bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Phylogenomics of novel clones of *Aeromonas veronii* recovered from a freshwater lake reveals unique biosynthetic gene clusters

Opeyemi U. Lawal[1], Noah Bryan[1,2], Valeria R. Parreira[1], Rebecca Anderson[1], Yanhong Chen[1], Melinda Precious[1], Lawrence Goodridge[1*]*

[1] Canadian Research Institute for Food Safety (CRIFS), University of Guelph, Guelph, ON, Canada N1G 2W1. [2] Bayview Secondary School, 10077 Bayview Ave, Richmond Hill, ON L4C 2L4.

*Email correspondence to Dr. Opeyemi U. Lawal – lawal@uoguelph.ca & Dr. Lawrence Goodridge – goodridl@uoguelph.ca

ABSTRACT

Aquatic ecosystems are important reservoirs for clinically relevant pathogens and antimicrobial resistance genes, thus present a significant risk to global health. Here, we assessed the phylogenomics of *Aeromonas veronii* (*A. veronii*) recovered from Lake Wilcox in Ontario using a combination of morphological, biochemical, and whole-genome sequencing (WGS) techniques.

Eleven distinct bacterial colonies were isolated and identified as *A. veronii* (n=9), and two other *Aeromonas* species (*A. caviae* and *A. allosaccharophila*), with significant discrepancies noted between biochemical and WGS identification methods. Of note, 67% (n=6/9) of *A. veronii* isolates were human pathogens (Pathogenicity score ≥ 0.50). The genomic analysis revealed high genetic diversity among the *A. veronii* isolates, including the discovery of 41 novel alleles and seven new sequence types (ST) suggesting the lake as a reservoir for multiple human pathogenic clones of this bacterium. The comparison of the newly isolated and sequenced *A. veronii* with 214 *A. veronii* genomes revealed significant genetic diversity and suggests potential broad geographical dissemination of strains. Chromosomal genes (*OXA-912* and *cphA* [*cphA3, cphA4, cphA7*]) genes encoding resistance to β-lactamases were detected in all isolates. Human and non-human pathogenic strains of *A. veronii* differed in their virulence gene content, with type III secretion systems being associated with human pathogenic isolates. Mobilome analysis revealed the absence of plasmids in *A. veronii* isolates and the presence of 13 intact the great majority of which were P22-like (Peduoviridae) phages, and nine different insertion sequence families. Novel biosynthetic gene clusters were identified and characterized, indicating the potential for unique secondary metabolite production in *A. veronii* with different pathogenic potential. Overall, this study underscores the importance of continuous surveillance of aquatic ecosystems for the presence of pathogens, contributing to our understanding of their evolution, potential for human pathogenicity, and the ecological roles of their genetic elements.

Keywords: *Aeromonas*, genomics, antimicrobial resistance, biosynthetic gene cluster, water quality, freshwater lake, public health

BACKGROUND

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The role of aquatic ecosystems as reservoirs for clinically relevant pathogens and antimicrobial resistance genes (ARG) has recently gained attention as the importance of assessing the quality of these ecosystem is paramount in public health (1, 2). Freshwater bodies like lakes and rivers used for recreational purposes can significantly impact the health of communities (3, 4). Poor water quality in these settings poses a substantial risk for the transmission of various waterborne diseases, including pathogenic viruses, protozoa, and bacteria including *Aeromonas* species that thrive in such contaminated water (2–4).

Aeromonas species are Gram-negative, facultative anaerobic rods, found in various aquatic environments (5–7), and known for their ability to survive in diverse environments, ranging from freshwater to the intestinal tracts of animals (5, 7). While some *Aeromonas* species including *Aeromonas salmonicida*, *Aeromonas hydrophila*, and *Aeromonas veronii* are known fish pathogens, *A. veronii* is one of the four species that are considered as potential human pathogens (8–10). *A. veronii* is an emerging human pathogen causing a wide range of diseases in human and animals including gastroenteritis, respiratory and skin infections and septicemia (9–11). In addition, *A. veronii* is increasing being recognized as a significant concern to food safety due to its frequent presence in different types of food, particularly in minimally processed ready-to-eat seafood (12, 13). Of note, the frequent and global occurrence of highly virulent strains of *A. veronii* has been detected in food samples such as meat, milk, catfish and fish in countries including Brazil (13), Egypt (14), India (15), Israel (11), and the United States (16, 17), among others. The adaptability of *A. veronii* to various conditions poses a challenge for water quality management, especially in environments with high anthropogenic activities, where the bacterium can be a potential source of infection (5, 18).

The mechanisms of pathogenicity of *A. veronii* involve the production of various toxins and virulence factors that contribute to its ability to infect host cells and cause disease (5, 18). A significant concern with *A. veronii* is its capacity for antimicrobial resistance (7, 19, 20). The presence of antimicrobial-resistant strains in aquatic environments is a public health concern, as it not only affects the treatment of *Aeromonas*-related infections but also represents a potential reservoir for the spread of resistance genes to other pathogenic bacteria (7, 19, 20). Studies on the population structure of *A. veronii* have described genetic diversity driven by its adaptability to various environmental conditions. These factors could drive variability in strains regarding pathogenicity and resistance to environmental stresses in this bacterium, with practical implications for public health and water management (5, 8, 18).

In recent years, advancements in sequencing technologies have greatly enhanced the genomic surveillance of known and emerging pathogens, such as *A. veronii*, across different environmental matrices(5). Despite these technological advancements, little importance has been given to *A. veronii,* especially in terms of its presence in freshwater, its impact on water quality, and its role in the dissemination of antimicrobial resistance (AMR) in both the environment and the food chain. Understanding the genomic surveillance and population structure of this bacterium is crucial for developing effective infection treatment strategies and ensuring public health safety.

We have previously reported the detection of clinically relevant pathogens in Lake Wilcox, including novel strains of *Bacillus anthracis* (21), and *Vibrio cholerae* (22) isolated at different time points. In

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

this study, we employed a combination of culture-based detection and whole-genome sequencing to assess the presence of *A. veronii*, its extensive genomic fingerprint, its population structure, and the genomic characterization of stress response genes in Lake Wilcox. The genetic relatedness of *A. veronii* isolates was assessed by comparing them with previously sequenced strains in public databases using a comparative genomic approach.

METHODS

Description of sampling site. Lake Wilcox is a small kettle lake located in Richmond Hill in Ontario (43º56'56.69" N, 79º26'9.45" W). Historically, the lake is used for recreational purposes by the surrounding community and tourists. Despite being impacted by feces of surrounding wildlife, recreational activities have continued, and users have reported skin rashes and gastrointestinal symptoms after recreational activities (https://projectboard.world/ysc/project/the-phage-takes-centre-stage-for-water-quality-testing) **(Figure 1)**.

Sample collection and processing. Freshwater samples were obtained from Lake Wilcox in the Summer of 2022 and Fall of 2023. Water samples were kept at 4° C and analyzed within 48 hours of collection. Samples were processed as described by Bryan et al, 2023 (23). Briefly, 1 mL of samples was serially diluted in 9 mL of lambda buffer (modified saline-magnesium buffer without gelatin) and plated onto tryptic soy agar (TSA). Following incubation for 24 h at 37° C, plates were analyzed for bacterial colonies. Distinct colonies of differing morphologies were sub-cultured onto TSA to obtain pure culture. The isolated colonies were *Gram* stained and taxonomic identification was performed using VITEK® (bioMérieux, Inc, Canada).

Genomic DNA extraction and whole-genome sequencing. Genomic DNA from isolated colonies was extracted using the DNeasy blood and tissue kit (Qiagen Hilden, Germany) according to the manufacturer's instructions. DNA libraries were prepared using the Illumina DNA prep tagmentation kit (#20018704) and IDT for Illumina DNA/RNA UD indexes (#20027213) following the manufacturer's instructions. Paired-end (2 x 150 bp) sequencing was performed using the high output flow cell on the Illumina MiniSeq instrument as described previously (23, 24).

Genome assembly and annotation. Raw paired-end reads were quality filtered using FastQC v0.11.9 (https://github.com/s-andrews/FastQC), and trimmed using Trimmomatic v0.39 (25). High quality reads with a Phred quality score above 20 were assembled *de novo* using the Skesa v2.4.0 pipeline (26). Assembly quality and genome completeness were assessed using QUAST v5.2 (27) and BUSCO (28), respectively. Genome annotation was performed using Prokka v1.14.6 (29).

Gene content analysis. The antimicrobial resistance gene profile of all the isolates was determined using AMRFinder Plus v3.10.45 (30) and CARD (31) databases, while the virulence genes were identified using VFDB (32, 33). To define the mobile genetic elements of the collection, the draft genomes were screened for plasmids and prophages using MOB-suite v3.1.6 (34) and PHASTEST (35), respectively. Biosynthetic gene clusters were assessed using the antiSMASH v6 pipeline (36).

Phylogenetic analysis. To construct the phylogeny, pangenome was generated from the annotated genomes using Roary v3.13.0 (37), core genome-based phylogenetic tree was constructed using

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

FastTree (38). The general time reversible model was performed with 1000 bootstrap resampling for node support. Except as otherwise stated, all bioinformatics tools were executed using the default settings.

Data availability. The whole-genome sequences reported in this study were deposited at DDBJ/ENA/GenBank under the BioProject accession numbers **PRJNA893208**. The raw sequence reads, and genome assembly accession numbers are listed in **Table 1**. In addition, accession numbers and associated metadata of genomes retrieved from NCBI are listed in **Table S3**.

RESULTS

Bacterial species were isolated from the freshwater samples recovered from a freshwater lake over a period of one year using the spread agar plate method. Overall, 11 distinct colonies with different morphologies were selected and further characterized using morphological, biochemical-based, and whole-genome sequencing. Taxonomic identification using the VITEK® Compact system identified the colonies as *Aeromonas sobria* (n=8), *Aeromonas hydrophila/punctata* (n=2), and one isolate with an inconclusive result **(Table 1)**. Sequencing of the 11 isolates yielded 1,024,248 - 2,725,402 paired-ended reads per isolate **(Table 1)**. Using pubMLST and rMLST (39), as well as k-*mer*-based species taxonomic classification with the Kraken2 database (40), isolates were identified as *Aeromonas veronii* (n=9), *A. caviae* (n=1) and *A. allosacharophila* (n=1) **(Table 1)**. The average nucleotide identity (ANI) analysis with fastANI (41) showed that the nine *A. veronii* strains had >96% ANI when *A. veronii* GCF_000820225.1 strain was used as a reference, *A. caviae* strain NB-180 had 97.9% ANI with *A. caviae* GCF_000819785.1, while *A. allosaccharophila* had 96.21% ANI with the reference strain *A. allosaccharophila* GCF_000819685.1. The draft genomes of *Aeromonas* species yielded between 28 and 113 contigs, with a G+C content of 58 - 59 %, except for *A. caviae* that had a higher G+C content of 61.26%, a value that was comparable to the reference strain *A. allosaccharophila* GCF_000819685.1. The genome size was comparable between the three *Aeromonas* species identified and ranged between 4,390,436 and 4,690,056 bp, with >50× genome coverage **(Table 1**).

