

Genomic characterization of four novel *Staphylococcus myoviruses*

Joseph M. Ochieng Oduor^{1,2*}, Saija Kiljunen¹, Ermir Kadija³, Marianne W. Mureithi², Atunga Nyachio⁴, Mikael Skurnik^{1,5}

¹ Department of Bacteriology and Immunology, Medicum, and Research Programs Unit, Human Microbiome Research Program, University of Helsinki, Finland,

² KAVI-Institute of Clinical Research and Department of Medical Microbiology, University of Nairobi-Kenya,

³ University "Luigj Gurakuqi" Shkoder-Albania,

⁴ Institute of Primate Research (IPR), Nairobi-Kenya,

⁵ Division of Clinical Microbiology, Helsinki University Hospital, HUSLAB, Helsinki, Finland

Joseph M. Ochieng Oduor: <https://orcid.org/0000-0002-0315-4051>

Saija Kiljunen: <https://orcid.org/0000-0003-0461-7270>

Ermir Kadija: <https://orcid.org/0000-0003-1150-0011>

Marianne W. Mureithi: <https://orcid.org/0000-0001-9119-3167>

Atunga Nyachio: <https://orcid.org/0000-0002-7906-5311>

Mikael Skurnik: <https://orcid.org/0000-0001-8791-9260>

Corresponding author:

Joseph M. Ochieng Oduor, josemislredo@gmail.com,

Abstract: We report here the annotation of the complete genomes of four novel lytic *Staphylococcus* phages; Stab20, Stab21, Stab22 and Stab23. These phages have double-stranded DNA genomes ranging between 153338 - 155962 bp in size with terminal repeats of 10814 – 12304 bp. The genome analysis suggests that they are new phage species within genus *Kayvirus* in the *Spounavirinae* subfamily of the *Myoviridae* family.

Introduction

Application of phages as therapeutic and biocontrol agents has once again caught the global attention. However, these centennial antibacterial agents also face resistance from bacteria. The bacteria are capable of developing resistance against phages just as they do against antibiotics [1-2]. Fortunately, the environment is a rich source of phages and it is easy to isolate new phages active against the phage-insensitive bacterial strains [3] and, therefore, the constant search for novel lytic phages of therapeutic and bio-control potential is justified.

Lytic phages Stab20, Stab21, Stab22 and Stab23 were isolated from sewage and river samples from Shkoder, Albania, using a natural sausage fermenter starter *Staphylococcus xylosus* as the host bacteria. Phage DNA was isolated using the phenol-chloroform procedure and purified using a commercial kit (Invisorb Spin Virus DNA Mini Kit [4]. The DNA samples were sequenced using Next Generation Sequencing (NGS) technology platform [5]. The paired-end 150 bp long reads were *de novo* assembled by A5-miseq pipeline to generate sequence scaffolds [6]. The termini of the phage genomes were identified using PhageTerm [7]. Preliminary annotation of the genome was carried out with the Rapid Annotation Subsystems Technology (RAST) [8] and confirmed manually using Artemis [9]. The predicted genes and gene products were further analyzed with BLASTP, smartBLAST, PSI-BLAST [10], HHpred, HMMERscan [11] and InterProScan [12]. The similarity of the phage genomes was visualized graphically with Mauve [13]. Identity with related phages was established pairwise using EMBOSS stretcher [14]. tRNA encoding genes were predicted with ARAGORN [15] and tRNAscan-SE [16]. Phylogenies using whole genome and predicted major tail sheath and DNA primase amino acid sequences were calculated with VICTOR [17] and Phylogen fr -“One Click” [18].

The analysis of the phage genomes revealed that they are related but distinct from each other, and most closely affiliated with genus *Kayvirus* in subfamily *Spounavirinae* of the *Myoviridae*. The characteristic properties of the phage genomes are summarized in Table 1 and in the Mauve alignment (Figure 1). The linear double stranded DNA genomes ranged between 153,338 and 155,962 bp in size with long terminal repeats ranging between 10,814 and 12,225 bp. The genomes of the phages are 67.4-81.6 % identical to that of phage K of the genus *Kayvirus* (Table 1). Phylogenetic

analysis further illustrates that Stab20, Stab21, Stab22 and Stab23 (Figure S1) qualify to be classified as the new species in genus *Kayvirus*.

