Detection of Lumpy Skin Disease Virus Reads in the Human Upper Respiratory Tract Microbiome Requires Further Investigation

Authors: Siddharth Singh Tomar[1,2] & Krishna Khairnar[1,2*]

Author Affiliation: [1]Environmental Epidemiology and Pandemic Management (EE&PM), Council of Scientific and Industrial Research-National Environmental Engineering Research Institute (CSIR-NEERI), Nagpur, India.

&

[2] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, UP, India.

*Corresponding author: Correspondence to Krishna Khairnar e-mail: k_khairnar@neeri.res.in, kskhairnar@gmail.com

Abstract

Lumpy skin disease virus (LSDV), a double-stranded DNA virus from the Capripoxvirus genus, primarily affects Bos indicus, Bos taurus breeds, and water buffalo. Arthropod vectors, including mosquitoes and biting flies, are the main LSDV transmitters. Although LSDV is not zoonotic, this study unexpectedly detected LSDV reads in the upper respiratory tract microbiome of humans from rural and urban areas in Maharashtra, India. Nasopharyngeal and oropharyngeal swab samples collected for SARS-CoV-2 surveillance underwent whole-genome metagenomics sequencing, revealing LSDV reads in 25% of samples. SKA analysis provided insights into sample relatedness despite the low coverage of LSDV reads with the reference genome. Our findings, which include the detection of LSDV contigs aligning to specific locations on the reference genome, suggest a common source for LSDV reads, potentially shared water sources or milk/milk products. Further investigation is needed to ascertain whether these reads indicate actual LSDV infections, environmental uptake, or cross-reactivity with related viral sequences.

Keywords: *Lumpy skin disease virus (LSDV), Metagenomics, Upper respiratory tract, Zoonoses, Human-animal interface*

1. Introduction

Lumpy skin disease virus (LSDV) is a double-stranded DNA virus belonging to the Capripoxvirus genus of the *Poxviridae* family. Within this genus are two other virus species: sheeppox and goatpox. LSDV exhibits a strong host specificity and causes disease mainly in *Bos indicus* and *Bos taurus* breeds and water buffalo (*Bubalus bubalis*).

Arthropod vectors primarily transmit LSDV. Studies have demonstrated that *Aedes aegypti* mosquitoes[1], *Stomoxys calcitrans*[2], and *Haematopota spp*. biting flies can mechanically transmit LSDV[3]. Additionally, other mosquitoes such as *Culex mirificens* and *Aedes natrionus*[4], flies like *Biomyia fasciata*, *Culicoides*, and ticks such as *Riphicephalus appendiculatus* and *Amblyomma hebraeum*, are likely to contribute to virus transmission in natural settings[5]. Contact with infected animals is considered to have a minor role in virus transmission[6]. Some studies suggest non-vector transmission such as ingesting contaminated feed and water with infected saliva[7,8].

LSDV is not a zoonotic virus as there is no evidence to suggest that it can infect humans or cause disease in them. However, in this study, we observed the presence of LSDV reads in the upper respiratory tract (URT) microbiome of humans. The Nasopharyngeal (NP) and Oropharyngeal (OP) samples belong to different locations in the central Indian region of Vidarbha (Maharashtra). These samples were collected as part of the Central government's SARS-CoV-2 genome surveillance program. These samples were subjected to non-targeted metagenomics to understand the microbial diversity of the samples. The collection period of samples coincides with the period of LSDV circulation in the region. While analyzing the data, the LSDV reads appeared in the samples.

This observation suggests the possibility of non-vector-borne transmission of LSDV. Despite LSDV not infecting humans, extensive bovine-human contact in India through activities such as agriculture, dairy production, and religious rituals raises concerns about the transmission of other diseases from cattle to humans. While LSDV itself isn't zoonotic, other diseases carried by cattle could pose risks to humans. There could also be a possibility of humans passively transmitting LSDV to cattle.

2. Materials and Methods

2.1. Sample Collection

NP-OP swab samples were collected in Viral Transport Medium (VTM) by trained paramedical staff at various primary healthcare centres and SARS-CoV-2 molecular testing laboratories from 5 districts of Vidarbha region during March-April 2023. The samples were then tested for SARS-CoV-2 using qRT-PCR. Positive samples ≤ 25 cycle threshold value were selected and transported for SARS-CoV-2 whole genome sequencing (WGS), aliquots of these positive samples were preserved at -80^{o}C. 48 (n) aliquots of these positive samples were selected for whole-genome metagenomics (WGMG) sequencing. The median (interquartile range [IQR]) age was 36 (65 to 16) years, and 24 (50%) of the samples were from females. Location data of every sample was collected and classified as urban and rural households, and the urban-rural distribution of samples stands at 20 (41.6%) and 28 (58.3%) respectively **(Supplementary data)**.

2.2. DNA extraction, Library preparation, & Metagenomic Sequencing

DNA extraction was performed with QIAamp DNA Microbiome Kit (51704) per the kit's protocol. Post-DNA extraction, the Preliminary quality control was done using Qubit to determine the concentration (ng/ul) and Nanodrop (A260/280, A260/230) to determine the purity of extracted metagenomic DNA. Library preparation was done with QIAseq FX DNA Library preparation kit[9] per the kit's protocol. The sequencing platform utilised was NextSeq550, employing a 2x150 chemistry high-output 300-cycle kit.

2.3. Data Analysis

Fastq files generated after sequencing were analyzed using the Chan Zuckerberg ID (v8.3)[10], a cloud-based bioinformatic data analysis software. The following steps were involved in the data analysis using Chan Zuckerberg ID (CZid)

2.3.1 Sequencing Data QC

External RNA Controls Consortium (ERCC) sequences were removed using Bowtie2[11]. Sequencing adapters, short reads, sequences with low quality, and low complexity

regions were filtered out using a customized version of fastp. Specifically, bases with quality scores below 17, reads shorter than 35 bp, sequences containing low complexity regions exceeding 40%, and sequences with more than 15 undetermined bases (Ns) were excluded.

2.3.2. Host Filtering

Human sequences were removed through Bowtie2[11] and HISAT2 alignments against reference human genomes. Only one representative read is retained for sequences that are 100% identical based on the first 70 bp, using CZID-dedup. Host reads were filtered out using the Spliced Transcripts Alignment to A Reference (STAR) algorithm. Duplicate, low-quality, and low-complexity sequences were removed in this process. The non-human reads were then aligned to the NCBI nucleotide and protein database (NCBI Index Date: 2021-01-22) using GSNAPL and RAPSearch, respectively.

2.3.3. Alignment

Host-filtered reads were aligned to the NCBI nucleotide (NT) database using Minimap2[12] and the NCBI protein (NR) database using the Diamond tool[13]. The hits from Minimap2 and Diamond were assigned their corresponding accession numbers. Combined taxon counts were generated by using Gsnap and Rapsearch.

2.3.4. Assembly

To improve sensitivity in mapping, short reads are de novo assembled using SPADES[14] to generate longer contigs. However, SPADES output lacks information on individual contigs associated with each read. To address this, bowtie2 maps original reads back to assembled contigs, restoring the link between reads and contigs. This process utilizes a Bowtie2 database constructed from contigs. BLAST analysis is then conducted separately for contigs, initially against the NT-BLAST database generated from short-read alignments to NCBI NT using GSNAP, followed by NR-BLAST against NCBI NR using Rapsearch2.