Prediction of human pathogenicity of *A. veronii* sequenced. Considering that Aeromonas species are commonly associated with diseases in fish, we evaluated the potential of these isolates to be pathogenic to humans. We did this by comparing the proteins of the new strains with a database composed of protein families associated with either pathogenic or non-pathogenic organisms in humans, using the PathogenFinder tool. (42). Six out of the nine *A. veronii* isolates in this study had a pathogenicity score greater than 0.5, suggesting that they may be pathogenic to humans. Other isolates, including *A. caviae* and *A. allosaccharophila*, were predicted to be non-human pathogens **(Table 1)**.

Population structure of *A. veronii* isolated from freshwater. To assess the genetic relatedness among isolates sequenced in this study, a combination of conventional MLST and whole genome-based phylogeny was employed. The *Aeromonas* MLST schema was used to determine the sequence types (STs) of all isolates. Of note, 41 novel alleles were identified among the 11 *Aeromonas* isolates and yielded nine unique allele profiles that were submitted together with the allele sequences and

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

assigned to nine new STs (ST2530 – ST2538) **(Table 1 and Table S1)**. Two STs (ST2530 and ST2535) contained two isolates each while others were singletons suggesting the uniqueness of the isolates understudy and high genetic diversity in the population. The core genome SNP-based phylogeny of the nine *A. veronii* sequenced was constructed to using the complete closed genome of *A. veronii* AP022281.1 as reference. *A. caviae* and *A. allosaccharophila* were used as outgroups to root the tree. Isolates were grouped into two main clusters irrespective of the period of isolation **(Figure 2)**. Isolates were distantly related by SNPs with ≥100 SNPs difference **(Table S2)** except for a pair of isolates from different timepoints (NB-2/NB-4, Summer, 2022; and NB-178/NB-181, Fall, 2023) that were highly related differing only by 9 and 11 SNPs, respectively **(Figure 2)**. Of note, the SNP-based clustering observed was similar to the MLST-based population structure suggesting a good concordance between these methods for typing *A. veronii*. Overall, the high genetic diversity observed in this study suggest that the freshwater lake could serve as a reservoir for multiple strains of *A. veronii* that are pathogenic to humans.

Global population structure of *A. veronii*. To assess the genetic relatedness of the sequenced isolates with global *A. veronii*, genomes and the associated metadata of 214 *A. veronii* in the *RefSeq* database (accessed on October 15, 2022) were downloaded and re-annotated (*see Method*). The 214 genomes were recovered from 18 different countries located in six continents between 1988 to 2022 from seven different sources including human, animal, aquatic ecosystem, fish, food, plant, and insect **(Table S3).** The pangenome size of the 214 *A. veronii* genomes together with the sequenced isolates (n=9) yielded 49,483 genes. A total of 2,248 core genes, defined as genes present in ≥95% of the genomes in the collection, were identified, whereas the shell and cloud genes totalled 2,294, and 44,941, respectively. The core genome-based maximum likelihood tree based on 2,248 core genes was constructed using *A. allosaccharophila* as outgroup to root the phylogenetic tree. The sequenced *A. veronii* isolates compared with global *A. veronii* species showed high genetic diversity which facilitated the clustering of the isolates into distinct clades **(Figure 3)**.

Isolates sequenced in this study were clustered into distinct subclades suggesting that they are distantly related to other global isolates. However, isolates from Turkey and Greece recovered in fish from different time points (2009, 2015 and 2016) were clustered together, a phenomenon that could suggest dissemination of *A. veronii* strains. Strain NB-188 belonged to the same subcluster as an isolate recovered from a similar aquatic system in France. Global *A. veronii* are distantly related. While some subclades were source-based, we could still observe a mixture of isolates from different countries and sources within others. Overall, *A. veronii* from different environments may have genetic signatures unique to pathogenic strains of this bacterium. This could also be important to determine or predict the source of isolates found in any matrix. Future evolutionary studies on this bacterium would be important to decipher this hypothesis.

Stress response genes among the *A. veronii* sequenced. Genes encoding resistance to antimicrobials and virulence were assessed in the sequenced isolates. Genes encoding resistance to β-lactams were detected in all the isolates. Different alleles of *cphA* (*cphA3, cphA4,* and *cphA7*) gene which belonged to the subclass B2 metallo-beta-lactamase that encodes resistance to carbapenem antibiotics were detected in all *A. veronii* isolates and in *A. allosaccharophila*. *OXA-912* that encodes resistance to penams, cephalosporins and carbapenems, and *cphA3* genes were

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

predominant in the collection. These data suggest that *A. veronii* recovered from freshwater sources are a reservoir for β-lactamase resistance genes. Additionally, detection of virulence genes using a gene homology approach and a curated virulence gene database [VFDB (32, 33)] detected eight to 48 virulence genes in each sequenced *A. veronii* isolate. Relative to isolates predicted as human pathogens that contained 41 – 48 virulence genes (except NB-3), all the *A. veronii* predicted as non-human pathogens carried less virulence genes (≤9 virulence genes). Human and non-human pathogenic strains of *A. veronii* differed in term of their virulence gene content. While the flagellar and type IV pili associated genes involved in biofilm formation (18, 33) were detected in all isolates, type III secretion system (T3SS) associated genes were detected only in the isolates predicted as human pathogens **(Figure 2)**.

Characterization of the mobile genetic elements among the sequenced *A. veronii*. Plasmids were not detected in any of the *A. veronii* isolates studied. However, *A. caviae* carried a small plasmid (3,976 bp, contig 44) that contains plasmid replication genes and uncharacterized proteins. Comparative sequence analysis with blast showed that the plasmid had the closest nucleotide sequence similarity (>99% coverage and identity) to *Aeromonas enteropelogenes* 9789_1_48 plasmid (LT635650.1). The detection and characterization of phage regions in the sequenced genomes yielded 13 unique intact phages among which four were predicted to be virulent phages (43). The completeness of all intact phage sequences was determined to be between 50 – 100% by CheckV (44). The phages were classified by PhaGCN (45) as Peduoviridae (n=10), and Chaseviridae (n=1) and two others unidentified according to the International Committee on Taxonomy of Viruses (ICTV) classification (46). In addition to *A. veronii* being predicted as host of the phages, other species of *Aeromonas* (*A. australiensis*, *A. diversa*, *Aeromonas* sp.) and *Serratia marcescens* could also serve as their hosts as determined by PhaBox (47, 48), suggesting that these phages could infect multiple hosts **(Table 2)**. Of note, the two pairs of isolates (NB-2/NB-4 and NB-178/NB-181) that were highly genetically related by SNP had the same phage content. No antibiotic resistance, toxin or related genes were detected in the intact phages. All the intact phages detected were screened for tailspike proteins (TSP) using TSPDB that contains 8,077 TSPs (49), but none was found. In addition to phages, other MGE identified in the collection include 16 different types of insertion sequence (IS) elements belonging to nine IS families **(Figure 4)**. IS*As19* (IS*481*), ISAs4 (IS*5*) were predominant in the sequenced *A. veronii*. The IS*4* subtype containing IS*Aeme13*, IS*As30*, IS*Apu1* was detected in isolates recovered from early sampling dates (Summer, 2022), whereas IS subtypes IS*630* (IS*Ahy2*), IS*200*/IS*605* (IS*As26*) and IS*21* (IS*As29*) were observed in isolates recovered in the Fall of 2023, suggesting a temporal distribution of IS elements in *A. veronii*.

Biosynthetic gene cluster profile in *A. veronii*.

The 53 biosynthetic gene clusters (BGCs) identified in all sequenced *Aeromonas* species were categorized into eight BGC families using sequence similarity network analysis with BiGSCAPE (50). The ribosomally synthesized and post-translationally modified peptides (RiPPs) were the most predominant BGC class, consisting of three gene families. In contrast, the non-ribosomal polyketide synthase (NRPS) included only one gene family. The remaining gene families were classified as "others" and included homoserine lactone (n=2) and aryl polyene (n=2). The three RiPPs detected were unique and conserved within the collection but exhibited low similarity scores to previously

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

described BGCs. For example, RiPP-1 (Figure 5a), comprising 11 open reading frames (ORFs), had a similarity score of 0.17 to angustmycin A/B/C (BGC0002621) described in *Streptomyces angustmyceticus* (Accession MZ151497.1) (51). Meanwhile, RiPP-2 and RiPP-3 (Figure 5b, Figure 5c), consisting of nine and seven ORFs respectively, had similarity scores of ≤0.08 to pseudopyronine A/B (BGC0001285) described in *Pseudomonas putida* (Accession KT373879.1) (52). Notably, RiPP-3 was also detected in *A. allosaccharophila* (NB-7), indicating that this BGC is not exclusive to *A. veronii* (Figure 5c). The identified NRPS had the highest similarity score of 0.9 to enterobactin (BGC0000343) previously described in *Pseudomonas* sp. J465 (Accession GQ370384.1) (53). This BGC was conserved in the *A. veronii* sequenced (Figure 5d). Further analysis of global *A. veronii* genomes confirmed that this BGC was conserved not only in this collection but also in all publicly available *A. veronii* genomes. A BGC encoding homoserine lactone, predominant in *A. veronii* (n=6/9), was also detected in *A. allosaccharophila*. This BGC had a low similarity score (0.14) to thioguanine (BGC0001992) in *Erwinia amylovora* CFBP1430 (Accession number: NC_013971.1) (54). Notably, a pair of *A. veronii* strains—one pathogenic (NB-6/NB-187) and one non-pathogenic to humans (NB-178/NB-181)—carried unique BGCs encoding aryl polyene (Figure 6a-b). The pathogenic pair consisted of 17 ORFs with a similarity score of 0.44 to aryl polyene (BGC0002008) described in *Xenorhabdus doucetiae* (Accession NZ_FO704550.1) (55), while the non-pathogenic pair contained 37 ORFs with a similarity score of 0.26 to bacilysin (BGC0000888) described in *Bacillus* sp. CS93 (Accession number: GQ889493.1) (56). Overall, *A. veronii* harbored putative unique BGCs that exhibited low similarity scores to previously described compounds.

