

Supplemental Table S3. Read error correction: Illumina-derived versus genome-derived list of valid k -mers. All values are 95% trimmed means (to remove outliers).

Organism	Number of reads*	Method†	Valid k -mer source	Total k -mers	Correct k -mers	% correct
<i>E. coli</i>	7,410	M	-	15,045.3	14,242.0	94.66
<i>E. coli</i>	7,410	L	Illumina (low frequency removed)	15,118.2	14,467.1	95.69
<i>E. coli</i>	7,410	L	genome (all k -mers)	15,118.2	14,467.1	95.69
<i>E. coli</i>	7,410	LH	Illumina (low and high frequency removed)	15,119.0	14,469.3	95.70
<i>E. coli</i>	7,410	LH	genome (single-copy k -mers)	15,119.0	14,469.4	95.70
<i>C. elegans</i>	128,710	M	-	18,699.7	17,391.2	93.03
<i>C. elegans</i>	128,710	L	Illumina (low frequency removed)	18,805.5	17,563.3	93.44
<i>C. elegans</i>	128,710	L	genome (all k -mers)	18,806.3	17,565.0	93.50
<i>C. elegans</i>	128,710	LH	Illumina (low and high frequency removed)	18,804.8	17,575.7	93.52
<i>C. elegans</i>	128,710	LH	genome (single-copy k -mers)	18,805.3	17,575.4	93.52

* Exactly the same reads were compared across the five combinations of method and valid k -mer source.

† "M", standard MHAP; "L", MHAP with low-frequency k -mer masking; "LH", MHAP with low and high frequency k -mer masking.