## **Supplemental Figure S1.**

Size distribution of the 4,993 contigs and 10,865 singletons resulting from the global assembly of 76,134 clones. Frequencies of singletons and contigs are indicated in blue for the full set of sequences, in red for sequences without CDS, in green for CDS with no detected similarity and in purple for CDS with homologs. We used a logarithmic scale for representation convenience.

