AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

METHODS

A High-Throughput Short Sequence Typing Scheme for *Serratia marcescens* Pure Culture and Environmental DNA

Thibault Bourdin,[a] Alizée Monnier,[a] Marie-Ève Benoit,[b] Emilie Bédard,[b] Michèle Prévost,[b] Caroline Quach,[c] Eric Déziel,[a] Philippe Constant[a]

[a]Institut National de la Recherche Scientifique, Centre Armand-Frappier Santé Biotechnologie, Laval, Québec, Canada
[b]Department of Civil, Geological, and Mining Engineering, Polytechnique Montréal, Montréal, Québec, Canada
[c]Université de Montréal, Montréal, Québec, Canada

ABSTRACT Molecular typing methods are used to characterize the relatedness between bacterial isolates involved in infections. These approaches rely mostly on discrete loci or whole-genome sequencing (WGS) analyses of pure cultures. On the other hand, their application to environmental DNA profiling to evaluate epidemiological relatedness among patients and environments has received less attention. We developed a specific, high-throughput short sequence typing (HiSST) method for the opportunistic human pathogen *Serratia marcescens*. Genes displaying the highest polymorphism were retrieved from the core genome of 60 *S. marcescens* strains. Bioinformatics analyses showed that use of only three loci (within *bssA*, *gabR*, and *dhaM*) distinguished strains with a high level of efficiency. This HiSST scheme was applied to an epidemiological survey of *S. marcescens* in a neonatal intensive care unit (NICU). In a first case study, a strain responsible for an outbreak in the NICU was found in a sink drain of this unit, by using HiSST scheme and confirmed by WGS. The HiSST scheme was also applied to environmental DNA extracted from sink-environment samples. Diversity of *S. marcescens* was modest, with 11, 6, and 4 different sequence types (ST) of *gabR*, *bssA*, and *dhaM* loci among 19 sink drains, respectively. Epidemiological relationships among sinks were inferred on the basis of pairwise comparisons of ST profiles. Further research aimed at relating ST distribution patterns to environmental features encompassing sink location, utilization, and microbial diversity is needed to improve the surveillance and management of opportunistic pathogens.

IMPORTANCE *Serratia marcescens* is an important opportunistic human pathogen, often multidrug resistant and involved in outbreaks of nosocomial infections in neonatal intensive care units. Here, we propose a quick and user-friendly method to select the best typing scheme for nosocomial outbreaks in relating environmental and clinical sources. This method, named high-throughput short sequence typing (HiSST), allows to distinguish strains and to explore the diversity profile of nonculturable *S. marcescens*. The application of HiSST profile analysis for environmental DNA offers new possibilities to track opportunistic pathogens, identify their origin, and relate their distribution pattern with environmental features encompassing sink location, utilization, and microbial diversity. Adaptation of the method to other opportunistic pathogens is expected to improve knowledge regarding their ecology, which is of significant interest for epidemiological risk assessment and elaborate outbreak mitigation strategies.

KEYWORDS molecular typing, HiSST, sink environment, opportunistic pathogens, neonatal intensive care unit (NICU), healthcare-associated infections (HAI), environmental DNA

Citation Bourdin T, Monnier A, Benoit M-È, Bédard E, Prévost M, Quach C, Déziel E, Constant P. 2021. A high-throughput short sequence typing scheme for *Serratia marcescens* pure culture and environmental DNA. Appl Environ Microbiol 87:e01399-21. https://doi.org/10.1128/AEM.01399-21.

Editor Christopher A. Elkins, Centers for Disease Control and Prevention

Address correspondence to Eric Déziel, eric.deziel@inrs.ca, or Philippe Constant, philippe.constant@inrs.ca.

Received 14 July 2021
Accepted 23 September 2021
Accepted manuscript posted online 29 September 2021
Published 24 November 2021

Interactions between patients and the hospital environment have gained attention in epidemiological studies aimed at identifying origins of nosocomial outbreaks. For instance, sink environments are recognized as a source of opportunistic pathogens in several health care-associated infections (HAI) (1–3). In preventive or outbreak investigations, molecular typing methods are commonly used to examine the relatedness of environmental or clinical isolates. Many typing techniques are available to achieve this goal (4–10), mostly based on multilocus sequence typing (MLST) methodologies initially developed by Maiden et al. (11). Democratization of high-throughput sequencing technologies have contributed to expand public genome databases, providing an unprecedented portrait of microbial diversity. This has led to the realization that the pangenome of bacterial species displays a mosaic landscape supporting the metabolic flexibility necessary to ensure species resistance and resilience toward disturbances. Such plasticity of microbial genome highlights the need to update and revisit conventional MLST schemes, typically relying on housekeeping genes. In some instances, these genes are not specific enough for accurate molecular typing of investigated strains (12, 13). In addition to lack of the resolution, sometimes the loci employed in the MLST schemes have been affected by recombination and do not properly genotype the isolates in question (14).

