

Data Cleanup: Contaminant Removal - Answer Key

Locate the Illumina adaptor and Phix contaminant sequences in the below set of 16 reads, and cross them out to remove them from your data set.

Illumina adaptor sequences are filled in with light grey below, Phix sequences are filled in with dark grey.

How many reads remain after filtering? What percentage of the original read set does this amount to?

12 of the original 16 reads remain, accounting for 75% of the original number of reads.

Illumina Adaptor Sequence

CTGTCT

Phix Sequences

CAAACATTGGGCCAAATGA

GGGCGTTGTATGGTTGCCA

AATACCCCCAGACGTCGGT

TACCGAGTTTCCGATTCGC

READ ID	SEQUENCE																								
1	C	T	G	T	C	T	A	T	G	A	T	A	A	A	A	G	T	T	C	A	A	T	T	G	A
2	C	T	G	T	C	T	T	G	C	A	A	T	G	C	G	C	T	T	A	T	T	G	A	A	A
3	C	T	G	T	C	T	T	G	C	A	A	T	G	C	G	C	T	T	A	T	T	G	A	A	A
4	C	T	G	T	C	T	T	G	C	A	A	T	G	C	G	C	T	T	A	T	T	G	A	A	A
5	C	T	G	T	C	T	T	G	C	A	A	T	G	C	G	C	T	T	A	T	T	G	A	A	A
6	C	T	G	T	C	T	T	G	C	A	A	T	G	C	G	C	T	T	A	T	T	G	A	A	A
7	C	T	G	T	C	T	T	G	C	A	A	T	G	C	G	C	T	T	A	T	T	G	A	A	A
8	C	T	G	T	C	T	T	G	C	A	A	T	G	C	G	C	T	T	A	T	T	G	A	A	A
9	C	T	G	T	C	T	T	G	C	A	A	T	G	C	G	C	T	T	A	T	T	G	A	A	A
10	C	T	G	T	C	T	T	G	C	A	A	T	G	C	G	C	T	T	A	T	T	G	A	A	A
11	C	T	G	T	C	T	T	G	C	A	A	T	G	C	G	C	T	T	A	T	T	G	A	A	A
12	C	T	G	T	C	T	A	G	T	C	T	T	T	C	T	T	T	C	C	A	A	T	T	T	G
13	C	T	G	T	C	T	G	T	A	G	G	T	T	T	A	G	T	A	A	A	G	C	A	T	G
14	C	T	G	T	C	T	A	C	T	T	G	A	C	C	C	G	G	G	G	C	A	A	A	G	T
15	C	T	G	T	C	T	A	A	T	A	C	C	C	C	C	A	G	A	C	G	T	C	G	G	T
16	C	T	G	T	C	T	A	C	G	A	T	A	G	G	G	A	C	A	C	C	A	G	C	A	A