DISCUSSION

Aquatic ecosystems are continually impacted by anthropogenic activities, making the microbial quality and safety of these water bodies, especially those used for recreational activities, paramount for public health (1). In this study, we assessed the presence of *A. veronii* in a recreational lake and determined the extensive genomic features of the isolates regarding their population structure and the genomic characterization of stress response genes, mobile genetic elements, and other gene content such as biosynthetic gene clusters that confer uniqueness to different *A. veronii* strains. We inferred the global population structure of *A. veronii* by assessing the genetic relatedness of the isolates sequenced with previously sequenced strains in public databases.

In the past decade, WGS has become the gold standard method for species identification, complementing existing biochemical-based methods (57). In this study, the WGS-based approach identified isolates as *A. veronii, A. caviae*, and *A. allosaccharophila*, whereas the biochemical identification system misidentified all isolates as either *A. sobria* or *A. hydrophila*. Misidentification of species of environmental bacteria by biochemical approaches is not uncommon (57). Studies comparing biochemical-based bacterial species identification systems to WGS have shown that species misidentification can vary by species and is common in specific bacteria, including *Pseudomonas fluorescens*, *Pseudomonas putida* (57), and *Enterococcus faecalis* (58).

A. veronii strains sequenced exhibited different pathogenic potentials, with the majority (67%, n=6/9) predicted to be pathogenic to humans and possessing a similar virulence determinant profile, including T3SS. T3SSs are crucial virulence mechanisms that allow bacteria to inject effector

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

proteins directly into the host cell cytoplasm. The activity of T3SSs closely correlates with infection progression and outcome in various infection models, and its presence is considered a general indicator of virulence in *A. veronii* (59–61). The detection of human pathogenic *A. veronii* in this study, along with other clinically relevant pathogens such as *Bacillus anthracis* (21) and *Vibrio cholerae* (22) in this recreational lake from previous studies, emphasizes the crucial role aquatic ecosystems play in disseminating pathogens. The recreational use of this water could pose a continuous risk to public health, serving as a reservoir and facilitating the transmission of waterborne diseases. This also underscores the significance of monitoring aquatic environments as reservoirs for pathogenic bacteria.

There was high genetic diversity among the nine *A. veronii* isolates sequenced, including the identification of novel sequence types and alleles. While some strains were indistinguishable by SNPs, others were genetically distant. This finding could imply that different *A. veronii* strains may have been introduced into the lake multiple times from various sources such as resident freshwater fish, domestic animals, and environmental samples (5). The integration of genomic data from the *A. veronii* isolates sequenced with global strains revealed that isolates from single sites formed smaller groups within the phylogeny. Interestingly, one isolate from this study (NB-188) was nested with an isolate recovered from an aquatic ecosystem in France. A previous study assessing the core genome-based phylogenetic analysis of *A. veronii* genomes deposited in NCBI from 18 countries revealed high genetic diversity (5). The admixture of *A. veronii* strains from different sources was observed, suggesting a lack of source- and timepoint-based clustering in the *A. veronii* population. However, strains from a single site tend to form small groups within the phylogenetic clusters. These observations concur with our findings. The genetic diversity observed in *A. veronii* reinforces the importance of continuous genomic surveillance to monitor the emergence and spread of virulent and/or resistant strains.

The AMR determinant profile observed in the isolates sequenced in this study was comparable and included only chromosome-borne genes encoding resistance to beta-lactams. The prevalence of β-lactam resistance genes, including those conferring resistance to carbapenems, is a known phenomenon in the *A. veronii* population (10). While these genes were chromosomal with no close proximity to mobile genetic elements, their spread to other strains or bacterial species is not entirely unlikely as bacterial cell lysis could release DNA into the environment where it could be taken up by other strains or bacterial species through the process of transformation. Indeed, natural transformation has been described as a common mechanism of horizontal gene transfer among *Aeromonas* species, including *A. veronii*. *Aeromonas* species are capable of competence and transformation (62). In addition, *A. veronii* is known to easily acquire and exchange AMR genes (7, 20, 63). Although there was a low occurrence of AMR in *A. veronii* in this study, Lake Wilcox is a potential reservoir for AMR genes encoding resistance to multiple antibiotics as evidenced by results from previous studies on the lake where other bacterial species isolated from the lake carried multiple AMR genes (21, 22).

The mobilome is known to facilitate gene gain and loss, a phenomenon that plays a crucial role in bacterial evolution and ecological adaptation and a probable change in bacterial fitness (64, 65). This change can contribute to the emergence of divergent bacterial populations with unique

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

features, including higher pathogenic potential (64, 66, 67). In this study, no plasmid was detected in the *A. veronii* sequenced, but other mobile genetic elements (MGEs) including prophages and insertion sequences were identified. The majority of the intact prophages were identified as P2-like phages (Peduoviridae) (46), and a few of them were predicted to have multiple host bacterial species. This observation is interesting and could suggest a broad host range of these phages, which could have applications in biocontrol (68–70), but further studies on the host range of these phages would be needed to ascertain this. Another factor that contributes to the rapid evolution and ecological adaptation and that could influence the pathogenicity of bacterial species is BGCs that encode the production of various secondary metabolites (52, 71, 72). This phenomenon is seldom studied in *A. veronii*.

In this study, we found a high abundance of novel BGCs and identified unique NRPS and RiPP that were conserved in *A. veronii*. Notably, NRPS with high similarity (0.9) to enterobactin found in *Pseudomonas* sp. J465 (53), that mediates high affinity for iron acquisition in stringent conditions (73, 74). Angustmycin A/B/C (51) and pseudopyronine A/B (52) homologs were found to be conserved in *A. veronii*. These RiPP products encode antimicrobial properties and contribute to the survival of their producers in their ecological niche (51, 75). These conserved clusters could be promising genomic markers for typing *A. veronii*. Of note, a bacilysin homolog gene (56) was detected in a pair of non-human pathogenic strains. Bacilysin is an antimicrobial dipeptide produced by *Bacillus* species that exhibits antagonistic activity against both Gram-negative and Gram-positive bacteria (56, 76, 77). Further studies would be required to decipher the antimicrobial activity of the bacilysin homolog identified in this study against human pathogenic strains of *A. veronii* and other pathogens, as well as their mechanism of actions.

CONCLUSION

The study presents a genomic analysis of *A. veronii* strains isolated from a freshwater lake and defined the population structure and characterized the genetic factors associated with stress and ecological adaptation. A significant finding is the pathogenic potential of *A. veronii* to humans that underscores the public health implications, especially considering the recreational use of the lake. Among the MGEs identified that could contribute to the genetic diversity, adaptability, and the pathogenicity to human, as well as the of *A. veronii* as a reservoir for AMR genes, the BGCs identified presents opportunities for the discovery of novel bioactive compounds. Overall, this study not only contributes to our understanding of the genetic diversity and ecological dynamics of *A. veronii* but also highlights the potential public health risks and AMR reservoir role of this bacterium. It underscores the need for continuous surveillance for pathogens in aquatic ecosystems.

Funding. We acknowledge support from the Canada First Research Excellence fund in support of this project.

Authors' contribution. OL, NB, VP, RA, MS, YC, MP conducted the sampling, isolation, and whole-genome sequencing; OL performed the bioinformatics analysis and wrote the original draft of the

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

manuscript. LG conceived the project and provided funding and resources. OL, VP and LG supervised the study. All authors read and approved the final manuscript.

REFERENCES

1. Rodrigues C, Cunha MÂ. 2017. Assessment of the microbiological quality of recreational waters: indicators and methods. Euro-Mediterr J Environ Integr 2:25.

2. Stec J, Kosikowska U, Mendrycka M, Stępień-Pyśniak D, Niedźwiedzka-Rystwej P, Bębnowska D, Hrynkiewicz R, Ziętara-Wysocka J, Grywalska E. 2022. Opportunistic Pathogens of Recreational Waters with Emphasis on Antimicrobial Resistance—A Possible Subject of Human Health Concern. IJERPH 19:7308.

3. Ramírez-Castillo F, Loera-Muro A, Jacques M, Garneau P, Avelar-González F, Harel J, Guerrero-Barrera A. 2015. Waterborne Pathogens: Detection Methods and Challenges. Pathogens 4:307–334.

4. Steven JAC, Thorn RMS, Robinson GM, Turner D, Lee JE, Reynolds DM. 2022. The control of waterborne pathogenic bacteria in fresh water using a biologically active filter. npj Clean Water 5:30.

5. Liu F, Yuwono C, Tay ACY, Wehrhahn MC, Riordan SM, Zhang L. 2022. Analysis of global Aeromonas veronii genomes provides novel information on source of infection and virulence in human gastrointestinal diseases. BMC Genomics 23:166.

6. Ramsamy Y, Mlisana KP, Amoako DG, Abia ALK, Allam M, Ismail A, Singh R, Essack SY. 2020. Comparative Pathogenomics of Aeromonas veronii from Pigs in South Africa: Dominance of the Novel ST657 Clone. Microorganisms 8:2008.

7. Sakulworakan R, Chokmangmeepisarn P, Dinh-Hung N, Sivaramasamy E, Hirono I, Chuanchuen R, Kayansamruaj P, Rodkhum C. 2021. Insight Into Whole Genome of Aeromonas veronii Isolated From Freshwater Fish by Resistome Analysis Reveal Extensively Antibiotic Resistant Traits. Front Microbiol 12:733668.

8. Janda JM, Abbott SL. 2010. The Genus *Aeromonas* : Taxonomy, Pathogenicity, and Infection. Clin Microbiol Rev 23:35–73.

9. Lamy B, Kodjo A, Laurent F. 2009. Prospective Nationwide Study of *Aeromonas* Infections in France. J Clin Microbiol 47:1234–1237.

10. Drk S, Puljko A, Dželalija M, Udiković-Kolić N. 2023. Characterization of Third Generation Cephalosporin- and Carbapenem-Resistant *Aeromonas* Isolates from Municipal and Hospital Wastewater. Antibiotics 12:513.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

11. Senderovich Y, Ken-Dror S, Vainblat I, Blau D, Izhaki I, Halpern M. 2012. A Molecular Study on the Prevalence and Virulence Potential of *Aeromonas* spp. Recovered from Patients Suffering from Diarrhea in Israel. PLoS ONE 7:e30070.

12. Hoel S, Vadstein O, Jakobsen A. 2019. The Significance of Mesophilic *Aeromonas* spp. in Minimally Processed Ready-to-Eat Seafood. Microorganisms 7:91.

13. Martins LM, Marquez RF, Yano T. 2002. Incidence of toxic *Aeromonas* isolated from food and human infection. FEMS Immunology & Medical Microbiology 32:237–242.

14. Zh E, Rn Z, Rr S. 2021. Prevalence and Characterization of *Aeromonas* Spp. Isolated from Some Meat Products in Egypt. Nutr Food Technol Open Access 7.

