



**Additional data file 8.** Percentage of sequence encoding genes on incorrectly binned fragments of 5 kb (blue) and 1 kb (red) length. Error bars represent one standard deviation. Dotted line indicates genome-wide average % of sequence coding for genes. \* Frac. seqs. < genome avg. is the fraction of incorrectly binned sequences with coding % less than one standard deviation below the genome average.