



Better lives through livestock

FASTQ File Format

Kennedy Mwangi

International Livestock Research Institute (ILRI)

Viral Pathogen Genome Sequencing and Bioinformatics Analysis Training Workshop
6th – 17th May, 2024

ILRI
INTERNATIONAL
LIVESTOCK RESEARCH
INSTITUTE



FASTQ File Data

@K00364:64:HTYYCBBXX:1:1108:4635:14133/1

TAGAATACGCCAGGTGTAAGAATAGTAGAATACGCCAGGTGTAAGAATAGTAGAACACGCCAGGTGTAAGAATAGTAGAA

+

AAAFJJJJJJFFJJJJFJJFFJJFJJJJJJJJFFJJJJFJJJFJJAJJJFJJJFJJJJ7JJJJF-<JAFJJ<F<AJAJJJ

@K00364:64:HTYYCBBXX:1:1108:5081:25527/1

AAAACACCAAAGAAAGATGCCAGGGTCCCTGCTCATCTGCGTGAAAGTGACTTAGGCATGCTGCAAGGAGGCATGAGG

+

[illegible]

@K00364:64:HTYYCBBXX:1:1108:5852:47295/1

AGGTGGCTCTAGAAGGTTCTCGGACCGAGAAACAGCCTCGTACATTTGCAATGATTTCAATTCATTTTGACCATTACGAA

+

[illegible]

← Sequence header

← Sequence

← Separator

← Base call quality score

Illumina versions 1.8+ of Casava

```
@A00600:191:HY75HDSX2:1:2624:27480:35744 2:N:0:AAGACCGT+CAATCGAC
```

@	: mandatory character that starts the ID line
A00600	: Unique sequencing machine ID
191	: Run number on instrument
HY75HDSX2	: Unique flow cell identifier
1	: Lane number
2624	: Tile number (section of the lane)
27480	: x-coordinate of the cluster within the tile
35744	: y-coordinate of the cluster within the tile
2	: Whether the sequence is from read 1 or read 2
N	: N if the sequence was not flagged as crappy by the machine, Y otherwise
0	: 0 when no control bits are turned on
AAGACCGT+CAATCGAC	: Index/barcode of the sample

Acknowledgments