The genus *Serratia* consists of Gram-negative bacteria classified as members of the *Enterobacteriaceae* that are ubiquitous in water, soil, plants and different hosts, including insects, humans and other vertebrates (15, 16). Among *Serratia* species, *Serratia marcescens* is the most important opportunistic human pathogen, often multidrug resistant and involved in outbreaks of HAI in neonatal intensive care units (NICUs) (17–27). No MLST scheme exists for the molecular typing of *S. marcescens*, but other typing techniques have been used during previous epidemiological studies, such as pulsed-field gel electrophoresis (17, 19, 23), ribotyping (28) or more recently whole-genome MLST (29, 30). Even though these techniques were proven efficient to distinguish strains, they are not tailored to epidemiological surveys involving a large sample size because they are technically demanding due to upstream cultivation and isolation efforts.

This study introduces a new molecular typing approach, which we called high-throughput short sequence typing (HiSST), to detect and identify *S. marcescens* relying on culture-dependent and culture-independent applications. The HiSST method was developed based on whole-genome sequences of *S. marcescens* available in public databases then validated with reference culture collections, clinical isolates and environmental DNA samples.

RESULTS AND DISCUSSION

Design of the HiSST scheme. Thirty-two out of the 3,301 genes of *S. marcescens* pangenome were identified as the most variable with 29 to 30 alleles per gene. Only the most specific and discriminatory genes were kept, after stepwise alignment, examination of polymorphism among *S. marcescens* genomes and specificity check (Fig. 1). This led to the selection of 15 loci for subsequent analyses listed in Table S4. The minimal number of loci included in the HiSST scheme was selected by topology data analysis of concatenated HiSST loci and genome similarity. We found that the dendrogram built from three loci has a similar topology to the cognate clustering analysis comprising the 15 most discriminatory loci, according to the downward loci selection procedure. Indeed, Pearson's Chi-squared test with simulated *P* value shows a significantly similar cluster ($P < 0.001$) between HiSST SNPs and core genome MLST dendrograms (Fig. 2). Also, cophenetic correlation (0.55) and Goodman-Kruskal-gamma index (0.93) confirm that the two phylogenetic trees are similar, however not identical (Table S3). This led to the selection of the three loci located in genes *gabR*, *bssA* and *dhaM* for the HiSST scheme. PHI tests (57) show no significant recombination within the sequence set for each selected locus ($P > 0.10$), implemented with the PhiPack software and based on compatibility of parsimoniously informative sites (Table S3). Each locus was discriminant, with more than 20% of nucleotide dissimilarity between *S. marcescens*