2.4. Threshold for Viral Reads Detection

Detection of virus reads is challenging in whole genome metagenome data due to their low abundance, lower genome size, and high genomic diversity. Therefore, selecting the threshold criteria is crucial for reliable detection. Grundy et al. (2023)[15] experimentally

established that 2.4 reads per million (rPM) is the optimum threshold for viral pathogen detection in a metagenomics dataset. We, however used the more conservative filters NT rPM >= 10 and NT L (alignment length) >=35. We utilized the NT and NR sequence filters to mitigate the risk of false identifications in the NCBI database due to misalignment or misannotation.

3. Results

3.1. Sequencing Data Quality Control (QC)

The data set consists of 48 NP-OP samples. On average (mean ± SD), these samples had 7,258,602 (7.26 million) ± 1,540,000 reads, of which 1,777,507 (1.77 million) ± 5,684,09 reads (25.63 % ± 10.17 %) were identified as nonhuman reads (passed filters). **(Figure 1)** The average minimum and maximum insert sizes were 35 and 557, respectively. The Average Duplicate Compression Ratio (DCR) of data was 1.01 (for metagenomics without enrichment, a DCR value less than 2 is ideal)[16]. Overall, 81% of reads passed the QC following filtering with fastp to eliminate low-quality bases, short reads (less than 35 base pairs in length), and reads with low complexity. **(Figure 2)**

3.2. Detection of Viral Reads in the Dataset

After applying specific thresholds for the detection of viral reads in the metagenomic data, including Nucleotide reads per million (NT rPM) >= 10, Nucleotide alignment length in base pairs (NT L) >= 35, and Nucleotide (NT) and Protein (NR) Z Scores >= 1 against a background model of (water only), the analysis revealed the presence of Lumpy Skin Disease Virus (LSDV) reads in 12 out of 48 (25%) samples **(Figure 3)**. (sample 23G214-5G-294_S21 was not considered as it had no aligned contigs with the LSDV reference genome KX683219.1) However, it showed 76% coverage and 2.2x depth when aligned to LSDV strain EGPT75 EEV126 gene, partial cds (MH639086.1).

The average alignment length of the reads is 746 ± 231 bases. However, these samples showed a low coverage and sequencing depth for LSDV reads. Samples had 0.6 ± 1.12 average (Coverage Depth X) and 0.87 ± 0.4 (Coverage Breadth %) **(Table 1)**

3.3. Consensus Genome

A consensus genome derived from samples containing LSDV (Lumpy Skin Disease Virus) reads would represent a composite sequence assembled from the overlapping sequences obtained from the dataset. The sample with the highest rPM value (1440)

was selected to construct the consensus genome. Czid's consensus genome feature for viral reads was used to create the consensus genome of LSDV [17]. The consensus genome was built against the LSDV reference sequence (AF409137.1). The consensus genome showed 27.7% GC content, with 2271 informative nucleotides covering 1.5% of the LSDV genome **(Figure 4)**.

3.4. General Relationships among the Samples

As the selected samples have contigs covering a small portion of the reference sequence of LSDV (<25%) therefore, the data is not sufficient for constructing a phylogenetic tree[18]; instead, we used Czid's web-based pipeline to create a pairwise distance matrix for understanding the relationships between identified taxa (LSDV) in the samples. The distance matrix depicts Mash-like distances between samples, calculated using the Jaccard Index (j) based on split kmer matches[19]. Mash-like distances are shown in a pairwise matrix, where each sample is compared to itself and others. Colors represent distance ranges, with dark red squares along the diagonal indicating identical sequences **(Figure 5)**.

3.5. Genomic Position and Classification of (LSDV) Contigs

Czid's coverage visualisation tool was used to determine the location of contigs over the reference genome. The corresponding accession numbers of reference genomes were also annotated. These samples were classified as wild-type (WT) and vaccine strain (V) based on the classifications as reported by Flannery et al. (2021)[20]. The table also includes information on each sample's total number of contigs and specifies its genomic position over the reference genome. It was observed that the contigs aligned to some specific regions of the reference genomes; these regions are 55109 to 57997 bp, 115916 to 117098 bp, 116144 to 117104 bp, and 124934 to 127961 bp. **(Table 2)** This observation is indicative of a probable common source of the LSDV reads.

3.6. CZid pipeline compared with Centrifuge

CZid uses the Genomic Short-read Nucleotide Alignment Program GSNAP[21] to align the reads with the reference genome using the NCBI index. Centrifuge[22] employs an indexing strategy derived from the Burrows-Wheeler transform (BWT) and the Ferragina-Manzini (FM) index for efficient metagenomic classification. We compared the results from CZid (rPM) and Centrifuge (Unique Reads) to avoid any pipeline-dependent biases;

both results followed a similar trend, as shown in **(Figure 6)** detailed results are shown in **(Table 3)**

4. Discussion

In this study, we have reported the detection of LSDV reads in the human upper respiratory tract microbiome. 48 NP-OP swab samples from 5 districts of Maharashtra were randomly selected for this study; these samples were collected from different healthcare centers and testing labs for SARS-CoV-2 genome surveillance. These samples were then subjected to enrichment-free whole-genome metagenomics (WGMG) sequencing to investigate upper respiratory tract microbiome composition.

The generated data set consists of 7.25 million reads on average, with an average of 25% of reads passing the host filters. The detection of LSDV reads was consistently observed among the samples; 20 out of 48 samples were detected with LSDV reads before applying the selection threshold criteria of NT rPM >= 10 and NT L (alignment length) >=35. After using this criterion, the number of samples with LSDV reads detection decreases to 12 (25%) of the total samples.

It was observed that the LSDV viral contigs showed less coverage with the reference genome. The low coverage of LSDV viral contigs with the reference genome could be attributed to several factors, including generally low biomass NP-OP sampling.[23] The samples involved in this study were OP-NP swab samples collected in VTM by primary healthcare workers from urban and rural regions of Vidarbha (Central India). Maintaining Cold chain transport is challenging for samples from rural areas; a break in the cold chain may contribute to sample degradation, compromising the coverage during sequencing.

Also, factors like the physical conditions of NP-OP, depth of sampling, the presence of mucous, and the age of the participants may determine the availability of biomass/tissue material while sampling. Metagenomics data generation, especially with low biomass NP-OP swab samples, presents challenges due to limited microbial material, complex microbial communities, and host DNA contamination.[24,25,26]

Another reason for the observed low coverage of LSDV reads could be genetic variability within the viral population that can lead to mismatches between the reference genome and sequenced reads, reducing alignment efficiency and coverage of viral contigs.[27,28]

As the contigs covered only a limited portion of the reference sequence of LSDV in the selected samples, this prevented the construction of a phylogenetic tree. Therefore, we used Czid's web-based pipeline, which allowed us to create a pairwise distance matrix. This pipeline used the split kmer analysis (SKA) method to determine the degree of relatedness among the samples.SKA enables quick comparison and alignment of shorter genome fragments, which makes it well-suited for surveillance and outbreak analyses. It can create split kmer files from fasta or fastq formats, cluster them, align them with a reference sequence, and offer various comparative and summary metrics. SKA has demonstrated accuracy with Illumina sequencing data.[29] Furthermore, a comparison between CZid and Centrifuge pipelines demonstrated similar trends in results, indicating consistency across different analysis methods.