Homology search identified putative functions for about 75 predicted gene products in all the phages. Capsid and scaffold protein, portal protein, prohead protease; membrane protein; tail tube, major tail sheath, tail morphogenetic, tail tape measure proteins, baseplate proteins, and carbohydrate binding domain-containing tail proteins were the structural proteins identified. The other likely non-structural proteins included DNA primase, helicase, exonuclease, DNA polymerases (A/I and B/II), RNA polymerase sigma factor, integration host factor, thioredoxin, ribonucleotide reductase large and small subunits, resolvase, ribonucleotide reduction protein, replication protein, nucleoside 2-deoxyribosyltransferase, RNA ligase, ribonuclease H, and a transcriptional regulator. Most of these nonstructural proteins belong to the DNA/RNA metabolism-oriented proteins. In addition, holins, endolysin, and CHAP domain containing proteins were identified. However, no integrase or other lysogenic associated genes were identified from these genomes. The genomic locations of some of the genes, including the genes encoding tRNAs, adsorption associated tail protein, Ig-like domain containing protein, major tail sheath protein and DNA primase, are indicated in Figure 1.

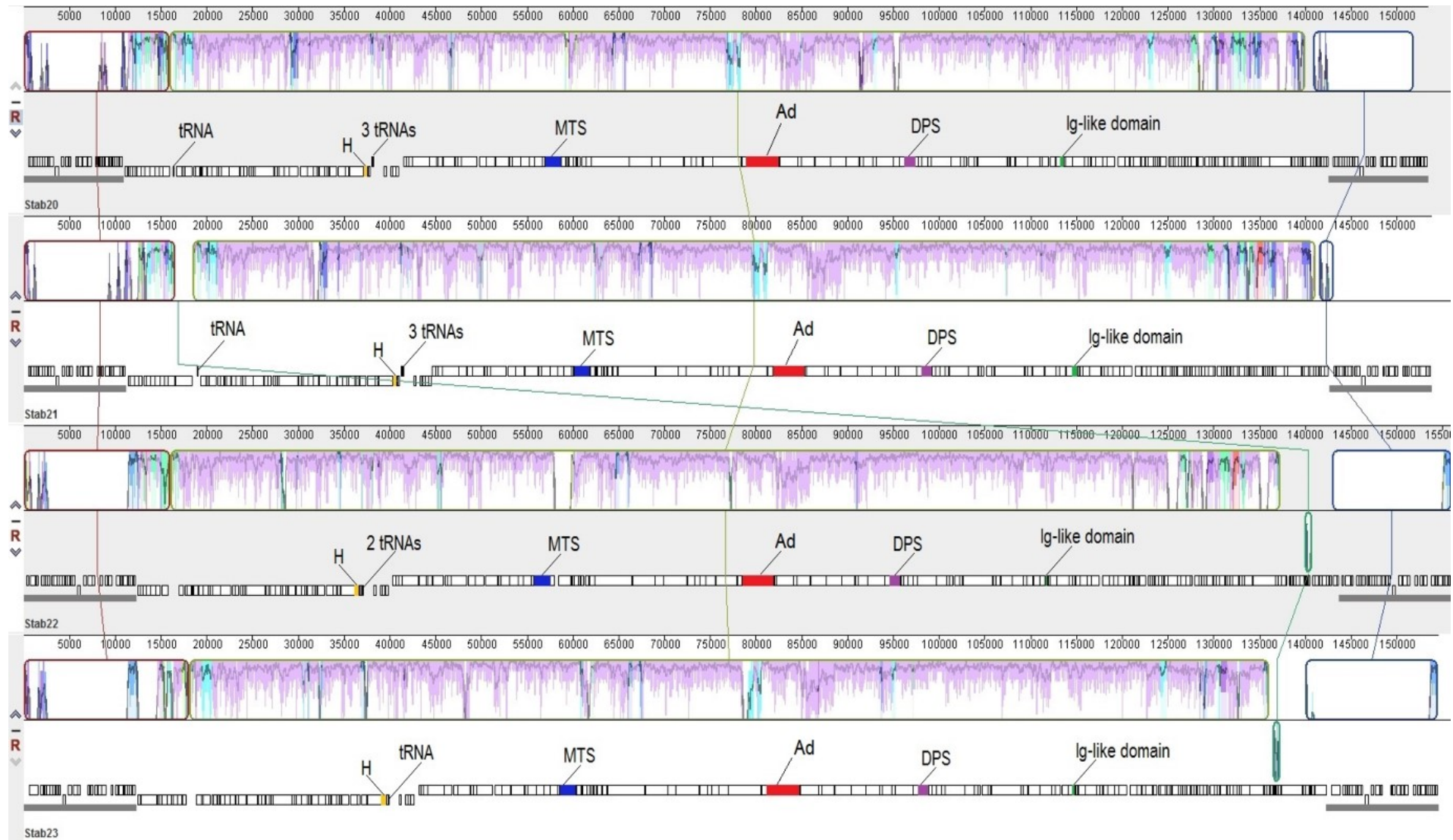
The genome sequences of *Staphylococcus* phages Stab20, Stab21, Stab22 and Stab23 have been submitted to the European Nucleotide Archives and their accession numbers are xxxx, xxxx, xxxx, and xxxx, respectively (pending).

