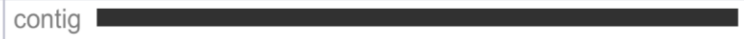
Supplemental Fig S1. Visualisation of BLAST evidence for lowest scoring 30 contigs according to each score component in Trinity assembly of yeast. To inspect the performance of the contig score components, we used the smallest of the assemblies from real data, the Trinity assembly of yeast from the Davidson & Oshlack (2014) dataset. Although contigs that score poorly on any given score component are likely to contain assembly problems, it does not follow that contigs scoring well on any given score component are well assembled, as they often score poorly on other components. Thus, the lowest scoring 30 contigs for each score component are visualised here with their supporting evidence.

$s(C_{nuc})$

$s(C_{nuc})$ measures how well the read evidence supports the identity of each base in a contig. A low score might indicate family collapse (Figure 1), which might be detectable as multiple reference transcripts aligning in an overlapping manner to a single contig. We aligned all contigs to both the reference transcriptome and the reference genome using blastn. Alignments were visualised so that a black bar represents the assembled contig, while coloured bars represent the aligned transcripts and/or genes. Each row in the stack of alignment represents a different gene, allowing the identification of cases where alignments overlap. The number of genes in the stack is limited to three. While many contigs received no hits at all - presumably due to the incompleteness of the reference, and biological novelty in the assembly - those that did receive hits usually received multiple hits, consistent with family collapse.

30 lowest scoring

	Component score	Plot
comp3423_c0_seq1	0.58	
comp309149_c0_seq1	0.59	
comp3111_c0_seq1	0.61	
comp3113_c1_seq1	0.65	
comp272622_c0_seq1	0.68	
comp24393_c0_seq1	0.69	
comp3600_c0_seq2	0.69	
comp3665_c1_seq1	0.70	
comp143163_c0_seq1	0.71	

comp129162_c0_seq1	0.71	
comp3009_c0_seq1	0.72	
comp3423_c0_seq3	0.73	
comp3562_c0_seq2	0.73	
comp156421_c0_seq1	0.74	
comp3640_c0_seq5	0.75	
comp111546_c0_seq1	0.75	
comp191305_c0_seq1	0.76	
comp1463_c0_seq1	0.76	
comp3603_c4_seq2	0.76	
comp3654_c0_seq16	0.77	
comp505_c0_seq1	0.77	
comp277323_c0_seq1	0.77	
comp162548_c0_seq1	0.78	
comp189983_c0_seq1	0.78	
comp157164_c0_seq1	0.78	
comp3637_c1_seq2	0.78	
comp201694_c0_seq1	0.78	
comp260862_c0_seq1	0.79	
comp256546_c0_seq1	0.80	
comp270985_c0_seq1	0.80	

$s(C_{seg})$

$s(C_{seg})$ estimates the probability that the coverage along the contig is best explained by a single Dirichlet distribution, as opposed to two or more. A low score might indicate chimerism (Figure 1), which could be detected as multiple distinct reference transcripts aligning to a contig in a non-overlapping manner. We aligned all contigs to the reference transcriptome using blastn. Alignments were visualised so that a black bar represents the assembled contig, while coloured bars represent the aligned transcripts and/or genes. All reference hits were placed on the same plane to allow identification of non-overlapping hits. Most contigs received hits, and the majority had non-overlapping hits from multiple distinct transcripts, consistent with chimerism.

30 lowest scoring

	Component score	Plot
comp2222_c0_seq1	0.02	
comp3736_c0_seq1	0.03	
comp2527_c0_seq1	0.03	
comp2420_c1_seq1	0.03	
comp3779_c0_seq1	0.04	
comp3698_c0_seq1	0.04	
comp2292_c0_seq1	0.04	
comp3016_c0_seq1	0.04	
comp3414_c0_seq1	0.04	
comp3669_c0_seq1	0.04	
comp2338_c0_seq1	0.04	
comp3807_c0_seq1	0.04	
comp3097_c1_seq1	0.04	
comp3665_c2_seq32	0.04	
comp4106_c0_seq1	0.04	
comp2456_c0_seq3	0.04	

comp2386_c0_seq1	0.05	
comp3685_c0_seq1	0.05	
comp2060_c0_seq1	0.05	
comp3448_c2_seq1	0.05	
comp3533_c0_seq3	0.05	
comp3191_c0_seq1	0.05	
comp3600_c1_seq1	0.05	
comp3377_c0_seq1	0.05	
comp3432_c0_seq1	0.05	
comp2320_c0_seq1	0.05	
comp4240_c0_seq1	0.05	
comp3252_c0_seq1	0.05	
comp3256_c0_seq1	0.05	
comp3179_c1_seq1	0.05	