The detection of LSDV reads in the human upper respiratory tract microbiome poses intriguing questions and warrants further investigation. While LSDV is traditionally associated with bovines and causes significant economic losses in the agricultural sector, its presence in the human respiratory tract microbiome indicates a potential uptake by humans from the environment. LSDV is primarily transmitted through arthropod vectors and direct contact with infected animals, thus making its presence in the human upper respiratory tract unexpected. Out of 12 samples with LSDV reads, seven belong to rural areas, and five belong to Urban areas. LSDV reads in the URT microbiome could also be attributed to frequent contact between humans and livestock in India.[30]

Another plausible explanation for LSDV reads detection in URT could be traced to consuming milk from infected bovines. Bedkovic (2018) [31] reported the detection of LSDV viral particles in Milk. These viral particles in milk could either be from naturally infected or animals vaccinated with Live, Attenuated Vaccines, as in the case of LSDV. DNA from LSDV wildtype and vaccine strains were also detected in the milk and body fluids of vaccinated animals.[31,32] Our data also showed the detection of contigs from both

wild-type and Vaccine strains of LSDV. These contigs were observed to be consistently aligning to specific locations on the reference genome, this observation is suggestive of a common source for LSDV reads. This common source is more likely to be a water source shared by both humans and animals or milk/milk-products from LSDV infected or Vaccinated bovines.

Therefore, it is crucial to investigate whether these reads represent actual LSDV infections in humans, take-up from the environment, or cross-reactivity with related viral sequences. Accurate taxonomic classification and identification of viral sequences require comprehensive reference databases and stringent alignment criteria to minimize false-positive results and ensure the reliability of findings.[33] Therefore, investigations in this direction further validate the presence of LSDV reads using alternative experimental approaches and studies in controlled environments.

Acknowledgement

The authors are thankful to CSIR-NEERI for providing funds under project OLP-57 (March 2023 - April 2024) for conducting this study

Data availability statement

The data for the samples in this study is available as **"Viral_Metagenome_Covid"** on CZid.

Author declaration

The author assures that the research has followed all ethical guidelines and received approvals from the Institutional Ethics Committee for Research on Human Subjects (IEC) of CSIR-NEERI, Nagpur-20, India. Necessary consent from patients/participants has been obtained and relevant institutional documentation is archived.

Confidentiality declaration

Sample IDs **(23G214-5G-264_S1 to 23G214-5G-311_S48)** are masked IDs and cannot be traced to participant details. Precise age of the participants is masked and non-ovelapping age ranges were used.

Figures

Figure 1: Distribution of total reads and nonhuman reads in the dataset. Each sample, on average, contained approximately 7,258,602 reads, with 25.6% identified as nonhuman reads after passing filters.

Figure 2: Quality control (QC) results showing the percentage of reads passing filtering with fastp, eliminating low-quality bases, short reads (<35 base pairs), and reads with low complexity. Overall, 81% of reads passed the QC process.

Figure 3: Detection of Lumpy Skin Disease Virus (LSDV) reads in metagenomic samples after applying specific thresholds for viral read detection.

Figure 4: Consensus genome analysis against the LSDV reference sequence (AF409137.1) showing 27.7% GC content and 2271 informative nucleotides, covering 1.5% of the LSDV genome.

Figure 5: Pairwise Mash-like distance matrix illustrating Jaccard Index (j) based on split kmer matches between samples. Dark red squares along the diagonal indicate identical sequences, with colors representing distance ranges.

Figure 6: Comparison of CZid (rPM) and Centrifuge (Unique Reads) pipeline results, demonstrating a consistent trend across both methods.

Tables

Sample	NCBI Reference	Reference Length	Aligned Contigs	Aligned Loose Reads	Coverage Depth X	Coverage Breadth %	Max Alignment Length	Avg. Mismatched %	% Matched
23G214-5G-280_S17	AF409137.1	150,793.00	1.00	0.00	0.03	0.40	617.00	0.20	99.80
23G214-5G-281_S18	AF325528.1	150,773.00	2.00	3.00	0.07	1.00	961.00	0.60	99.40
23G214-5G-282_S19	KY702007.1	150,661.00	2.00	1.00	0.23	1.00	722.00	0.40	99.60
23G214-5G-285_S22	MT130502.2	146,159.00	2.00	1.00	0.07	1.00	722.00	0.00	100.00
23G214-5G-286_S23	MN072619.1	150,663.00	1.00	3.00	0.06	0.80	961.00	0.20	99.80
23G214-5G-287_S24	MH893760.2	150,751.00	1.00	0.00	0.01	0.20	331.00	0.30	99.70
23G214-5G-288_S25	AF409137.1	150,793.00	4.00	13.00	3.00	1.50	961.00	0.20	99.80
23G214-5G-289_S26	MN072619.1	150,329.00	2.00	5.00	0.35	1.10	961.00	0.00	100.00
23G214-5G-290_S27	KX894508.1	150,562.00	2.00	2.00	0.02	0.60	348.00	0.00	100.00
23G214-5G-291_S28	MN995838.1	150,721.00	2.00	2.00	0.50	0.90	772.00	0.00	100.00
23G214-5G-292_S29	KX683219.1	150,663.00	1.00	1.00	0.02	0.50	644.00	0.20	99.80
23G214-5G-294_S31	AF409137.1	150,793.00	4.00	14.00	3.00	1.50	961.00	0.20	99.80
Mean		**150,305.08**	**2.00**	**3.75**	**0.61**	**0.88**	**746.75**	**0.19**	**99.81**
SD		**1,312.33**	**1.04**	**4.77**	**1.12**	**0.40**	**231.72**	**0.18**	**0.18**

Table 1: Summary of Aligned Contigs, Aligned Loose Reads, Coverage Depth X Coverage Breadth %, and Max Alignment Length of samples containing Lumpy Skin Disease Virus (LSDV) reads.

336

337

S. No.	Sample	NCBI Acession No.	Name of Accession	Vaccine Strain (V) / Wild type Strain (WT)	Total Contigs	Contig 1 Position (bp)	Contig 2 Position (bp)	Contig 3 Position (bp)	Contig 4 Position (bp)
1	23G214-5G-280_S17	AF409137.1	Lumpy skin disease virus NW-LW isolate Neethling Warmbaths LW, complete genome	WT	1		116306-116922		
2	23G214-5G-281_S18	AF325528.1	Lumpy skin disease virus NI-2490 isolate Neethling 2490, complete genome	WT	2	57401-57699	116138-117098		
3	23G214-5G-282_S19	KY702007.1	Lumpy skin disease virus isolate SERBIA/Bujanovac/2016, complete genome	WT	2		115916-116876		127172-127893
4	23G214-5G-285_S22	MT130502.2	Lumpy skin disease virus strain Neethling-RIBSP vaccine, partial genome	V	2	55109-55703			124934-125655
5	23G214-5G-286_S23	MN072619.1	Lumpy skin disease virus isolate Kenya, complete genome	WT	1		115916-116876		
6	23G214-5G-287_S24	MH893760.2	Lumpy skin disease virus strain LSDV/Russia/Dagestan/2015, complete genome	WT	1	57395-57725			
7	23G214-5G-288_S25	AF409137.1	Lumpy skin disease virus NW-LW isolate Neethling Warmbaths LW, complete genome	WT	4	57403-57997	116144-116475	116634-117104	127240-127961
8	23G214-5G-289_S26	MN072619.1	Lumpy skin disease virus isolate Kenya, complete genome	WT	2		115916-116876		127007-127728
9	23G214-5G-290_S27	KX894508.1	Lumpy skin disease virus isolate 155920/2012, complete genome	WT	2	57404-57739			127252-127599
10	23G214-5G-291_S28	MN995838.1	Lumpy skin disease virus isolate pendik, complete genome	WT	2	57361-57955			127196-127917
11	23G214-5G-292_S29	KX683219.1	Lumpy skin disease virus strain KSGP 0240, complete genome	V	1		116256-116896		
12	23G214-5G-294_S31	AF409137.1	Lumpy skin disease virus NW-LW isolate Neethling Warmbaths LW, complete genome	WT	4	57403-57997	116199-116244	116144-117104	127240-127961