15. Manna SK, Maurye P, Dutta C, Samanta G. 2013. Occurrence and Virulence Characteristics of *A. aeromonas* Species in Meat, Milk and Fish in I ndia. Journal of Food Safety 33:461–469.

16. Nawaz M, Khan SA, Khan AA, Sung K, Tran Q, Kerdahi K, Steele R. 2010. Detection and characterization of virulence genes and integrons in *Aeromonas veronii* isolated from catfish. Food Microbiology 27:327–331.

17. Nawaz M, Sung K, Khan SA, Khan AA, Steele R. 2006. Biochemical and Molecular Characterization of Tetracycline-Resistant *Aeromonas veronii* Isolates from Catfish. Appl Environ Microbiol 72:6461–6466.

18. Talagrand-Reboul E, Colston SM, Graf J, Lamy B, Jumas-Bilak E. 2020. Comparative and Evolutionary Genomics of Isolates Provide Insight into the Pathoadaptation of *Aeromonas*. Genome Biology and Evolution 12:535–552.

19. Hatrongjit R, Kerdsin A, Takeuchi D, Wongsurawat T, Jenjaroenpun P, Chopjitt P, Boueroy P, Akeda Y, Hamada S. 2020. Genomic Analysis of *Aeromonas veronii* C198, a Novel Mcr-3.41-Harboring Isolate from a Patient with Septicemia in Thailand. Pathogens 9:1031.

20. Zhu Z, Wu S, Zhu J, Wang T, Wen Y, Yang C, Lv J, Zhang H, Chen L, Du H. 2023. Emergence of *Aeromonas veronii* strain co-harboring blaKPC–2, mcr-3.17, and tmexC3.2-tmexD3.3-toprJ1b cluster from hospital sewage in China. Front Microbiol 14:1115740.

21. Bryan N, Anderson R, Lawal OU, Parreira VR, Goodridge L. 2023. Draft Genome Sequence of *Bacillus anthracis* N1, Isolated from a Recreational Freshwater Kettle Lake in Ontario, Canada. Microbiol Resour Announc e01262-22.

22. Lawal OU, Bryan N, Parreira VR, Soni M, Chen Y, Precious M, Goodridge L. 2024. Whole Genome Sequence of *Vibrio cholerae* NB-183 isolated from freshwater in Ontario, Canada harbors unique gene repertoire. BMC Genomic Data 25(8).

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

23. Bryan N, Rebecca Anderson, Lawal OU, Parreira VR, Goodridge L. 2023. Draft Genome Sequence of *Exiguobacterium* sp. Strain N5, Isolated from a Recreational Freshwater Kettle Lake in Ontario. Microbiol Resour Announc e01261-22.

24. Karama M, Lawal OU, Parreira VR, Soni M, Chen Y, Cenci-Goga BT, Grispoldi L, Greyling J, Goodridge L. 2024. Draft genome sequences of two *Salmonella* Uzaramo isolates from poultry in South Africa. Microbiol Resour Announc 13:e01026-23.

25. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

26. Souvorov A, Agarwala R, Lipman DJ. 2018. SKESA: strategic k-mer extension for scrupulous assemblies. Genome Biology 19:153.

27. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075.

28. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.

29. Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

30. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, Tyson GH, Zhao S, Hsu C-H, McDermott PF, Tadesse DA, Morales C, Simmons M, Tillman G, Wasilenko J, Folster JP, Klimke W. 2019. Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. Antimicrob Agents Chemother 63:e00483-19.

31. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48:D517–D525.

32. Chen L, Zheng D, Liu B, Yang J, Jin Q. 2016. VFDB 2016: Hierarchical and refined dataset for big data analysis - 10 years on. Nucleic Acids Research 44:D694–D697.

33. Liu B, Zheng D, Zhou S, Chen L, Yang J. 2022. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Research 50:D912–D917.

34. Robertson J, Nash JHE. 2018. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microbial Genomics 4:1–7.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

453 35. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, Stothard P, Gautam V. 2023. PHASTEST:
454 faster than PHASTER, better than PHAST. Nucleic Acids Research 51:W443–W450.

455 36. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T.
456 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids
457 Research 49:W29–W35.

458 37. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane
459 JA, Parkhill J. 2015. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics
460 31:3691–3693.

461 38. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - Approximately maximum-likelihood trees for
462 large alignments. PLoS ONE 5:1–10.

463 39. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb
464 software, the PubMLST.org website and their applications. Wellcome Open Res 3:124.

465 40. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using
466 exact alignments. Genome Biol 15:R46.

467 41. Yoon S-H, Ha S, Lim J, Kwon S, Chun J. 2017. A large-scale evaluation of algorithms to calculate
468 average nucleotide identity. Antonie van Leeuwenhoek 110:1281–1286.

469 42. Cosentino S, Voldby Larsen M, Møller Aarestrup F, Lund O. 2013. PathogenFinder -
470 Distinguishing Friend from Foe Using Bacterial Whole Genome Sequence Data. PLoS ONE
471 8:e77302.

472 43. Hockenberry AJ, Wilke CO. 2021. BACPHLIP: predicting bacteriophage lifestyle from conserved
473 protein domains. PeerJ 9:e11396.

474 44. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV
475 assesses the quality and completeness of metagenome-assembled viral genomes. Nat
476 Biotechnol 39:578–585.

477 45. Shang J, Jiang J, Sun Y. 2021. Bacteriophage classification for assembled contigs using graph
478 convolutional network. Bioinformatics 37:i25–i33.

479 46. Turner D, Shkoporov AN, Lood C, Millard AD, Dutilh BE, Alfenas-Zerbini P, Van Zyl LJ, Aziz RK,
480 Oksanen HM, Poranen MM, Kropinski AM, Barylski J, Brister JR, Chanisvili N, Edwards RA,
481 Enault F, Gillis A, Knezevic P, Krupovic M, Kurtböke I, Kushkina A, Lavigne R, Lehman S, Lobocka
482 M, Moraru C, Moreno Switt A, Morozova V, Nakavuma J, Reyes Muñoz A, Rūmnieks J, Sarkar B,
483 Sullivan MB, Uchiyama J, Wittmann J, Yigang T, Adriaenssens EM. 2023. Abolishment of
484 morphology-based taxa and change to binomial species names: 2022 taxonomy update of the
485 ICTV bacterial viruses subcommittee. Arch Virol 168:74.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

47. Shang J, Peng C, Liao H, Tang X, Sun Y. 2023. PhaBOX: A web server for identifying and characterizing phage contigs in metagenomic data. arXiv:2303.15707. arXiv.

48. Shang J, Sun Y. 2022. CHERRY: a Computational metHod for accuratE pRediction of virus–pRokarYotic interactions using a graph encoder–decoder model. Briefings in Bioinformatics 23:bbac182.

49. Lawal OU, Goodridge L. 2024. TSPDB: A curated resource of tailspike proteins with potential applications in phage research. Preprint.

50. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, Roeters A, Lokhorst W, Fernandez-Guerra A, Cappelini LTD, Goering AW, Thomson RJ, Metcalf WW, Kelleher NL, Barona-Gomez F, Medema MH. 2020. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol 16:60–68.

51. Yu L, Zhou W, She Y, Ma H, Cai Y-S, Jiang M, Deng Z, Price NPJ, Chen W. 2021. Efficient biosynthesis of nucleoside cytokinin angustmycin A containing an unusual sugar system. Nat Commun 12:6633.

52. Bauer JS, Ghequire MGK, Nett M, Josten M, Sahl H, De Mot R, Gross H. 2015. Biosynthetic Origin of the Antibiotic Pseudopyronines A and B in *Pseudomonas putida* BW11M1. ChemBioChem 16:2491–2497.

53. Curson ARJ, Sullivan MJ, Todd JD, Johnston AWB. 2010. Identification of genes for dimethyl sulfide production in bacteria in the gut of Atlantic Herring (*Clupea harengus*). The ISME Journal 4:144–146.

54. Litomska A, Ishida K, Dunbar KL, Boettger M, Coyne S, Hertweck C. 2018. Enzymatic Thioamide Formation in a Bacterial Antimetabolite Pathway. Angew Chem Int Ed 57:11574–11578.

55. Grammbitter GLC, Schmalhofer M, Karimi K, Shi Y-M, Schöner TA, Tobias NJ, Morgner N, Groll M, Bode HB. 2019. An Uncommon Type II PKS Catalyzes Biosynthesis of Aryl Polyene Pigments. J Am Chem Soc 141:16615–16623.

56. Moran S, Robertson K, Paradisi F, Rai DK, Murphy CD. 2010. Production of lipopeptides in Bacillus sp. CS93 isolated from Pozol: Lipopeptides from *Bacillus*. FEMS Microbiology Letters 304:69–73.

57. Tohya M, Teramoto K, Watanabe S, Hishinuma T, Shimojima M, Ogawa M, Tada T, Tabe Y, Kirikae T. 2022. Whole-Genome Sequencing-Based Re-Identification of *Pseudomonas putida* / *fluorescens* Clinical Isolates Identified by Biochemical Bacterial Identification Systems. Microbiol Spectr 10:e02491-21.

58. Garza-González E, Bocanegra-Ibarias P, Dinh A, Morfín-Otero R, Camacho-Ortiz A, Rojas-Larios F, Rodríguez-Zulueta P, Arias CA. 2020. Species identification of *Enterococcus* spp:

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Whole genome sequencing compared to three biochemical test-based systems and two Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) systems. Clinical Laboratory Analysis 34:e23348.

59. Chacon MR, Soler L, Groisman EA, Guarro J, Figueras MJ. Type III Secretion System Genes in Clinical Aeromonas Isolates.

60. Stuber K, Frey J, Burnens AP, Kuhnert P. 2003. Detection of type III secretion genes as a general indicator of bacterial virulence. Molecular and Cellular Probes 17:25–32.

61. Coburn B, Sekirov I, Finlay BB. 2007. Type III Secretion Systems and Disease. Clin Microbiol Rev 20:535–549.

62. Huddleston JR, Brokaw JM, Zak JC, Jeter RM. 2013. Natural transformation as a mechanism of horizontal gene transfer among environmental *Aeromonas* species. Systematic and Applied Microbiology 36:224–234.

63. Algammal AM, Ibrahim RA, Alfifi KJ, Ghabban H, Alghamdi S, Kabrah A, Khafagy AR, Abou-Elela GM, Abu-Elala NM, Donadu MG, El-Tarabili RM. 2022. A First Report of Molecular Typing, Virulence Traits, and Phenotypic and Genotypic Resistance Patterns of Newly Emerging XDR and MDR *Aeromonas veronii* in Mugil seheli. Pathogens 11:1262.

64. Hazen TH, Pan L, Gu J-D, Sobecky PA. 2010. The contribution of mobile genetic elements to the evolution and ecology of Vibrios: Vibrio HGT. FEMS Microbiology Ecology 74:485–499.