ACKNOWLEDGEMENTS

The study has been supported by research grants from Finnish National Agency for Education (CIMO), International Society for Infectious Diseases and the European Society of Clinical Microbiology and Infectious Diseases (ISID\ESCMID) and Jane and Aatos Erkkö Foundation.

Reference

- [1] R. T. Schooley *et al.*, “Development and Use of Personalized Bacteriophage-Based Therapeutic Cocktails To Treat a Patient with a Disseminated Resistant *Acinetobacter baumannii* Infection,” *Antimicrob. Agents Chemother.*, vol. 61, no. 10, 2017.
- [2] S. J. Labrie, J. E. Samson, and S. Moineau, “Bacteriophage resistance mechanisms,” *Nat. Rev. Microbiol.*, vol. 8, no. 5, pp. 317–327, May 2010.
- [3] X. Wittebole, S. De Roock, and S. M. Opal, “A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens,” *Virulence*, vol. 5, no. 1, pp. 226–235, Jan. 2014.
- [4] “Invisorb Spin Virus DNA Mini Kit - Stratec.” [Online]. Available: <https://www.molecular.stratec.com/products/details/invisorb-spin-virus-dna-mini-kit>. [Accessed: 14-Nov-2018].
- [5] “Eurofins GATC Biotech,” *Eurofins Scientific*. [Online]. Available: <https://www.eurofins.com/contact-us/worldwide-interactive-map/germany/eurofins-gatc-biotech/>. [Accessed: 08-Jan-2019].
- [6] F. Prestinaci, P. Pezzotti, and A. Pantosti, “Antimicrobial resistance: a global multifaceted phenomenon,” *Pathog. Glob. Health*, vol. 109, no. 7, pp. 309–318, Oct. 2015.
- [7] J. R. Garneau, F. Depardieu, L.-C. Fortier, D. Bikard, and M. Monot, “PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data,” *Sci. Rep.*, vol. 7, no. 1, p. 8292, Aug. 2017.
- [8] R. K. Aziz *et al.*, “The RAST Server: rapid annotations using subsystems technology,” *BMC Genomics*, vol. 9, p. 75, Feb. 2008.
- [9] T. Carver *et al.*, “Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database,” *Bioinforma. Oxf. Engl.*, vol. 24, no. 23, pp. 2672–2676, Dec. 2008.
- [10] “BLAST: Basic Local Alignment Search Tool.” [Online]. Available: <https://blast.ncbi.nlm.nih.gov/Blast.cgi#>. [Accessed: 14-Nov-2018].
- [11] L. Zimmermann *et al.*, “A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core,” *J. Mol. Biol.*, vol. 430, no. 15, pp. 2237–2243, Jul. 2018.
- [12] P. Jones *et al.*, “InterProScan 5: genome-scale protein function classification,” *Bioinformatics*, vol. 30, no. 9, pp. 1236–1240, May 2014.
- [13] A. E. Darling, B. Mau, and N. T. Perna, “progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement,” *PLOS ONE*, vol. 5, no. 6, p. e11147, Jun. 2010.
- [14] “EMBOSS Stretcher < Pairwise Sequence Alignment < EMBL-EBI.” [Online]. Available: https://www.ebi.ac.uk/Tools/psa/emboss_stretcher/. [Accessed: 14-Nov-2018].
- [15] D. Laslett and B. Canback, “ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences,” *Nucleic Acids Res.*, vol. 32, no. 1, pp. 11–16, 2004.
- [16] T. M. Lowe and P. P. Chan, “tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes,” *Nucleic Acids Res.*, vol. 44, no. Web Server issue, pp. W54–W57, Jul. 2016.
- [17] J. P. Meier-Kolthoff and M. Goeker, “VICTOR: Genome-based Phylogeny and Classification of Prokaryotic Viruses,” *bioRxiv*, p. 107862, Feb. 2017.
- [18] A. Dereeper *et al.*, “Phylogeny.fr: robust phylogenetic analysis for the non-specialist,” *Nucleic Acids Res.*, vol. 36, no. Web Server issue, pp. W465–469, Jul. 2008.