Table 2: Details of LSDV Samples: Classification of Vaccine Strain (V) / Wild type Strain (WT), Contig Details, and their Genomic Position

S. No.	sample	Taxon name (Centrifuge)	taxID (Centrifuge)	taxRank (Centrifuge)	numReads (Centrifuge)	numUniqueReads (Centrifuge)	Abundance (Centrifuge)	rPM (CZid)
1	23G214-5G-280_S17	Lumpy skin disease virus NI-2490	376849	leaf	154	134	0.000272504	17.1864
2	23G214-5G-281_S18	Lumpy skin disease virus NI-2490	376849	leaf	538	508	0.049054	10.5188
3	23G214-5G-282_S19	Lumpy skin disease virus NI-2490	376849	leaf	1096	963	0.0503631	336.258
4	23G214-5G-285_S22	Lumpy skin disease virus NI-2490	376849	leaf	185	166	0.00782118	52.9468
5	23G214-5G-286_S23	Lumpy skin disease virus NI-2490	376849	leaf	376	319	0.0106748	70.5397
6	23G214-5G-287_S24	Lumpy skin disease virus NI-2490	376849	leaf	279	238	0.00439487	19.1551
7	23G214-5G-288_S25	Lumpy skin disease virus NI-2490	376849	leaf	1083	956	0.195623	342.871
8	23G214-5G-289_S26	Lumpy skin disease virus NI-2490	376849	leaf	751	683	0.0185272	160.216
9	23G214-5G-290_S27	Lumpy skin disease virus NI-2490	376849	leaf	49	39	0.0120165	14.2801
10	23G214-5G-291_S28	Lumpy skin disease virus NI-2490	376849	leaf	1024	903	0.0321114	329.775
11	23G214-5G-292_S29	Lumpy skin disease virus NI-2490	376849	leaf	159	141	0.000831785	15.5906
12	23G214-5G-294_S31	Lumpy skin disease virus NI-2490	376849	leaf	5956	5332	0.0544535	1440.89

Table 3: Summary of Taxon names, taxon IDs, taxonomic ranks, number of reads, number of unique reads, abundance, and relative abundance (reads per million) of samples with LSDV reads as determined by Centrifuge and CZid.

346

347 **References**

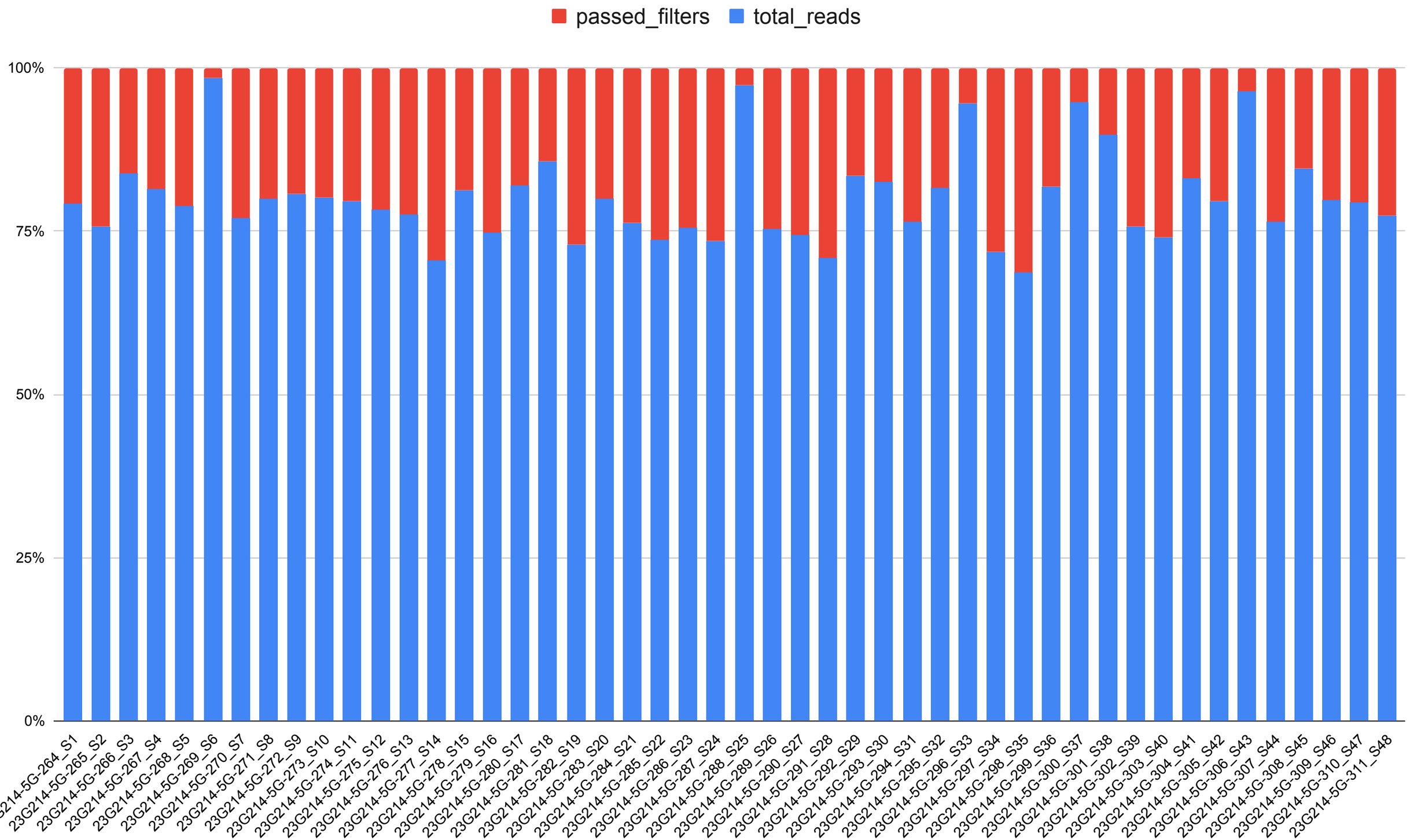
348 1. Chihota, C. M., Rennie, L. F., Kitching, R. P., & Mellor, P. S. (2001). Mechanical
349 transmission of lumpy skin disease virus by Aedes aegypti (Diptera: Culicidae).
350 Epidemiology and Infection, 126(2), 317-321.
351 https://doi.org/10.1017/s0950268801005179
352 2. Haegeman, A., Sohier, C., Mostin, L., De Leeuw, I., Van Campe, W., Philips, W., De
353 Regge, N., & De Clercq, K. (2023). Evidence of Lumpy Skin Disease Virus Transmission
354 from Subclinically Infected Cattle by Stomoxys calcitrans. Viruses, 15(6), 1285.
355 https://doi.org/10.3390/v15061285
356 3. Sohier, C., Haegeman, A., Mostin, L., De Leeuw, I., Campe, W. V., De Vleeschauwer,
357 A., Tuppurainen, E. S., De Regge, N., & De Clercq, K. (2019). Experimental evidence of
358 mechanical lumpy skin disease virus transmission by Stomoxys calcitrans biting flies and
359 Haematopota spp. Horseflies. *Scientific Reports*, *9*(1), 1-10.
360 https://doi.org/10.1038/s41598-019-56605-6
361 4. Gupta, T. D., Patial, V., Bali, D., Angaria, S., Sharma, M., & Chahota, R. (2020). A
362 review: Lumpy skin disease and its emergence in India. Veterinary Research
363 Communications, 44(3–4), 111–118. https://doi.org/10.1007/s11259-020-09780-1
364 5. Lubinga, J. C., Clift, S. J., Tuppurainen, E., Stoltsz, W. H., Babiuk, S., Coetzer, J. A., &
365 Venter, E. H. (2014). Demonstration of lumpy skin disease virus infection in Amblyomma
366 hebraeum and Rhipicephalus appendiculatus ticks using immunohistochemistry. Ticks
367 and Tick-borne Diseases, 5(2), 113–120. https://doi.org/10.1016/j.ttbdis.2013.09.010
368 6. Tuppurainen, E., Alexandrov, T. & Beltrán-Alcrudo, D. 2017. Lumpy skin disease field
369 manual – A manual for veterinarians. FAO Animal Production and Health Manual No. 20.
370 Rome. Food and Agriculture Organization of the United Nations (FAO). 60 pages
371 7. Aleksandr, K., Olga, B., David, W. B., Pavel, P., Yana, P., Svetlana, K., Alexander, N.,
372 Vladimir, R., Dmitriy, L., & Alexander, S. (2020). Non-vector-borne transmission of lumpy
373 skin disease virus. Scientific Reports, 10(1), 1-12. https://doi.org/10.1038/s41598-020-
374 64029-w
375 8. Shumilova, I., Nesterov, A., Byadovskaya, O., Prutnikov, P., Wallace, D. B., Mokeeva,
376 M., Pronin, V., Kononov, A., Chvala, I., & Sprygin, A. (2022). A Recombinant Vaccine-
377 like Strain of Lumpy Skin Disease Virus Causes Low-Level Infection of Cattle through
378 Virus-Inoculated Feed. Pathogens, 11(8). https://doi.org/10.3390/pathogens11080920