65. Vale FF, Lehours P, Yamaoka Y. 2022. Editorial: The Role of Mobile Genetic Elements in Bacterial Evolution and Their Adaptability. Front Microbiol 13:849667.

66. Huang S, Li H, Ma L, Liu R, Li Y, Wang H, Lu X, Huang X, Wu X, Liu X. 2023. Insertion sequence contributes to the evolution and environmental adaptation of *Acidithiobacillus*. BMC Genomics 24:282.

67. Kozitskaya S, Cho SH, Dietrich K, Marre R, Naber K, Ziebuhr W. 2004. The Bacterial Insertion Sequence Element IS*256* Occurs Preferentially in Nosocomial *Staphylococcus epidermidis* Isolates: Association with Biofilm Formation and Resistance to Aminoglycosides. Infection and Immunity 72:1210–1215.

68. Ross A, Ward S, Hyman P. 2016. More Is Better: Selecting for Broad Host Range Bacteriophages. Front Microbiol 7.

69. Myers J, Davis Ii J, Lollo M, Hudec G, Hyman P. 2023. More's the Same—Multiple Hosts Do Not Select for Broader Host Range Phages. Viruses 15:518.

70. Fong K, Wong CWY, Wang S, Delaquis P. 2021. How Broad Is Enough: The Host Range of Bacteriophages and Its Impact on the Agri-Food Sector. PHAGE 2:83–91.

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

71. Gavriilidou A, Kautsar SA, Zaburannyi N, Krug D, Müller R, Medema MH, Ziemert N. 2022. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. Nat Microbiol 7:726–735.

72. Xia L, Miao Y, Cao A, Liu Y, Liu Z, Sun X, Xue Y, Xu Z, Xun W, Shen Q, Zhang N, Zhang R. 2022. Biosynthetic gene cluster profiling predicts the positive association between antagonism and phylogeny in Bacillus. Nat Commun 13:1023.

73. Moynié L, Milenkovic S, Mislin GLA, Gasser V, Malloci G, Baco E, McCaughan RP, Page MGP, Schalk IJ, Ceccarelli M, Naismith JH. 2019. The complex of ferric-enterobactin with its transporter from *Pseudomonas aeruginosa* suggests a two-site model. Nat Commun 10:3673.

74. Saha P, Xiao X, Yeoh BS, Chen Q, Katkere B, Kirimanjeswara GS, Vijay-Kumar M. 2019. The bacterial siderophore enterobactin confers survival advantage to *Salmonella* in macrophages. Gut Microbes 10:412–423.

75. Bouthillette LM, Darcey CA, Handy TE, Seaton SC, Wolfe AL. 2017. Isolation of the antibiotic pseudopyronine B and SAR evaluation of C3/C6 alkyl analogs. Bioorganic & Medicinal Chemistry Letters 27:2762–2765.

76. Islam T, Rabbee MF, Choi J, Baek K-H. 2022. Biosynthesis, Molecular Regulation, and Application of Bacilysin Produced by *Bacillus* Species. Metabolites 12:397.

77. Nannan C, Vu HQ, Gillis A, Caulier S, Nguyen TTT, Mahillon J. 2021. Bacilysin within the *Bacillus subtilis* group: gene prevalence versus antagonistic activity against Gram-negative foodborne pathogens. Journal of Biotechnology 327:28–35.

Legends

Table 1 - Summary of sequence metrics of *Aeromonas* isolates recovered from a freshwater lake.

Table 2 – Features of intact phages detected in *Aeromonas veronii* sequenced in the study.

Figure 1 - Location of Lake Wilcox in Ontario, Canada. Image from http://maps.google.com

Figure 2 – Maximum likelihood tree of *Aeromonas veronii* recovered from freshwater lake in Ontario. Core genome sequence alignment was generated from draft genomes using MAFFT v7.453, and the tree was constructed using FastTree v2.1.11 using generalized time reversible nucleotide substitution model with 1000 bootstraps for node support. The dotted line depicts *Aeromonas* veronii. NB-7 (*A. allosaccharophila*) and NB-180 (*A. caviae*) were outgroups used to re-root the tree. The draft genomes were screened for genes encoding antimicrobial resistance and virulence using CARD and VFDB databases, respectively. The scale bar at the bottom represents nucleotide substitution per site. The tree was visualized using iTOL; (https://itol.embl.de).

bioRxiv preprint doi: https://doi.org/10.1101/2024.05.08.593017; this version posted May 10, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 3 - Core-genome based maximum-likelihood tree of global *Aeromonas veronii* from different sources. Each node represents a strain. Core-genome sequence alignment was generated from draft genomes using MAFFT v7.453. Maximum likelihood tree was constructed using FastTree v2.1.11. The generalized time reversible nucleotide substitution was performed with 1000 bootstraps random resampling for support. *A. allosaccharophila* (NB-7) and *A. caviae* (NB-180) were outgroups used for re-rooting the tree. The scale bar at the bottom represents nucleotide substitution per site. The figure was generated using iTOL (https://itol.embl.de). *Aeromonas* isolates recovered in this study were labelled in red.

Figure 4 – Distribution of insertion sequence elements in *Aeromonas veronii* recovered from freshwater lake in Ontario. Core genome sequence alignment was generated from draft genomes using MAFFT v7.453, and the tree was constructed using FastTree v2.1.11 using generalized time reversible nucleotide substitution model with 1000 random resampling for node support. The dotted line depicts *A.* veronii. NB-7 (*A. allosaccharophila*) and NB-180 (*A. caviae*) were outgroups used to re-root the tree. Mobile elements were detected using Mob-suites. The scale bar at the bottom represents nucleotide substitution per site. The tree was visualized using iTOL; (https://itol.embl.de).

Figure 5 – Novel biosynthetic gene clusters identified in *Aeromonas veronii* sequenced. **A**. ribosomally synthesized and post-translationally modified peptides (RiPP) biosynthetic gene cluster (BGC) with low similarity score of 0.17 to angustmycin A/B/C in *Streptomyces angustmyceticus* (Accession MZ151497.1). **B**. RiPP gene cluster with low similarity score of 0.04 pseudopyronine A/B found in *Pseudomonas putida* (Accession KT373879.1). **C**. RiPP gene cluster detected in *A. veronii* and *A. allosaccharophila* with low similarity score of 0.08 to pseudopyronine A/B found in *Pseudomonas putida* (Accession KT373879.1). **(D).** Non-ribosomally polyketide synthase (NRPS) BGC with 0.9 similarity to enterobactin in *Pseudomonas* sp. J465 (Accession GQ370384.1). **E**. Homoserine lactone detected in *A. veronii* and in *A. allosaccharophila* with low similarity score of 0.14 to thioguanine found in *Erwinia amylovora* CFBP1430 (Accession number: NC_013971.1).

Figure 6 – Biosynthetic gene clusters identified in *Aeromonas veronii* predicted to be human or non-human pathogens. (A). Aryl polyene gene cluster detected in *A. veronii* isolates predicted to be non-human pathogen with similarity score of 0.26 to bacilysin found in *Bacillus* sp. CS93 (Accession number: GQ889493.1). (B). Aryl polyene gene cluster detected in *A. veronii* isolates predicted to be human pathogen with similarity score of 0.44 to aryl polyene found in *Xenorhabdus doucetiae* (Accession number: NZ_FO704550.1).

Supplementary Data

Table S1 - Summary of the novel alleles and sequence types identified in *Aeromonas* isolates sequenced in this study.

Table S2 - Distance matrix between *Aeromonas veronii* isolates sequenced in this study.

Table S3 - List of accession number and associated metadata of publicly available *Aeromonas veronii* genomes retrieved from the GenBank NCBI database.

➔ **Tables 1 and 2, and Supplementary Tables S1, S2 & S3 are accessible here - Tables.xlsx**

