



#### Additional file 7 – Effects of data size on MDL plots

The ten variously downsized IMR90 datasets were used for MDL plots under four different thresholds of minimal read depth (i.e.,  $\geq 3$ ,  $\geq 5$ ,  $\geq 10$  or  $\geq 20$  reads). The number in parenthesis indicates the mean depth of genomic coverage.