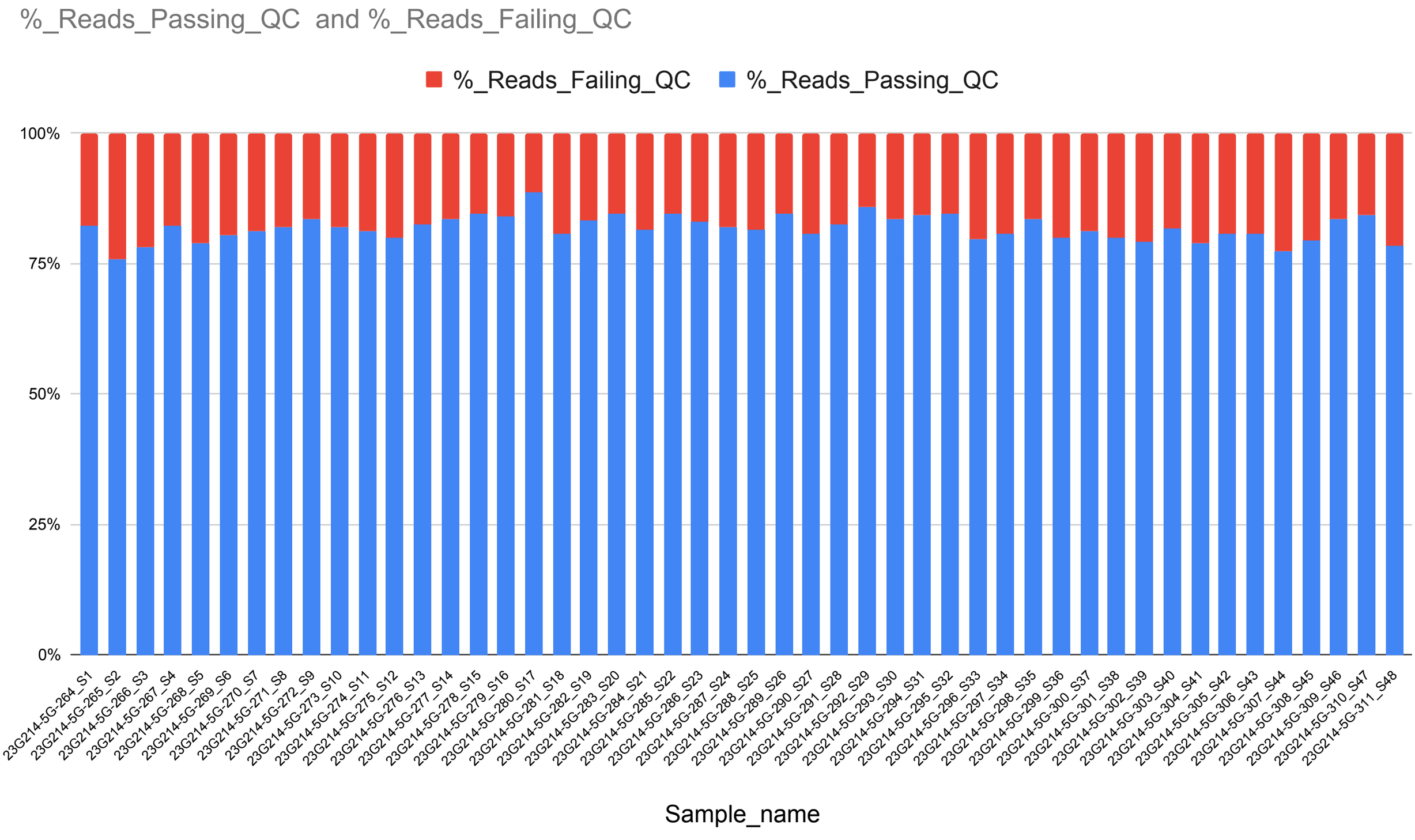
9. QIASEq FX DNA Library Kit. (2023). Qiagen.com. https://www.qiagen.com/zh-us/products/discovery-and-translational-research/next-generation-sequencing/metagenomics/qiaseq-fx-dna-library-kit
10. Simmonds, S. E., Ly, L., Beaulaurier, J., Lim, R. G., Morse, T., Thakku, S. G., Rosario, K., Perez, J. C. R., Puschnik, A. S., Mwakibete, L., Hickey, S. E., Tato, C. M., Team, C. I., & Kalantar, K. (2024). CZ ID: a cloud-based, no-code platform enabling advanced long read metagenomic analysis. bioRxiv (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2024.02.29.579666
11. Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4), 357–359. https://doi.org/10.1038/nmeth.1923
12. Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics, 34(18), 3094-3100. https://doi.org/10.1093/bioinformatics/bty191
13. Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. Nature Methods, 12(1), 59-60. https://doi.org/10.1038/nmeth.3176
14. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology, 19(5), 455-477. https://doi.org/10.1089/cmb.2012.0021
15. Grundy, B., Parikh, H. I., Jacob, S. T., Banura, P., Moore, C. C., Liu, J., & Houpt, E. R. (2023). Pathogen Detection Using Metagenomic Next-Generation Sequencing of Plasma Samples from Patients with Sepsis in Uganda. Microbiology Spectrum, 11(1). https://doi.org/10.1128/spectrum.04312-22
16. Sample QC. (2024). CZ ID Help Center. https://chanzuckerberg.zendesk.com/hc/en-us/articles/360053758913-Sample-QC#Passed-QC
17. Assemble Viral Consensus Genomes from Sample Report. (n.d.). CZ ID Help Center. https://chanzuckerberg.zendesk.com/hc/en-us/articles/13618630200084-Assemble-Viral-Consensus-Genomes-from-Sample-Report
18. Why a pairwise distance matrix instead of a tree? (2024). CZ ID Help Center. https://chanzuckerberg.zendesk.com/hc/en-us/articles/13725914751380-Why-a-Pairwise-Distance-Matrix-Instead-of-a-Tree
19. Simonrharris. (2024). ska distance. GitHub. https://github.com/simonrharris/SKA/wiki/ska-distance#distancetxt-output-columns