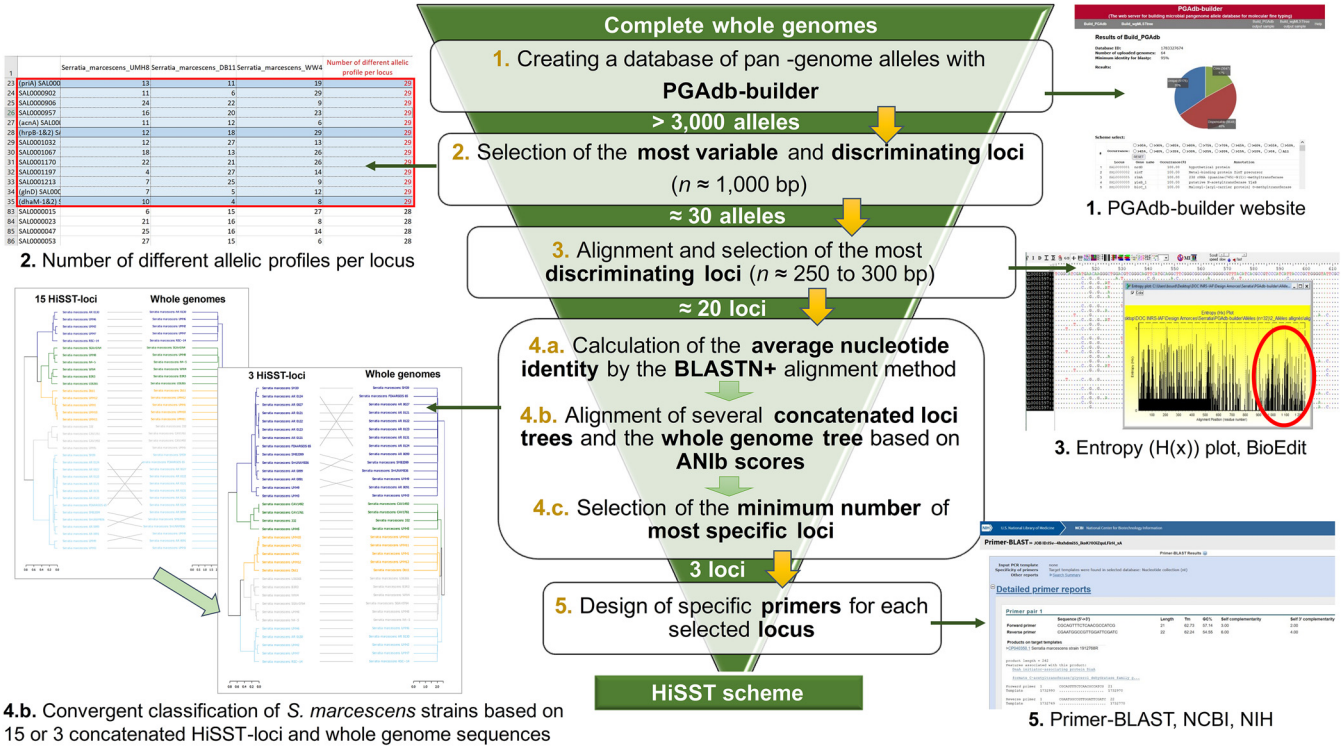

FIG 1 Step-by-step approach of the method used to develop the *S. marcescens* HiSST scheme. (A) A pan-genome allele database was created with the web service tool PGAdb-builder. (B) The 30 most variable and discriminating loci were selected by sorting loci with the highest number of different alleles using Excel. (C) Selected loci were aligned with the software BioEdit, and short gene fragments (<350 bp) displaying the highest variability were kept. (D) Three successive steps were necessary to define the trade-off between the number of different loci and specificity of the assay by comparison of concatenated HiSST loci and genome similarity. (E) Finally, oligonucleotides comprising 18- to 22-mers with either a single or no substituted base were designed to target discriminant internal loci and tested with the software tool Primer-BLAST.

strains and other species (Fig. S1). The locus *gabR* is more specific to *S. marcescens* species than *bssA*, followed by *dhaM*, with, respectively, more than 26%, 17% and 14% of nucleotide dissimilarity with *S. ficaria*, the closest relative species of *S. marcescens* for these loci.

The HiSST scheme based on these 3 loci differentiates most *S. marcescens* strains better than the average nucleotide identity (ANI) score based on their whole genomes (Fig. 3). The pairwise ANIb genome similarity score is over 94% in its ability to distinguish *S. marcescens* strains from other *Serratia* species while only a few strains of *Serratia* spp. have more than 70% nucleotide identity with the three selected loci of *S. marcescens*. Classification of *S. marcescens* strains based on the HiSST scheme is congruent with classification scheme relying on complete genome sequences (Fig. 2). Some strains sharing an ANI score of more than 99.9% are not properly clustered by the HiSST scheme compared to those with a complete genome. This difference is expected due to the overall limitation of molecular methods based on few marker genes. At such a high level of similarity, the threshold delineating species, strains or clones is empirical, depending on the examined species. For example, *P. aeruginosa* has high genomic plasticity mainly due to frequent horizontal gene transfers (58, 59), while *S. marcescens* has a higher genetic diversity at the sequence level according to PGAdb-builder results, with also genome flexibility (60). Additional factors to consider include the study context (e.g., the precautionary principle for epidemiological studies tends to identify highly similar but not identical strains as a nonclonal strain) and the method used (i.e., depending on the sensitivity of the molecular typing method and the sequencing platform used, the evolution of the technology, and knowledge). As a whole, the minimal similarity threshold among *S. marcescens* strains is 89% for the three loci (Fig. S2 and based on BLAST results).