Adsorption-associated tail protein (Ad)

Major tail sheath protein (MTS)

Ig-like domain containing protein

DNA primase (DPS)

Terminal_repeat region

Holin (H)

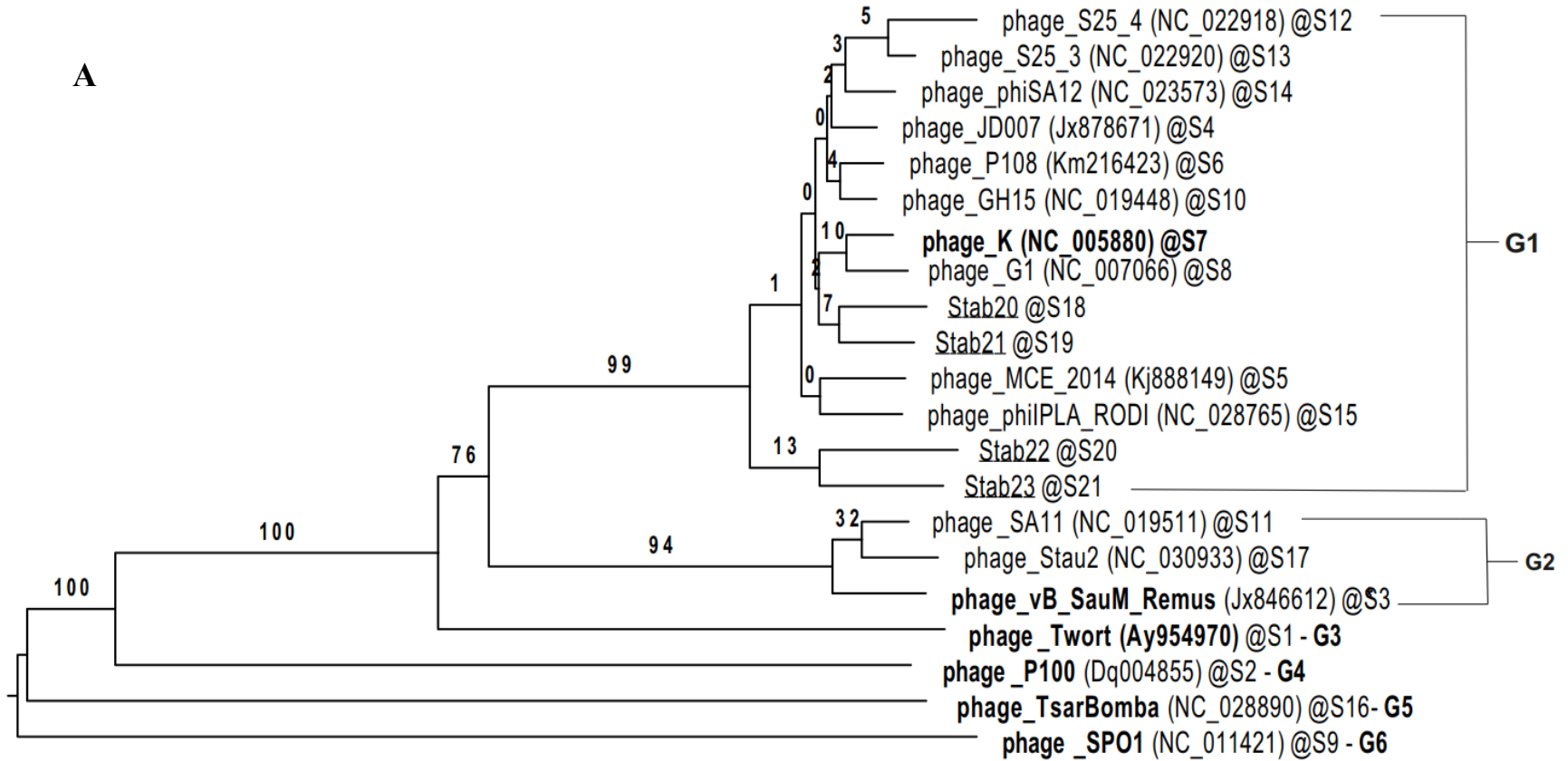
Figure 1: Mauve alignment of annotated complete genomes of Stab20, Stab21, Stab22 and Stab23 from top to bottom; showing the locations of tRNAs, and the genes for proteins such as major tail sheath protein (blue), holin (yellow), DNA primase (purple), Ig-like domain containing protein (green) and adsorption-associated tail protein (red) within the genomes. The terminal-repeat regions are indicated by thick grey bars. The mauve plots show conserved genomic regions of the four phages. The similarity levels vary with the heights of the curves and the intensities within the blocks that are proportional to the average nucleotide identities. The white spaces inside or outside the blocks represent regions of difference between the genomes of these phages.