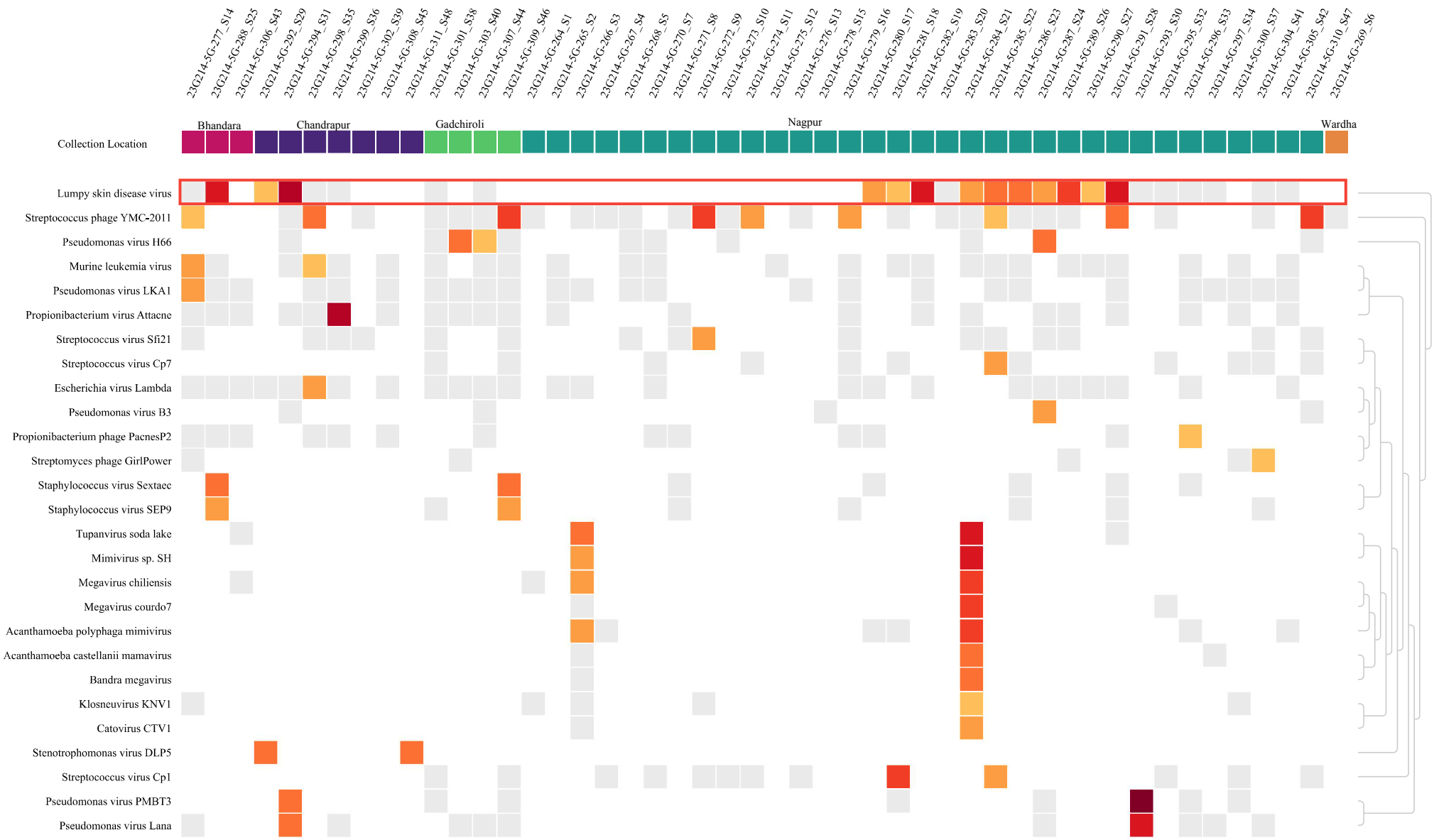
20. Flannery, J., Shih, B., Haga, I. R., Ashby, M., Corla, A., King, S., Freimanis, G., Polo, N., Tse, A. C., Brackman, C. J., Chan, J. Y., Pun, P. H., Ferguson, A. D., Law, A., Lycett, S., Batten, C., & Beard, P. M. (2021). A novel strain of lumpy skin disease virus causes clinical disease in cattle in Hong Kong. Transboundary and Emerging Diseases, 69(4). https://doi.org/10.1111/tbed.14304
21. Wu, T. D., & Nacu, Ş. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics, 26(7), 873–881. https://doi.org/10.1093/bioinformatics/btq057
22. Kim, D., Song, L., Breitwieser, F. P., & Salzberg, S. L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Research, 26(12), 1721–1729. https://doi.org/10.1101/gr.210641.116
23. Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., Oliveira, G., Robles□Sikisaka, R., Rogers, T. F., Beutler, N., Burton, D. R., Lewis-Ximenez, L. L., De Jesus, J. G., Giovanetti, M., Hill, S. C., Black, A., Bedford, T., Carroll, M. W., Nunes, M. R. T., . . . Loman, N. J. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nature Protocols, 12(6), 1261–1276. https://doi.org/10.1038/nprot.2017.066
24. Shi, Y., Wang, G., Lau, H. C., & Yu, J. (2022). Metagenomic sequencing for microbial DNA in human samples: Emerging technological advances. International Journal of Molecular Sciences, 23(4), 2181. https://doi.org/10.3390/ijms23042181
25. Jervis-Bardy, J., Leong, L. E. X., Marri, S., Smith, R. J., Choo, J. M., Smith□Vaughan, H., Nosworthy, E., Morris, P. S., O'Leary, S., Rogers, G. B., & Marsh, R. L. (2015). Deriving accurate microbiota profiles from human samples with low bacterial content through post-sequencing processing of Illumina MiSeq data. Microbiome, 3(1). https://doi.org/10.1186/s40168-015-0083-8
26. Ferretti, P., Farina, S., Cristofolini, M., Girolomoni, G., Tett, A., & Segata, N. (2017). Experimental metagenomics and ribosomal profiling of the human skin microbiome. Experimental Dermatology, 26(3), 211-219. https://doi.org/10.1111/exd.13210
27. García-López, R., Vázquez-Castellanos, J. F., & Moyá, A. (2015). Fragmentation and coverage variation in viral metagenome assemblies, and their effect in diversity calculations. Frontiers in Bioengineering and Biotechnology, 3. https://doi.org/10.3389/fbioe.2015.00141
28. Depledge, D. P., Palser, A. L., Watson, S. J., Lai, I. Y., Gray, E. R., Grant, P., Kanda, R. K., LeProust, E., Kellam, P., & Breuer, J. (2011). Specific Capture and Whole-Genome