Figure 1

Figure 2

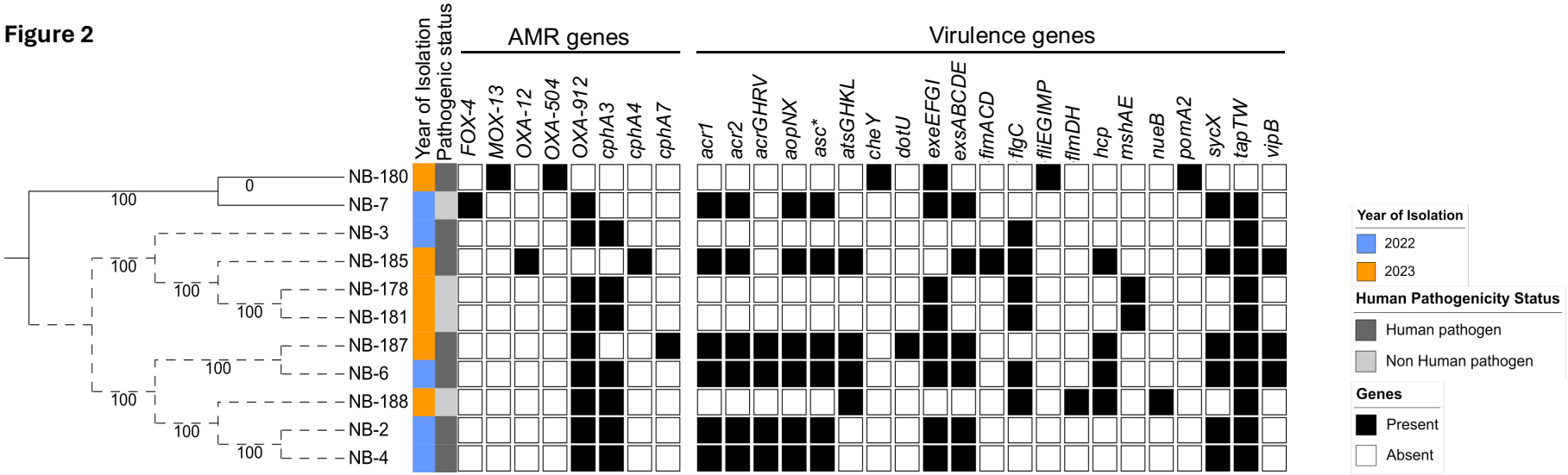


Figure 3

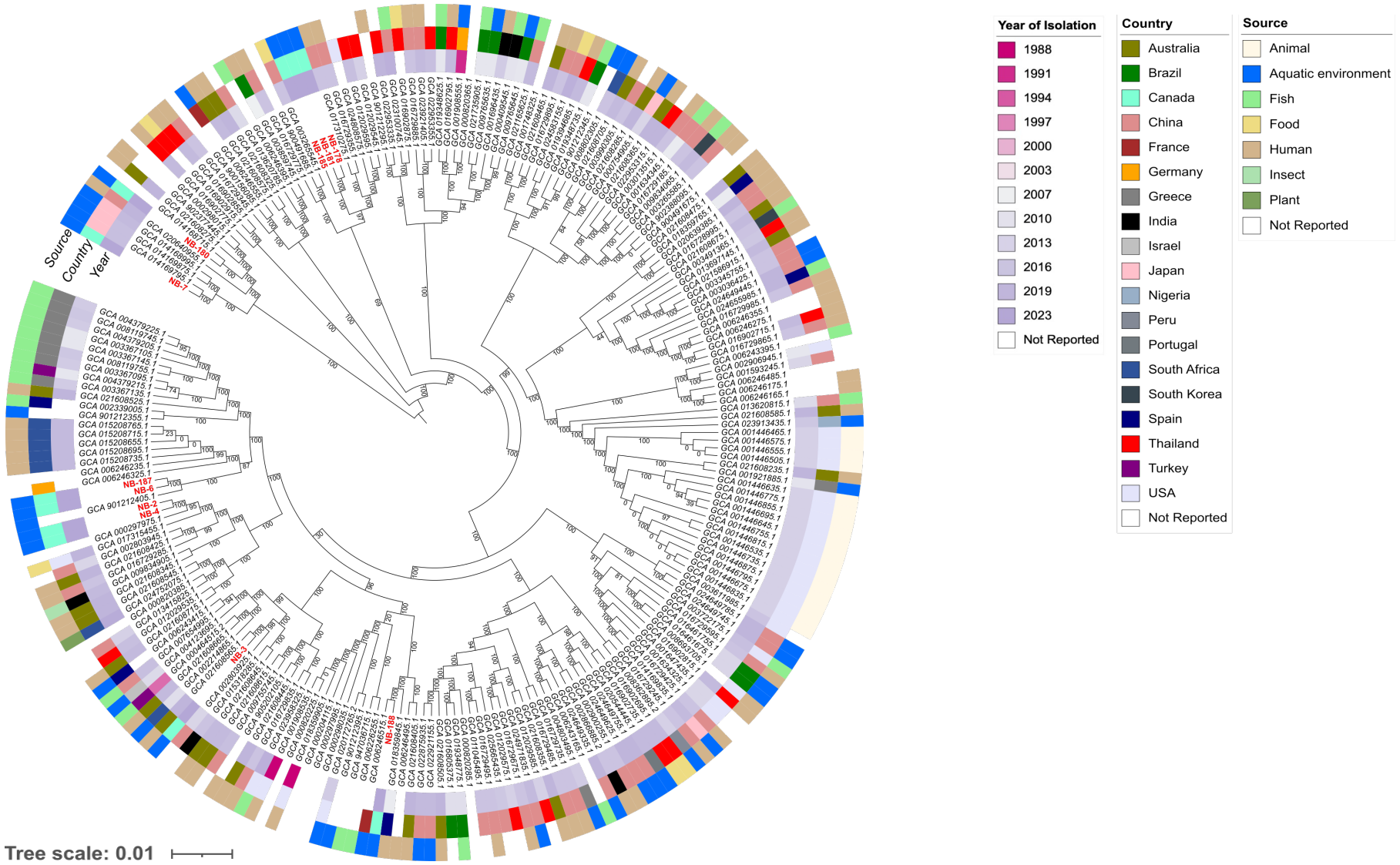

Source
Country
Year
Year of Isolation
1988
1991
1994
1997
2000
2003
2007
2010
2013
2016
2019
2023
Not Reported
Country
Australia
Brazil
Canada
China
France
Germany
Greece
India
Israel
Japan
Nigeria
Peru
Portugal
South Africa
South Korea
Spain
Thailand
Turkey
USA
Not Reported
Source
Animal
Aquatic environment
Fish
Food
Human
Insect
Plant
Not Reported
Tree scale: 0.01

Figure 4

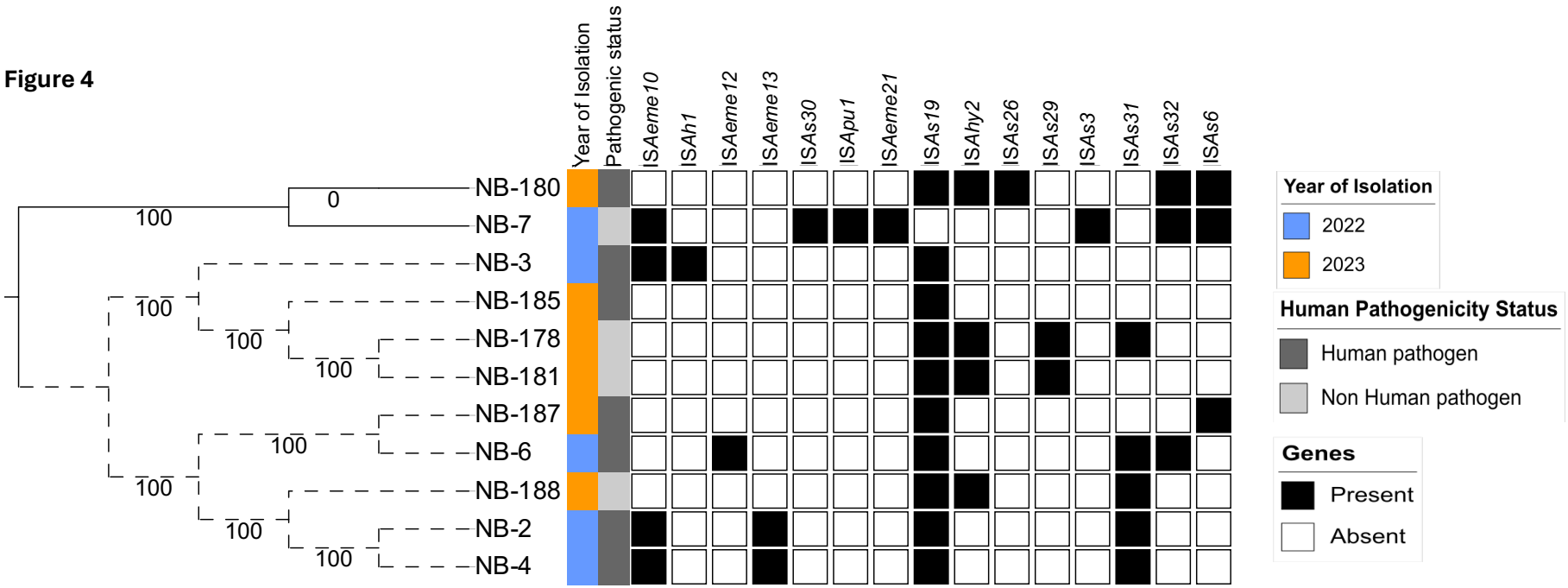
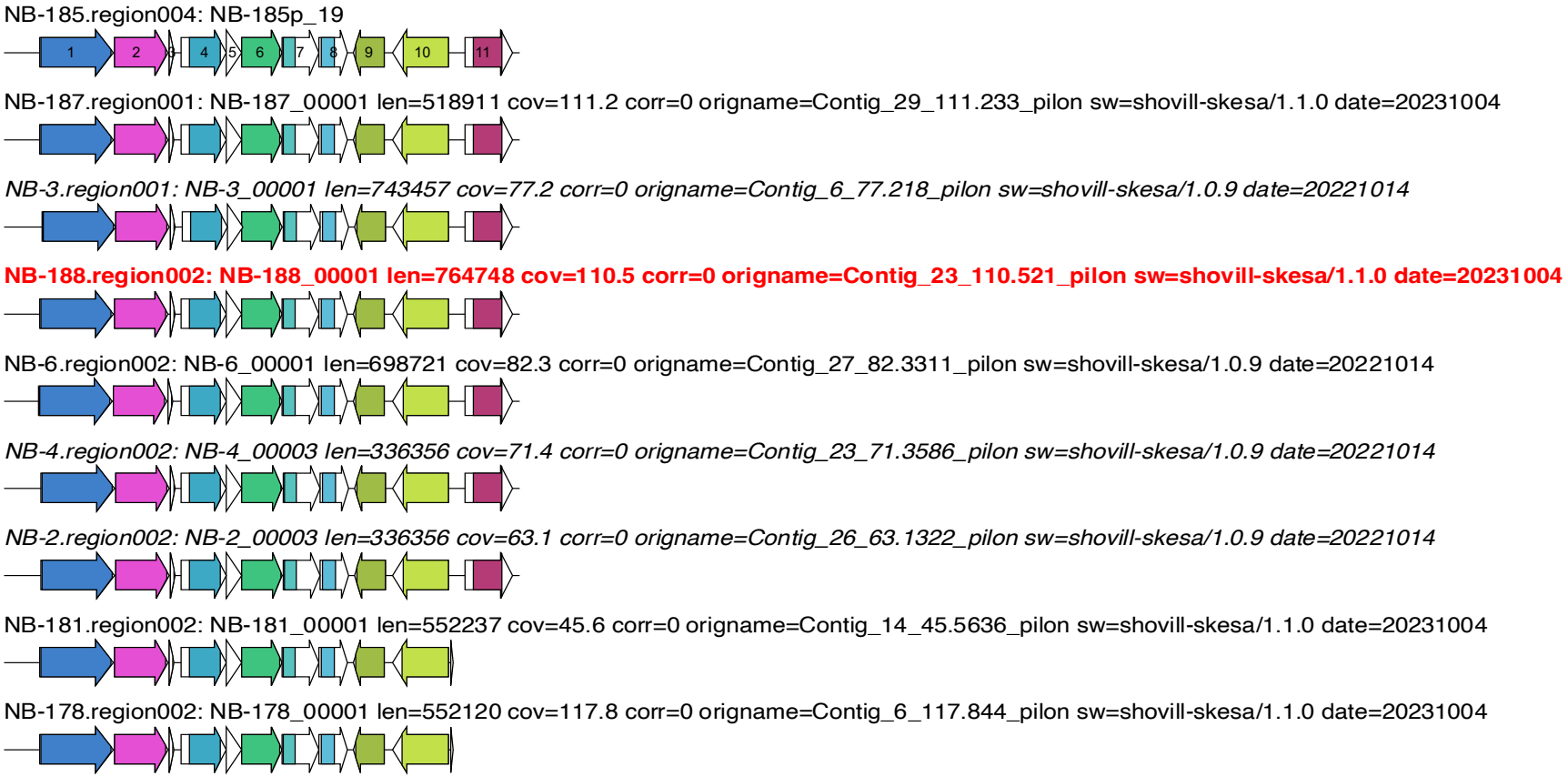

