Completion of genome of *Aeromonas salmonicida* subsp. *salmonicida* 01-B526 reveals how sequencing technologies can influence sequence quality and result interpretations

A. T. Vincent[1] **and S. J. Charette**[2,3,4]

1) *INRS-Institut Armand-Frappier, Bacterial Symbionts Evolution, Laval City, QC, H7V 1B7, Canada,* 2) *Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Quebec City, QC, G1V 0A6, Canada,* 3) *Département de biochimie, de microbiologie et de bio-informatique, Faculté des sciences et de génie, Université Laval, Quebec City, QC, G1V 0A6, Canada and* 4) *Centre de recherche de l'Institut universitaire de cardiologie et de pneumologie de Québec (IUCPQ), Quebec City, QC, G1V 4G5, Canada*

Abstract

Aeromonas salmonicida subsp. *salmonicida* is a pathogen that primarily infects salmonids. A strain of this bacterium, 01-B526, has been used in several studies as a reference. The genomic sequence of this strain is available, but comes from pyrosequencing and is the second most fragmented assembly for this bacterium. We generated its closed genome sequence and found a pitfall in result interpretations associated with low-quality genomic sequences.
© 2018 The Author(s). Published by Elsevier Ltd.

Keywords: Aeromonas salmonicida, average nucleotide identity, Genomics, PacBio, sequencing technologies
Original Submission: 24 February 2018; **Accepted:** 24 May 2018
Article published online: 31 May 2018

Corresponding author: A. T. Vincent, INRS-Institut Armand-Frappier, 531 boul. des Prairies, Laval, QC, H7V 1B7, Canada.
E-mail: Antony.Vincent@iaf.inrs.ca

The bacterium *Aeromonas salmonicida* subsp. *salmonicida* is a fish pathogen causing furunculosis [1]. Given its veterinary importance, a sequencing effort was made during the last decade to obtain the genomic sequences of strains of this bacterium. Mainly two technologies were used, pyrosequencing (Roche 454) and sequencing by synthesis (Illumina MiSeq). Pyrosequencing is conducive in generating several errors in the allocation of bases, especially in homopolymeric regions [2]. Even if this technology is no longer available, sequences originating from it are still abundant in databases.

Here, the differences between *A. salmonicida* subsp. *salmonicida* genomic sequences resulting from 454 pyrosequencing or Illumina MiSeq were assessed. With the Illumina MiSeq, it is possible to obtain fewer contigs than with 454 pyrosequencing (Fig. 1A). As expected, the N50 values are generally higher for assemblies from Illumina MiSeq than 454 pyrosequencing (Fig. 1B). Only two assemblies, from strains isolated in China [3], do not respect the trend. This may reflect a difference in the assembly tools and parameters used, or even the presence of repeated regions [4,5].

After the publication of the first complete genome of *A. salmonicida* subsp. *salmonicida,* that of the French strain A449 [6], the genome of a virulent strain from Canada was sequenced in order to find biogeographical markers [7]. This made it possible to highlight a genomic island, *AsaGEI,* whose variants correlate with the geographical locations where the strains were isolated [3,8,9]. Since then, strain 01-B526 has been repeatedly used as a reference, both for bioinformatics studies and for wet-lab experiments. Unfortunately, the sequence of this strain is from pyrosequencing data and is the second assembly with the highest number of contigs for this bacterium (Supplementary Table S1).

Given the importance of strain 01-B526 as a reference, we resequenced its genome with Illumina MiSeq as well as with PacBio single-molecule real-time (SMRT) technology, which is known for producing bacterial genome sequences that are often closed [10]. As such, it was possible to obtain the closed

© 2018 The Author(s). Published by Elsevier Ltd
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
https://doi.org/10.1016/j.nmni.2018.05.007