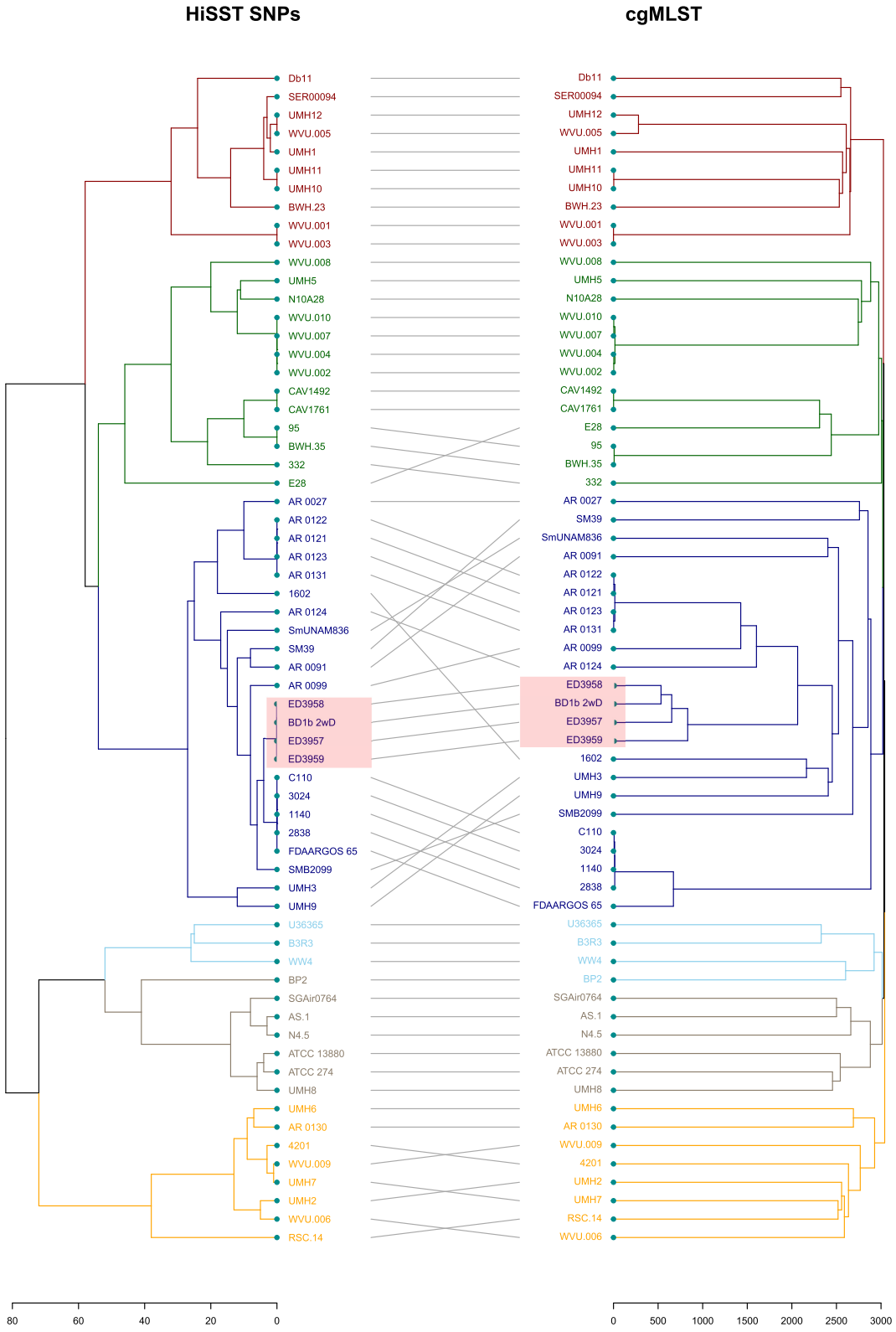

FIG 2 Convergent classification of *S. marcescens* strains based on the HiSST scheme and core genome MLST. Complete linkage dendrograms based on maximum distance of concatenated loci selected for the HiSST scheme (A) and the core genome MLST difference to discriminate strains of *Serratia marcescens* (B). Label colors correspond to the optimal number of clusters for HiSST SNPs, according to Rousseeuw's Silhouette quality index. Labels in the shaded red box correspond to the cluster of four strains isolated in this study, suggesting that the environmental BD1b-2wD strain and clinical isolates (ED3957, ED3958, and ED3959) descend from a single cell.