Year of Isolation
Pathogenic status
ISAeme10
ISAh1
ISAeme12
ISAeme13
ISAs30
ISApu1
ISAeme21
ISAs19
ISAhy2
ISAs26
ISAs29
ISAs3
ISAs31
ISAs32
ISAs6
NB-180
NB-7
NB-3
NB-185
NB-178
NB-181
NB-187
NB-6
NB-188
NB-2
NB-4
0
100
Year of Isolation
2022
2023
Human Pathogenicity Status
Human pathogen
Non Human pathogen
Genes
Present
Absent

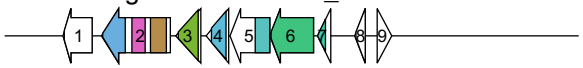
Figure 5

A

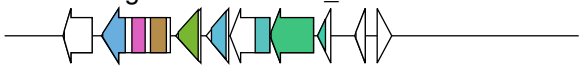

NB-185.region004: NB-185p_19
1
2
3
4
5
6
7
8
9
10
11
NB-187.region001: NB-187_00001 len=518911 cov=111.2 corr=0 origname=Contig_29_111.233_pilon sw=shovill-skesa/1.1.0 date=20231004
NB-3.region001: NB-3_00001 len=743457 cov=77.2 corr=0 origname=Contig_6_77.218_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-188.region002: NB-188_00001 len=764748 cov=110.5 corr=0 origname=Contig_23_110.521_pilon sw=shovill-skesa/1.1.0 date=20231004
NB-6.region002: NB-6_00001 len=698721 cov=82.3 corr=0 origname=Contig_27_82.3311_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-4.region002: NB-4_00003 len=336356 cov=71.4 corr=0 origname=Contig_23_71.3586_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-2.region002: NB-2_00003 len=336356 cov=63.1 corr=0 origname=Contig_26_63.1322_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-181.region002: NB-181_00001 len=552237 cov=45.6 corr=0 origname=Contig_14_45.5636_pilon sw=shovill-skesa/1.1.0 date=20231004
NB-178.region002: NB-178_00001 len=552120 cov=117.8 corr=0 origname=Contig_6_117.844_pilon sw=shovill-skesa/1.1.0 date=20231004

B

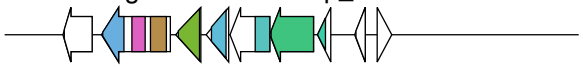
NB-178.region004: NB-178_00006 len=239147 cov=123.0 corr=0 origname=Contig_30_123.022_pilon sw=shovill-skesa/1.1.0 date=20231004



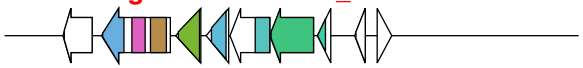
NB-181.region003: NB-181_00003 len=270486 cov=46.0 corr=0 origname=Contig_28_46.0484_pilon sw=shovill-skesa/1.1.0 date=20231004



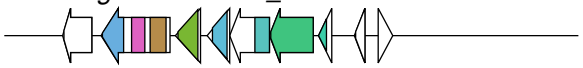
NB-185.region001: NB-185p_1



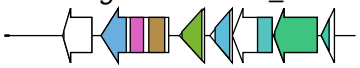
NB-188.region003: NB-188_00002 len=602127 cov=111.2 corr=0 origname=Contig_6_111.181_pilon sw=shovill-skesa/1.1.0 date=20231004



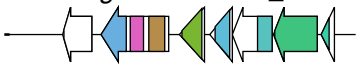
NB-6.region001: NB-6_00001 len=698721 cov=82.3 corr=0 origname=Contig_27_82.3311_pilon sw=shovill-skesa/1.0.9 date=20221014



NB-4.region003: NB-4_00014 len=105762 cov=72.2 corr=0 origname=Contig_6_72.1807_pilon sw=shovill-skesa/1.0.9 date=20221014

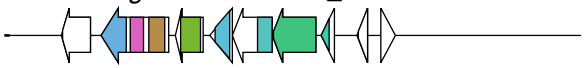


NB-2.region003: NB-2_00014 len=105762 cov=64.0 corr=0 origname=Contig_3_64.0166_pilon sw=shovill-skesa/1.0.9 date=20221014

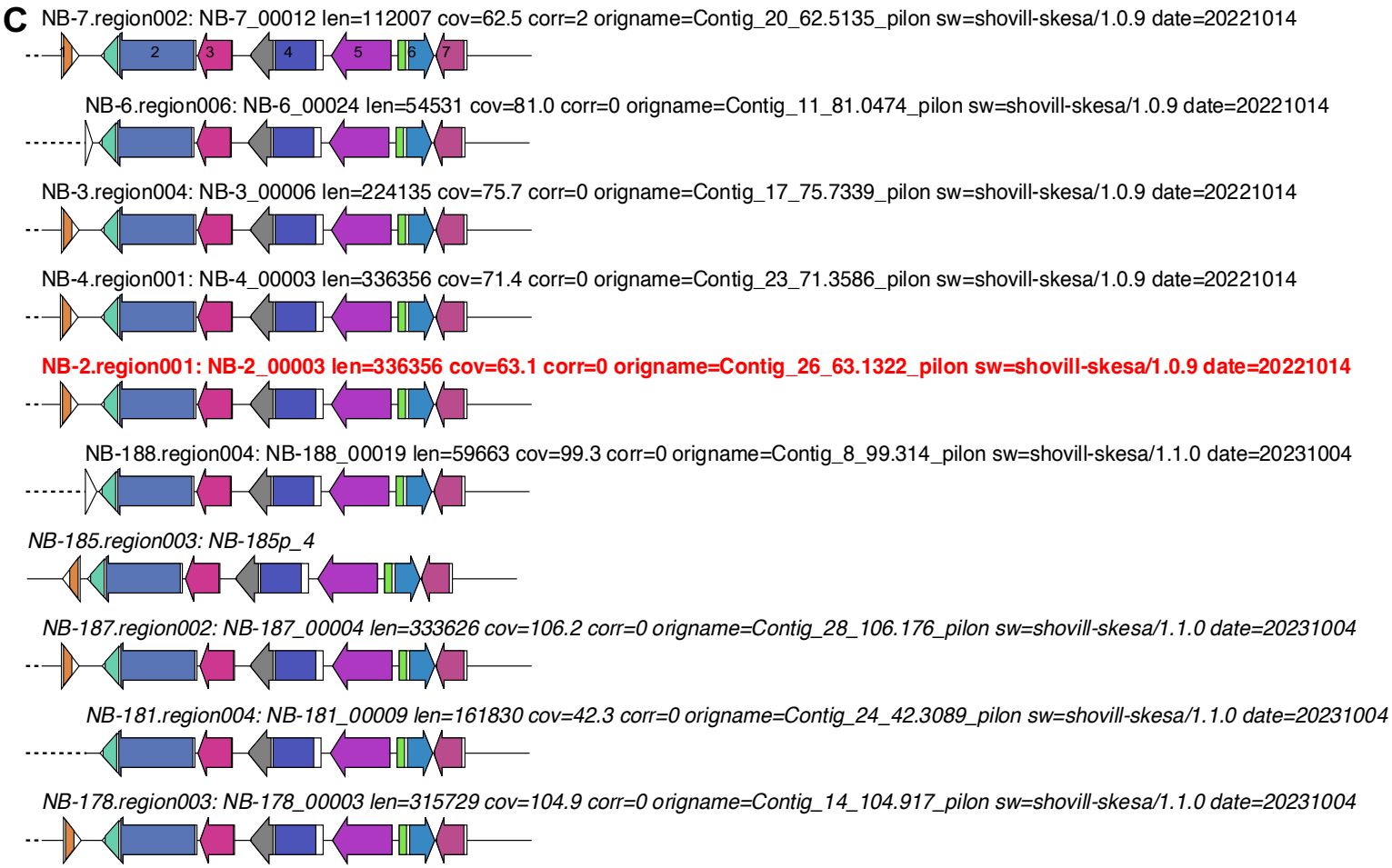


NB-3.region003: NB-3_00002 len=655010 cov=78.6 corr=0 origname=Contig_9_78.5934_pilon sw=shovill-skesa/1.0.9 date=20221014

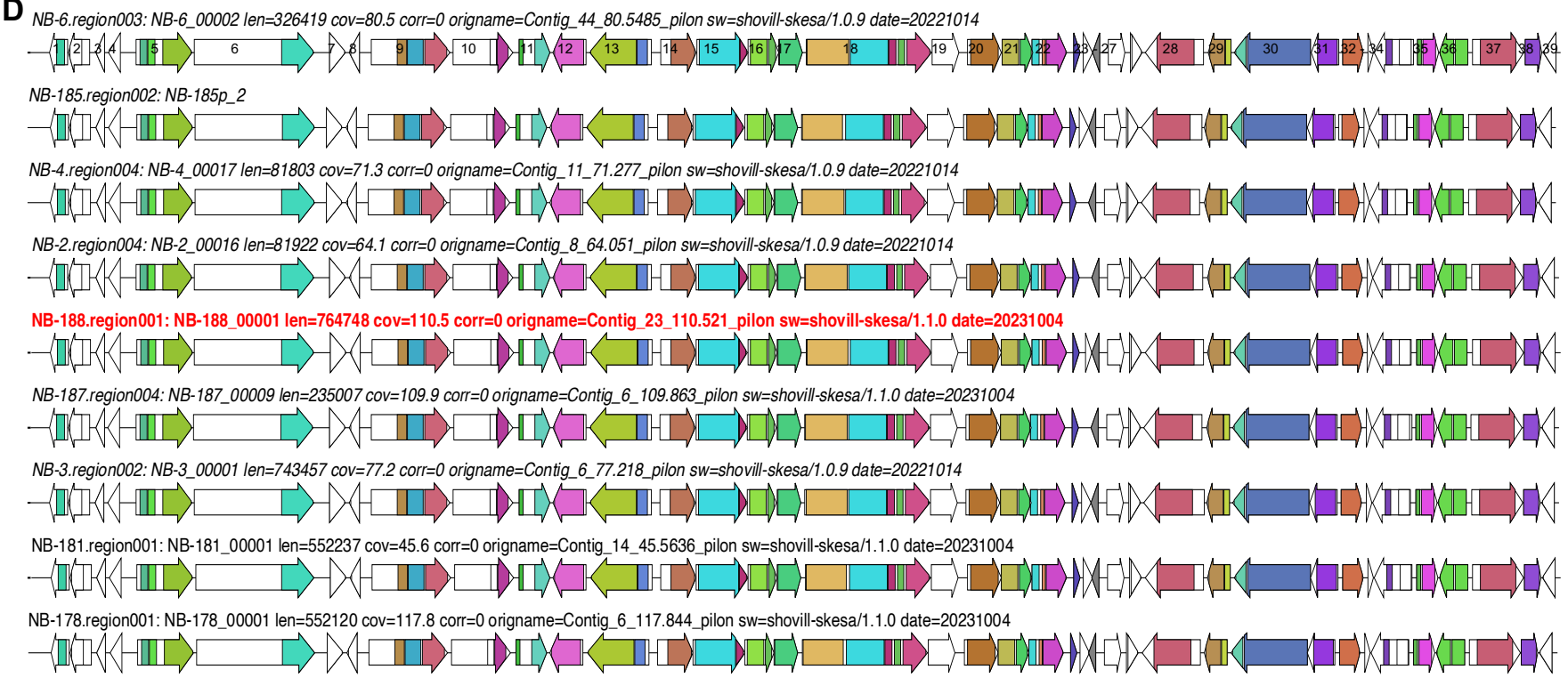
NB-187.region003: NB-187_00006 len=300356 cov=124.6 corr=0 origname=Contig_18_124.617_pilon sw=shovill-skesa/1.1.0 date=20231004