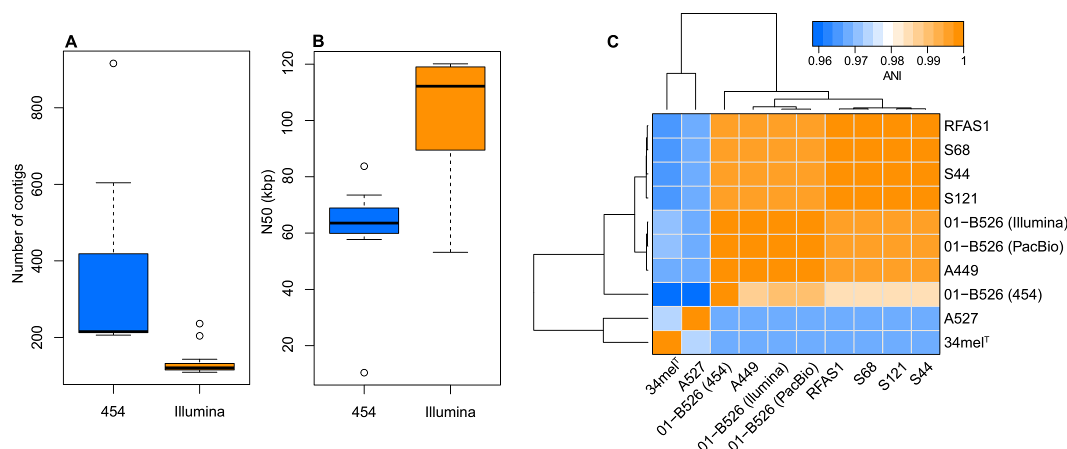

FIG. 1. Effects of sequencing technologies on genome sequences of *Aeromonas salmonicida* subsp. *salmonicida*. Box plots showing distribution of contig size (A) and of N50 values (B) for genome assemblies of *A. salmonicida* subsp. *salmonicida* available in GenBank. Information concerning genome sequences used is available in Supplementary Table S1. (C) Clustering and heat map based on average nucleotide identity values. Sequences used are: *A. salmonicida* subsp. *pectinolytica* 34mel^T (assembly: GCA_002735225.1), *A. salmonicida* subsp. *masoucida* RFAS1 (assembly: GCA_002313065.1), *A. salmonicida* A527 (assembly: GCA_002764135.1), *A. salmonicida* S44 (assembly: GCA_002214305.1), *A. salmonicida* S68 (assembly: GCA_002214265.1), *A. salmonicida* S121 (assembly: GCA_002214245.1), *A. salmonicida* subsp. *salmonicida* A449 (assembly: GCA_000196395.1), *A. salmonicida* subsp. *salmonicida* 01-B526 (454) (assembly: GCA_000234845.2) and *A. salmonicida* subsp. *salmonicida* 01-B526 (PacBio) (GenBank: CP027000). 01-B526 (Illumina MiSeq) sequences were only used for comparison and therefore were not deposited.

chromosome sequence using a combination of Canu version 1.6 [11] and Pilon version 1.22 [12].

To illustrate the biases caused by different sequencing technologies, the three assemblies of strain 01-B526 (454 pyrosequencing, Illumina MiSeq and PacBio SMRT) were compared to other sequences of *A. salmonicida* found in databases. For this purpose, a clustering was made using average nucleotide identity values (Fig. 1C). As expected, the sequence of strain 01-B526 generated by Illumina MiSeq and PacBio SMRT clustered together, close to the one of strain A449. Surprisingly, the sequence of strain 01-B526 obtained by 454 pyrosequencing was much more distant, even basal to the clade composed of sequences of *A. salmonicida* strains RFAS1, S44, S68 and S121, which are not of the *salmonicida* subspecies [13]. The tool used, pyani, aligns fragments of 1020 nucleotides by default for performing the comparisons. Knowing that 77% of the contigs generated from 454 pyrosequencing data were smaller than 1020 nt, this result was not totally unexpected and illustrates the influence of sequencing technologies on subsequent analyses and interpretations.

Accession number

The chromosome sequence of *Aeromonas salmonicida* subsp. *salmonicida* 01-B526 has been deposited in GenBank under accession number CP027000.

Acknowledgements

ATV received an Alexander Graham Bell Canada Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC). This project was funded by an NSERC Discovery grant (RGPIN-2014-04595) to SJC. SJC is also a research scholar of the Fonds de la Recherche du Québec–Santé. The authors thank Jeff Gauthier (Université Laval, Canada) for his critical reading of the report and for helpful comments.

Conflict of interest

None declared.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.nmni.2018.05.007.

References

[1] Dallaire-Dufresne S, Tanaka KH, Trudel MV, Lafaille A, Charette SJ. Virulence, genomic features, and plasticity of *Aeromonas salmonicida* subsp. *salmonicida*, the causative agent of fish furunculosis. Vet Microbiol 2014;169:1–7.

© 2018 The Author(s). Published by Elsevier Ltd, *NMNI*, **25**, 24–26
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[2] Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS One 2012;7, e30087.

[3] Long M, Nielsen TK, Leisner JJ, Hansen LH, Shen ZX, Zhang QQ, et al. Aeromonas salmonicida subsp. *salmonicida* strains isolated from Chinese freshwater fish contain a novel genomic island and possible regional-specific mobile genetic elements profiles. FEMS Microbiol Lett 2016;363. fnw190.

[4] Vincent AT, Boyle B, Derome N, Charette SJ. Improvement in the DNA sequencing of genomes bearing long repeated elements. J Microbiol Methods 2014;107:186–8.

[5] Vincent AT, Tanaka KH, Trudel MV, Frenette M, Derome N, Charette SJ. Draft genome sequences of two *Aeromonas salmonicida* subsp. *salmonicida* isolates harboring plasmids conferring antibiotic resistance. FEMS Microbiol Lett 2015;362:1–4.

[6] Reith ME, Singh RK, Curtis B, Boyd JM, Bouevitch A, Kimball J, et al. The genome of *Aeromonas salmonicida* subsp. *salmonicida* A449: insights into the evolution of a fish pathogen. BMC Genomics 2008;9:427.

[7] Charette SJ, Brochu F, Boyle B, Filion G, Tanaka KH, Derome N. Draft genome sequence of the virulent strain 01-B526 of the fish pathogen *Aeromonas salmonicida*. J Bacteriol 2012;194:722–3.

[8] Emond-Rheault JG, Vincent AT, Trudel MV, Brochu F, Tanaka KH, Attéré SA, et al. Variants of a genomic island in *Aeromonas salmonicida* subsp. *salmonicida* link isolates with their geographical origins. Vet Microbiol 2015;175:68–76.

[9] Emond-Rheault JG, Vincent AT, Trudel MV, Frey J, Frenette M, Charette SJ. AsaGEI2b: a new variant of a genomic island identified in the *Aeromonas salmonicida* subsp. *salmonicida* JF3224 strain isolated from a wild fish in Switzerland. FEMS Microbiol Lett 2015;362. fnv093.

[10] Bleidorn C. Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. Syst Biodivers 2016;14: 1–8.

[11] Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. Genome Res 2017;27:722–36.

[12] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 2014;9: e112963.

[13] Vincent AT, Rouleau FD, Moineau S, Charette SJ. Study of mesophilic *Aeromonas salmonicida* A527 strain sheds light on the species' lifestyles and taxonomic dilemma. FEMS Microbiol Lett 2017;364. fnx239.

© 2018 The Author(s). Published by Elsevier Ltd, *NMNI*, **25**, 24–26
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).