Table 1: The properties and whole genome sequence comparisons of Stab20, Stab21, Stab22, Stab23 and *Staphylococcus* phage K. The identities were determined by EMBOSS stretcher.

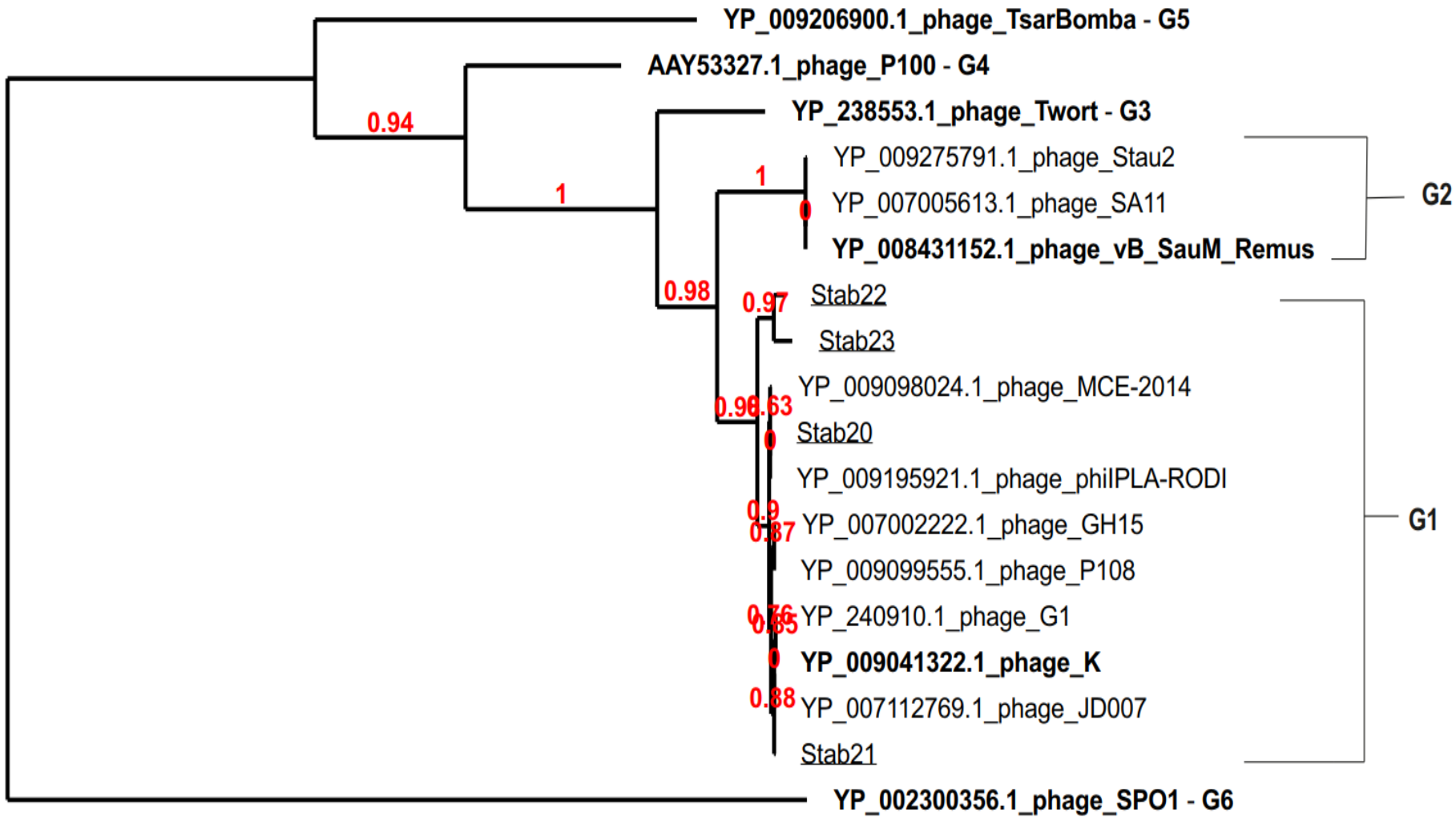
| Properties/ Identity % | <i>Staphylococcus</i> phages | | | | |
|---------------------------|------------------------------|--------------|--------------|--------------|--------------|
| | Stab20 | Stab21 | Stab22 | Stab23 | Phage K* |
| Genome size (bp) | 153338 | 153797 | 155962 | 154499 | 148317 |
| Terminal repeat size (bp) | 10814 | 11149 | 12304 | 12225 | 8486 |
| GC content % | 30.21 | 30.32 | 30.61 | 30.88 | 30.4 |
| Number of predicted genes | 223 | 217 | 218 | 206 | 233 |
| Number of tRNAs | 4 | 4 | 2 | 1 | 4 |
| Stab20 identity % | 100.0 | 84.4 | 49.7 | 49.6 | 81.6 |
| Stab21 identity % | 84.4 | 100.0 | 73.4 | 76.9 | 76.2 |
| Stab22 identity % | 49.7 | 73.4 | 100.0 | 77.5 | 72.2 |
| Stab23 identity % | 49.6 | 76.9 | 77.5 | 100.0 | 67.4 |
| Phage K identity % | 81.6 | 76.2 | 72.2 | 67.4 | 100.0 |

* Phage K refers to *Staphylococcus* phage K that is type species recognized by the International committee on taxonomy of viruses (ICTV).