Sequencing of Viruses from Clinical Samples. PloS One, 6(11), e27805. https://doi.org/10.1371/journal.pone.0027805

29. Simonrharris. (2024). SKA/README.md at master · simonrharris/SKA · GitHub. https://github.com/simonrharris/SKA/blob/master/README.md
30. Ali, J. (2007). Commentary: Livestock, common property resources and rural smallholders in India. International Journal of Agricultural Sustainability, 5, 265 - 268. https://doi.org/10.1080/14735903.2007.9684826.
31. Bedeković, T., Šimić, I., Krešić, N., & Lojkić, I. (2018). Detection of lumpy skin disease virus in skin lesions, blood, nasal swabs and milk following preventive vaccination. *Transboundary and emerging diseases*, *65*(2), 491–496. https://doi.org/10.1111/tbed.12730
32. Katsoulos, P. D., Chaintoutis, S. C., Dovas, C. I., Polizopoulou, Z. S., Brellou, G., Agianniotaki, E. I., Tasioudi, K. E., Chondrokouki, E., Papadopoulos, O., Karatzias, H., & Boscos, C. (2017). Investigation on the incidence of adverse reactions, viraemia and haematological changes following field immunization of cattle using a live attenuated vaccine against lumpy skin disease. Transboundary and Emerging Diseases, 65(1), 174–185. https://doi.org/10.1111/tbed.12646
33. Rose, R., Constantinides, B., Tapinos, A., Robertson, D. L., & Prosperi, M. (2016). Challenges in the analysis of viral metagenomes. Virus Evolution, 2(2), vew022. https://doi.org/10.1093/ve/vew022


passed_filters
total_reads
100%
75%
50%
25%
0%
23G214-5G-264_S1
23G214-5G-265_S2
23G214-5G-266_S3
23G214-5G-267_S4
23G214-5G-268_S5
23G214-5G-269_S6
23G214-5G-270_S7
23G214-5G-271_S8
23G214-5G-272_S9
23G214-5G-273_S10
23G214-5G-274_S11
23G214-5G-275_S12
23G214-5G-276_S13
23G214-5G-277_S14
23G214-5G-278_S15
23G214-5G-279_S16
23G214-5G-280_S17
23G214-5G-281_S18
23G214-5G-282_S19
23G214-5G-283_S20
23G214-5G-284_S21
23G214-5G-285_S22
23G214-5G-286_S23
23G214-5G-287_S24
23G214-5G-288_S25
23G214-5G-289_S26
23G214-5G-290_S27
23G214-5G-291_S28
23G214-5G-292_S29
23G214-5G-293_S30
23G214-5G-294_S31
23G214-5G-295_S32
23G214-5G-296_S33
23G214-5G-297_S34
23G214-5G-298_S35
23G214-5G-299_S36
23G214-5G-300_S37
23G214-5G-301_S38
23G214-5G-302_S39
23G214-5G-303_S40
23G214-5G-304_S41
23G214-5G-305_S42
23G214-5G-306_S43
23G214-5G-307_S44
23G214-5G-308_S45
23G214-5G-309_S46
23G214-5G-310_S47
23G214-5G-311_S48

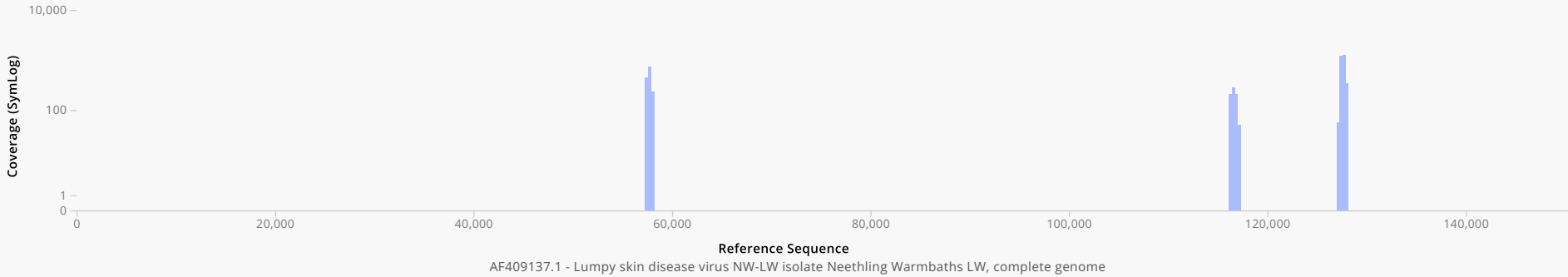

%_Reads_Passing_QC and %_Reads_Failing_QC
%_Reads_Failing_QC
%_Reads_Passing_QC
100%
75%
50%
25%
0%
23G214-5G-264_S1
23G214-5G-265_S2
23G214-5G-266_S3
23G214-5G-267_S4
23G214-5G-268_S5
23G214-5G-269_S6
23G214-5G-270_S7
23G214-5G-271_S8
23G214-5G-272_S9
23G214-5G-273_S10
23G214-5G-274_S11
23G214-5G-275_S12
23G214-5G-276_S13
23G214-5G-277_S14
23G214-5G-278_S15
23G214-5G-279_S16
23G214-5G-280_S17
23G214-5G-281_S18
23G214-5G-282_S19
23G214-5G-283_S20
23G214-5G-284_S21
23G214-5G-285_S22
23G214-5G-286_S23
23G214-5G-287_S24
23G214-5G-288_S25
23G214-5G-289_S26
23G214-5G-290_S27
23G214-5G-291_S28
23G214-5G-292_S29
23G214-5G-293_S30
23G214-5G-294_S31
23G214-5G-295_S32
23G214-5G-296_S33
23G214-5G-297_S34
23G214-5G-298_S35
23G214-5G-299_S36
23G214-5G-300_S37
23G214-5G-301_S38
23G214-5G-302_S39
23G214-5G-303_S40
23G214-5G-304_S41
23G214-5G-305_S42
23G214-5G-306_S43
23G214-5G-307_S44
23G214-5G-308_S45
23G214-5G-309_S46
23G214-5G-310_S47
23G214-5G-311_S48
Sample_name

