, inferred multiplicity, corresponding contigs, alignments to the reference genome, and overlaps with other edges is shown at the left. At the top, corresponding edge alignments to the selected chromosome are displayed. Red blocks contain detected assembly errors, while green blocks were aligned correctly. The alignment of the selected edge is highlighted with dark green color.

References

Koren, S. *et al.* (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.*, **27**(5), 722–736.

Mikheenko, A. *et al.* (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, **34**(13), i142–i150.