$s(C_{ord})$

measures the structural correctness of the contig. A low score might indicate various kinds of mis-assembly, which could be detected as multiple hits inverted relative to one another, or no hits at all. Alignments and visualisation were carried out as for. No contigs received hits, consistent with mis-assembly.

30 lowest scoring

	Component score	Plot
comp53_c0_seq1	0.00	
comp233_c0_seq1	0.00	

comp356_c0_seq1	0.00	contig hits 
comp399_c0_seq1	0.00	contig hits 
comp443_c0_seq1	0.00	contig hits 
comp578_c0_seq1	0.00	contig hits 
comp607_c0_seq1	0.00	contig hits 
comp999_c0_seq1	0.00	contig hits 
comp1080_c0_seq1	0.00	contig hits 
comp1179_c0_seq1	0.00	contig hits 
comp1210_c1_seq2	0.00	contig hits 
comp1225_c0_seq1	0.00	contig hits 
comp1233_c0_seq1	0.00	contig hits 
comp1283_c0_seq1	0.00	contig hits 
comp1395_c0_seq1	0.00	contig hits 
comp1407_c0_seq1	0.00	contig hits 
comp1463_c0_seq1	0.00	contig hits 
comp1471_c0_seq1	0.00	contig hits 
comp1518_c0_seq1	0.00	contig hits 
comp2518_c1_seq1	0.00	contig hits 
comp2520_c0_seq1	0.00	contig hits 
comp2567_c0_seq1	0.00	contig hits 
comp2578_c2_seq1	0.00	contig hits 

comp2616_c1_seq1	0.00	contig hits 
comp2651_c1_seq1	0.00	contig hits 
comp2671_c0_seq1	0.00	contig hits 
comp2858_c0_seq1	0.00	contig hits 
comp2958_c0_seq2	0.00	contig hits 
comp2976_c1_seq1	0.00	contig hits 
comp2998_c0_seq1	0.00	contig hits 

$s(C_{cov})$

measures the proportion of bases that have read support. A low score indicates unsupported noise sequence in the contig. Although this score component captures the information by definition, we performed an independent verification of the analysis by using SAMtools mpileup to generate per-base coverage from the assigned read alignments. We visualised contigs as a black bar with a red bar below each position that received at least one supporting read. All contigs received no supporting reads.

30 lowest scoring

	Component score	Plot
comp2036_c0_seq1	0.00	
comp3654_c0_seq1	0.00	
comp3654_c0_seq5	0.00	
comp3654_c0_seq6	0.00	
comp3654_c0_seq7	0.00	
comp3654_c0_seq8	0.00	
comp3654_c0_seq9	0.00	
comp3654_c0_seq11	0.00	
comp3654_c0_seq13	0.00	

comp3654_c0_seq14	0.00	contig hits 
comp3654_c0_seq17	0.00	contig hits 
comp3654_c0_seq18	0.00	contig hits 
comp3654_c0_seq19	0.00	contig hits 
comp3654_c0_seq21	0.00	contig hits 
comp3663_c3_seq2	0.00	contig hits 
comp3665_c2_seq13	0.00	contig hits 
comp3665_c2_seq15	0.00	contig hits 
comp3665_c2_seq16	0.00	contig hits 
comp3665_c2_seq17	0.00	contig hits 
comp3665_c2_seq18	0.00	contig hits 
comp3665_c2_seq19	0.00	contig hits 
comp3665_c2_seq20	0.00	contig hits 
comp3665_c2_seq21	0.00	contig hits 
comp3665_c2_seq23	0.00	contig hits 
comp3665_c2_seq24	0.00	contig hits 
comp3665_c2_seq25	0.00	contig hits 
comp3665_c2_seq26	0.00	contig hits 
comp3665_c2_seq27	0.00	contig hits 
comp3665_c2_seq30	0.00	contig hits 
comp3665_c2_seq36	0.00	contig hits