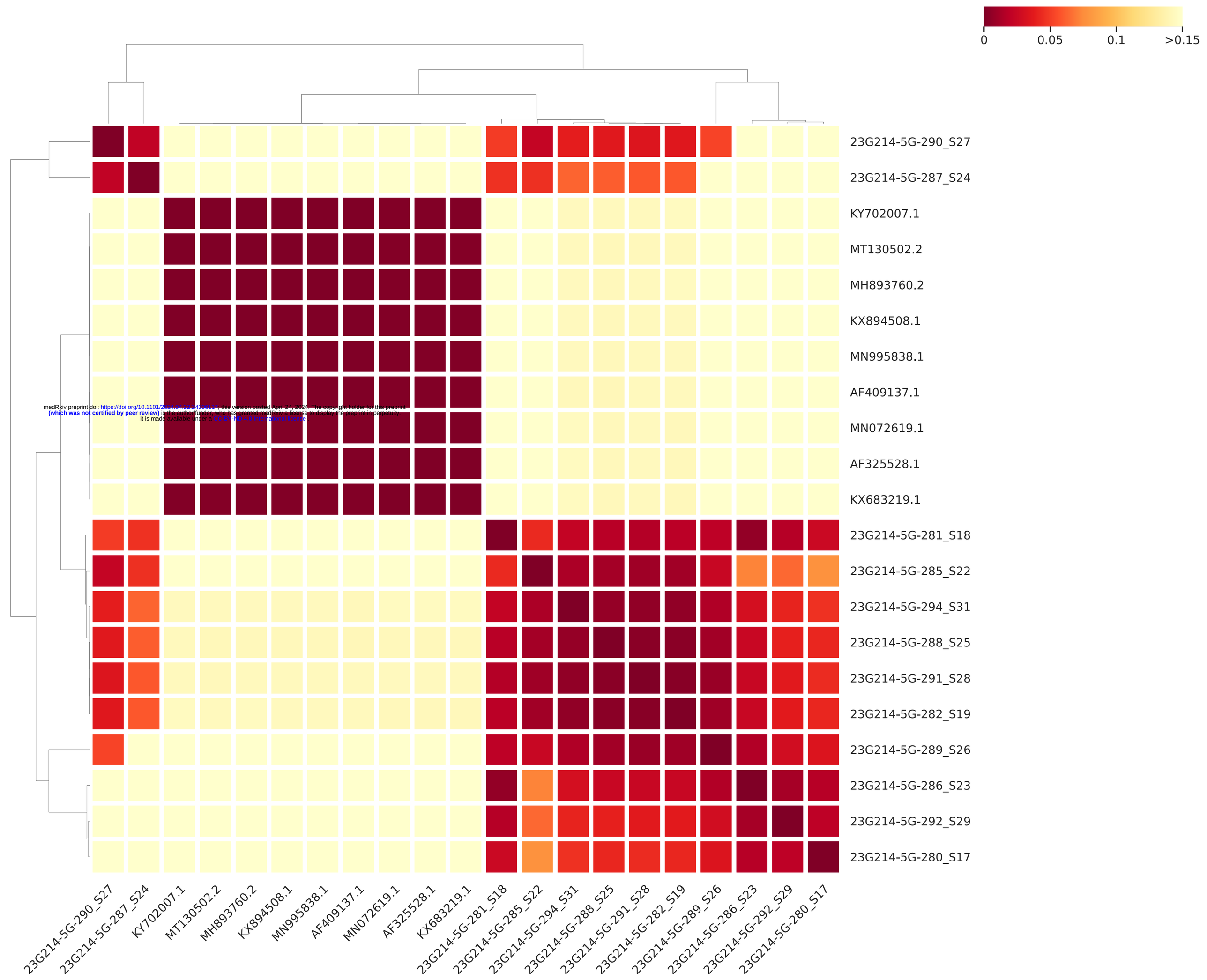


Showing NT.rpm for 10 taxa per sample on a symlog scale; Background: Water_only_control.
Taxa which are not present in the sample are represented by white cells. Taxa filtered out in this view have been hidden or grayed out. 6 filter(s) applied:
Categories: Viruses,Phage; Thresholds: NT rPM >= 10,NT L (alignment length in bp) >= 35,NR Z Score >= 1,NT Z Score >= 1

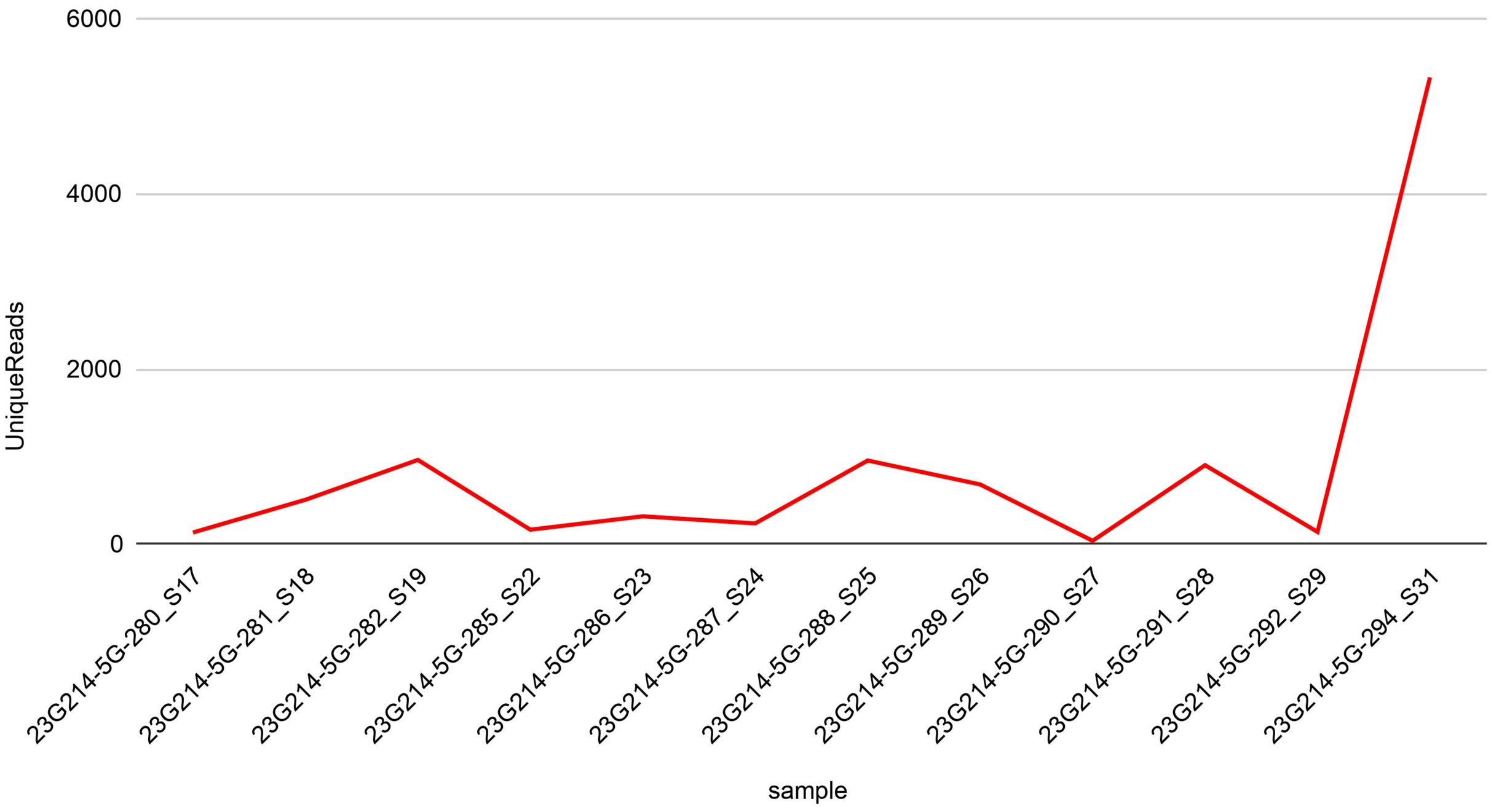
Taxon	Mapped Reads	GC Content	SNPs	%id	Informative Nucleotides	% Genome Called	Missing Bases	Ambiguous Bases
Lumpy skin disease virus	12441	27.7%	1	100%	2271	1.5%	68280	10

NCBI Reference	AF409137.1	Reference Length	150793	Coverage Depth	10.0x	Coverage Breadth	1.5%





UniqueReads (Centrifuge) vs. sample



rPM (Czid) vs. sample