

Supplemental Fig. 1. Sequence similarity search results for the entire ~1.2 Mb contig is shown in conjunction with RepeatMasker analysis. The 60-kb gap estimated between clones RP11-336L20 and RP11-32B5 is included for illustrative purposes. Above the black line, which represents the contig sequence, the output from RepeatMasker has been depicted using color coded boxes, however the output has been filtered for LTR, Simple Repeats/Low-Complexity Repeats, LINEs, SINEs and both centromeric and telomeric Satellite sequences only. All other elements detected by RepeatMasker have been removed for illustrative purposes. Note the block of centromeric satellite sequence at the proximal end (left) of the contig. The duplicons identified in Figure 1 have been added for reference purposes. Below the black line, the alignments produced by a sequence similarity search (using MEGABLAST) against the human genome assembly (build34, July 2003) are shown color coded according to the percent identity of the alignment. Each horizontal box therefore depicts an alignment. For global analysis of segmental duplications, the genome is fractioned into 400-kb pieces, and the identity of each 400-kb piece is indicated to the left hand side of the alignments produced from that segment of the assembly, alignments to the respective chromosomal random bins are indicated by the "r". Note the overall complexity of the paralogous relationships depicted by this analysis and variation in the number of paralogous sites, which is indicative of wide variation in copy number of these duplicated segments. For simplicity, only alignments >1.5 kb are shown.