C


NB-7.region002: NB-7_00012 len=112007 cov=62.5 corr=2 origname=Contig_20_62.5135_pilon sw=shovill-skesa/1.0.9 date=20221014
2
3
4
5
6
7
NB-6.region006: NB-6_00024 len=54531 cov=81.0 corr=0 origname=Contig_11_81.0474_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-3.region004: NB-3_00006 len=224135 cov=75.7 corr=0 origname=Contig_17_75.7339_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-4.region001: NB-4_00003 len=336356 cov=71.4 corr=0 origname=Contig_23_71.3586_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-2.region001: NB-2_00003 len=336356 cov=63.1 corr=0 origname=Contig_26_63.1322_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-188.region004: NB-188_00019 len=59663 cov=99.3 corr=0 origname=Contig_8_99.314_pilon sw=shovill-skesa/1.1.0 date=20231004
NB-185.region003: NB-185p_4
NB-187.region002: NB-187_00004 len=333626 cov=106.2 corr=0 origname=Contig_28_106.176_pilon sw=shovill-skesa/1.1.0 date=20231004
NB-181.region004: NB-181_00009 len=161830 cov=42.3 corr=0 origname=Contig_24_42.3089_pilon sw=shovill-skesa/1.1.0 date=20231004
NB-178.region003: NB-178_00003 len=315729 cov=104.9 corr=0 origname=Contig_14_104.917_pilon sw=shovill-skesa/1.1.0 date=20231004

D


NB-6.region003: NB-6_00002 len=326419 cov=80.5 corr=0 origname=Contig_44_80.5485_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-185.region002: NB-185p_2
NB-4.region004: NB-4_00017 len=81803 cov=71.3 corr=0 origname=Contig_11_71.277_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-2.region004: NB-2_00016 len=81922 cov=64.1 corr=0 origname=Contig_8_64.051_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-188.region001: NB-188_00001 len=764748 cov=110.5 corr=0 origname=Contig_23_110.521_pilon sw=shovill-skesa/1.1.0 date=20231004
NB-187.region004: NB-187_00009 len=235007 cov=109.9 corr=0 origname=Contig_6_109.863_pilon sw=shovill-skesa/1.1.0 date=20231004
NB-3.region002: NB-3_00001 len=743457 cov=77.2 corr=0 origname=Contig_6_77.218_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-181.region001: NB-181_00001 len=552237 cov=45.6 corr=0 origname=Contig_14_45.5636_pilon sw=shovill-skesa/1.1.0 date=20231004
NB-178.region001: NB-178_00001 len=552120 cov=117.8 corr=0 origname=Contig_6_117.844_pilon sw=shovill-skesa/1.1.0 date=20231004

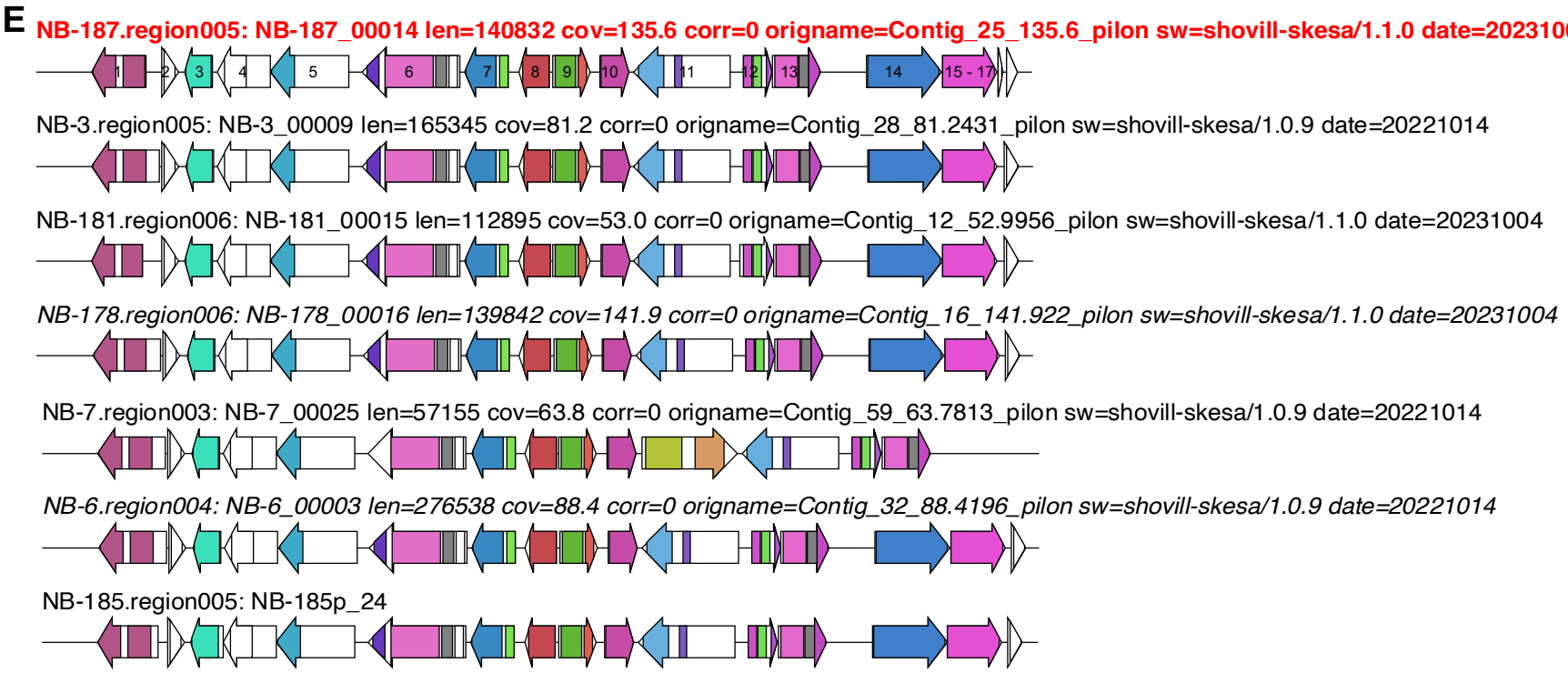
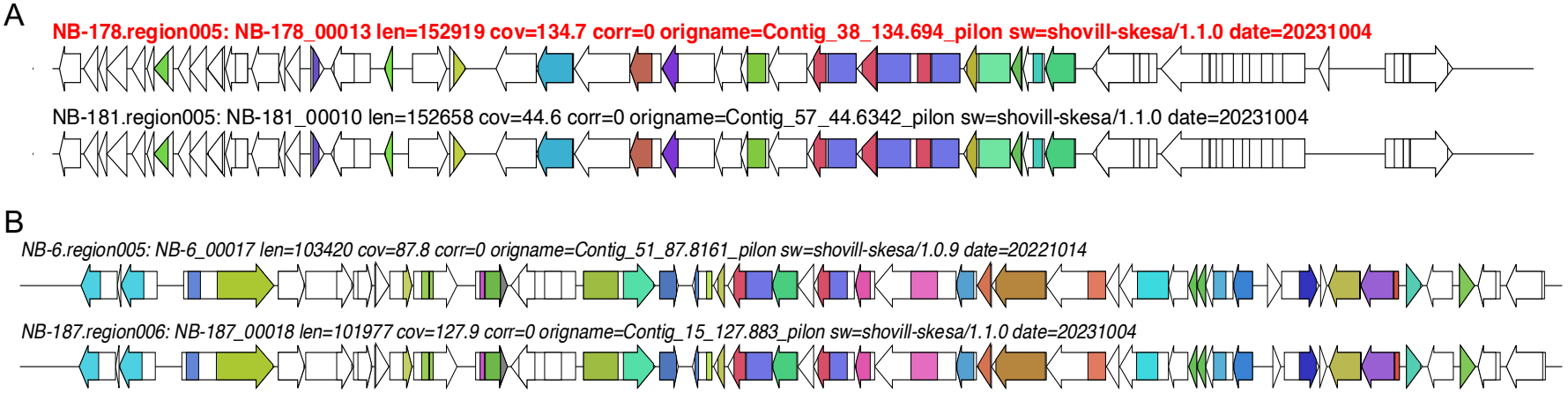

E
NB-187.region005: NB-187_00014 len=140832 cov=135.6 corr=0 origname=Contig_25_135.6_pilon sw=shovill-skesa/1.1.0 date=20231004
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15 - 17
NB-3.region005: NB-3_00009 len=165345 cov=81.2 corr=0 origname=Contig_28_81.2431_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-181.region006: NB-181_00015 len=112895 cov=53.0 corr=0 origname=Contig_12_52.9956_pilon sw=shovill-skesa/1.1.0 date=20231004
NB-178.region006: NB-178_00016 len=139842 cov=141.9 corr=0 origname=Contig_16_141.922_pilon sw=shovill-skesa/1.1.0 date=20231004
NB-7.region003: NB-7_00025 len=57155 cov=63.8 corr=0 origname=Contig_59_63.7813_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-6.region004: NB-6_00003 len=276538 cov=88.4 corr=0 origname=Contig_32_88.4196_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-185.region005: NB-185p_24

Figure 6


A
NB-178.region005: NB-178_00013 len=152919 cov=134.7 corr=0 origname=Contig_38_134.694_pilon sw=shovill-skesa/1.1.0 date=20231004
NB-181.region005: NB-181_00010 len=152658 cov=44.6 corr=0 origname=Contig_57_44.6342_pilon sw=shovill-skesa/1.1.0 date=20231004
B
NB-6.region005: NB-6_00017 len=103420 cov=87.8 corr=0 origname=Contig_51_87.8161_pilon sw=shovill-skesa/1.0.9 date=20221014
NB-187.region006: NB-187_00018 len=101977 cov=127.9 corr=0 origname=Contig_15_127.883_pilon sw=shovill-skesa/1.1.0 date=20231004