A



B



C

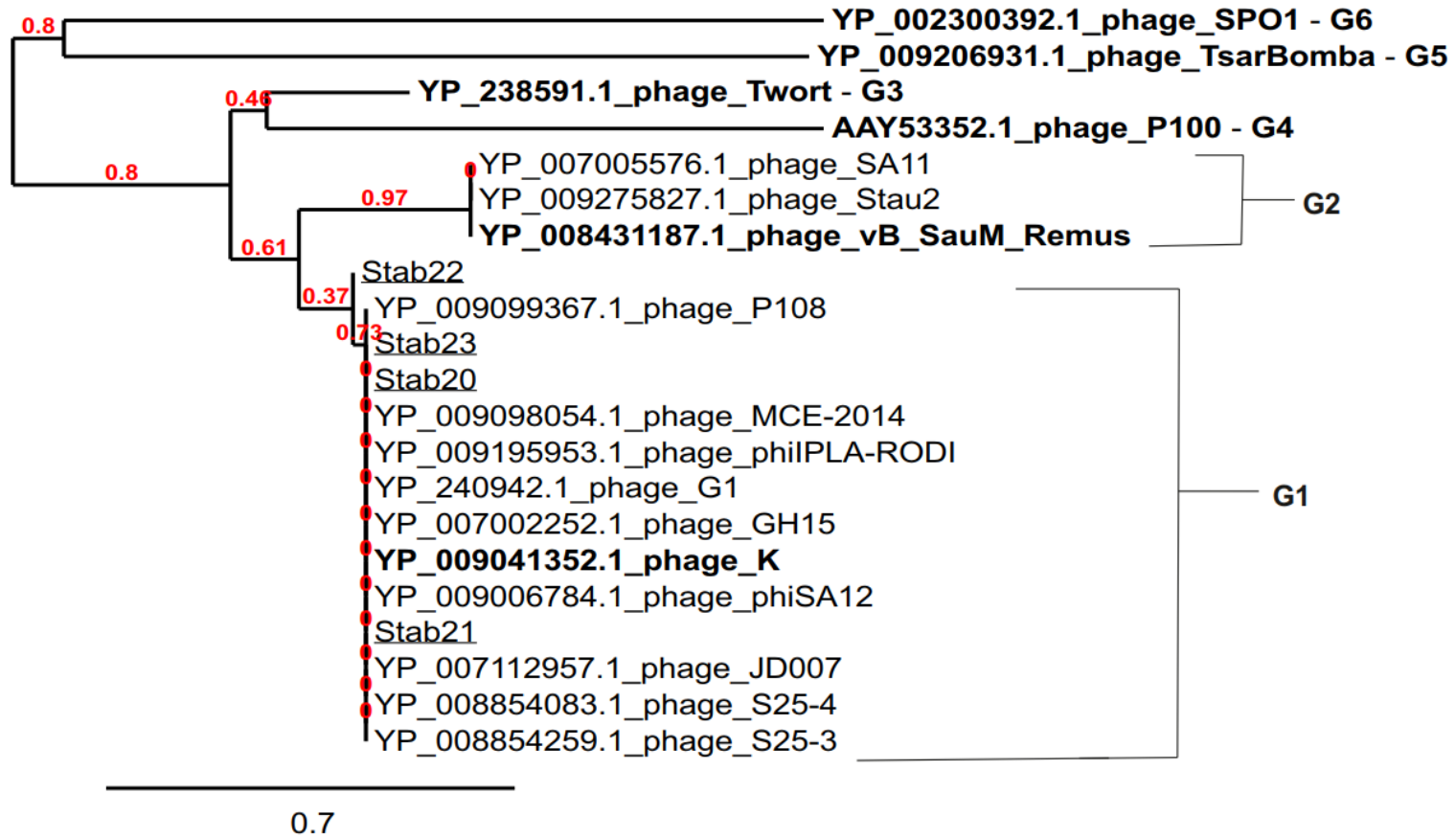


Figure S1: These are phylogenetic trees of the *Spounavirinae* subfamily illustrating the Stab phages genus. Phylogenetic trees were created with whole genomes (**A**), predicted tail sheath (**B**) and DNA primase amino acid sequences (**C**) of the Stab phages and representatives of closely related phages selected from the ICTV database. The branch length is proportional to the number of substitutions per site. The abbreviations **G1** – genus *Kayvirus*; **G2** - genus *Silviavirus*; **G3** - genus *Twort*; **G4** – genus *P100virus*; **G5** – genus *Tsarbombavirus*; **G6** – genus *Spolvirus*; and (**S**) – species.