Validation and application of the HiSST scheme. Specificity of primers targeting *gabR*, *bssA* and *dhaM* loci was first confirmed by BLAST searches against the RefSeq database of nonredundant protein sequences. The efficacy and the specificity of the PCR assays was further confirmed with reference strains (Fig. S3). A PCR amplicon of the correct size was observed for all *S. marcescens* strains ($n = 15$), but not for other *Serratia* species or other species of gammaproteobacteria. An accuracy test of the HiSST scheme, including the bioinformatic procedure utilized to assign alleles to SST, was realized with *S. marcescens* reference strains Db11 and Db10. Genomic DNA from both strains was subjected to PCR amplicon sequencing with an average allocation of 1,000 reads per library. The HiSST profile (sequence type 1 [ST 1]) of both strains corresponded to the expected profile with a single amplicon sequence variant (ASV) for each gene, supporting the accuracy of the HiSST procedure and parameters utilized in sequence quality control (Fig. 4).

The method was applied to two different case studies conducted in the same NICU. The first case study sought to compare the ST profile of a strain isolated from the sink-drain environment (BD1b-2wD) with three clinical strains (ED3657, ED3658, and ED3659) from patients admitted in that NICU, where an outbreak occurred. The infection prevention and control team determined their relatedness based on the PFGE profiles. Molecular typing of the environmental strain BD1b-2wD and the clinical strains revealed very close relatedness, with all four having an identical HiSST profile, ST 47 (Fig. 5A). WGS was done for each strain to challenge the HiSST scheme result. Pairwise comparison of contiguated genomes confirmed the high degree of similarity between each strain (ANIm $>$ 99.7%). In principle, *bssA* and *gabR* are sufficient to ensure diversity coverage of ST represented in the genome database (Fig. 5B), but inclusion of *dhaM* in the HiSST scheme is included to prevent false-negative results (i.e., in the case where the targeted gene is absent or subject to unknown mutations) and allows the distinguishing of an environmental or clinical origin of *S. marcescens* strains for culture-based diagnostics (Fig. 5A). These results suggest that the environmental BD1b-2wD strain and clinical isolates descend from a single cell, while providing supplementary experimental evidence supporting the specificity of the HiSST scheme (Fig. 2).

The second case study was conducted to explore the diversity of *S. marcescens* by applying the HiSST method to environmental DNA (eDNA). PCR amplicon sequencing of each locus was done to report the diversity of each "locus short sequence type" number (locus-SST) separately for a culture-independent epidemiological investigation. All retrieved ASV sequences were specific to *S. marcescens*. Diversity among the 19 sinks was modest, with 11, 6 and 4 different alleles of *gabR*, *bssA* and *dhaM* found, respectively (Fig. 4). A single allele was dominant in each sample, with a relative abundance of 70–100% (Table S2). Either a single or two allele(s) per sample were observed for *gabR* and *dhaM* loci, whereas *gabR* was represented by up to three different alleles per sample. For the three HiSST loci, rare alleles differ from the dominant allele in the same sample by 1 to 6 SNPs, suggesting the presence of other strains in the drain. Artificial inflation in diversity caused by sequencing errors is less likely due to the stringent filtering process of sequences (cf. Materials and Methods) and the low error probability of incorrect base-call for short sequences (61). The intercomparison of ST profiles among the 19 sinks of the NICU was done to infer potential epidemiological links (Fig. 6). The most straightforward link between sink environments is the case where ST profiles are identical. This situation was observed in sinks #72 and #73 for dominant ASVs (*bssA*-SST 36, *gabR*-SST 18, and *dhaM*-SST 2) that are likely colonized by the same *S. marcescens* strain. This link is supported by the proximity of both sinks in the NICU, with the same drain connection and interconnection through handwashing (62, 63). The sink PLM shared two loci-SST detected in sinks #72 and #73 (*bssA*-SST 36 and *dhaM*-SST 2) and two loci-SST in sink #80 (*gabR*-SST 2 and *dhaM*-ST 2). This result suggests an epidemiological link between the four sinks related to one another by the sink PLM (that is used for the initial handwashing at the NICU entrance). Finally, the HiSST profile (ST 2) of sink #80 is identical to that of *S. marcescens* 95 and BWH-35 strains included in the reference genome database,

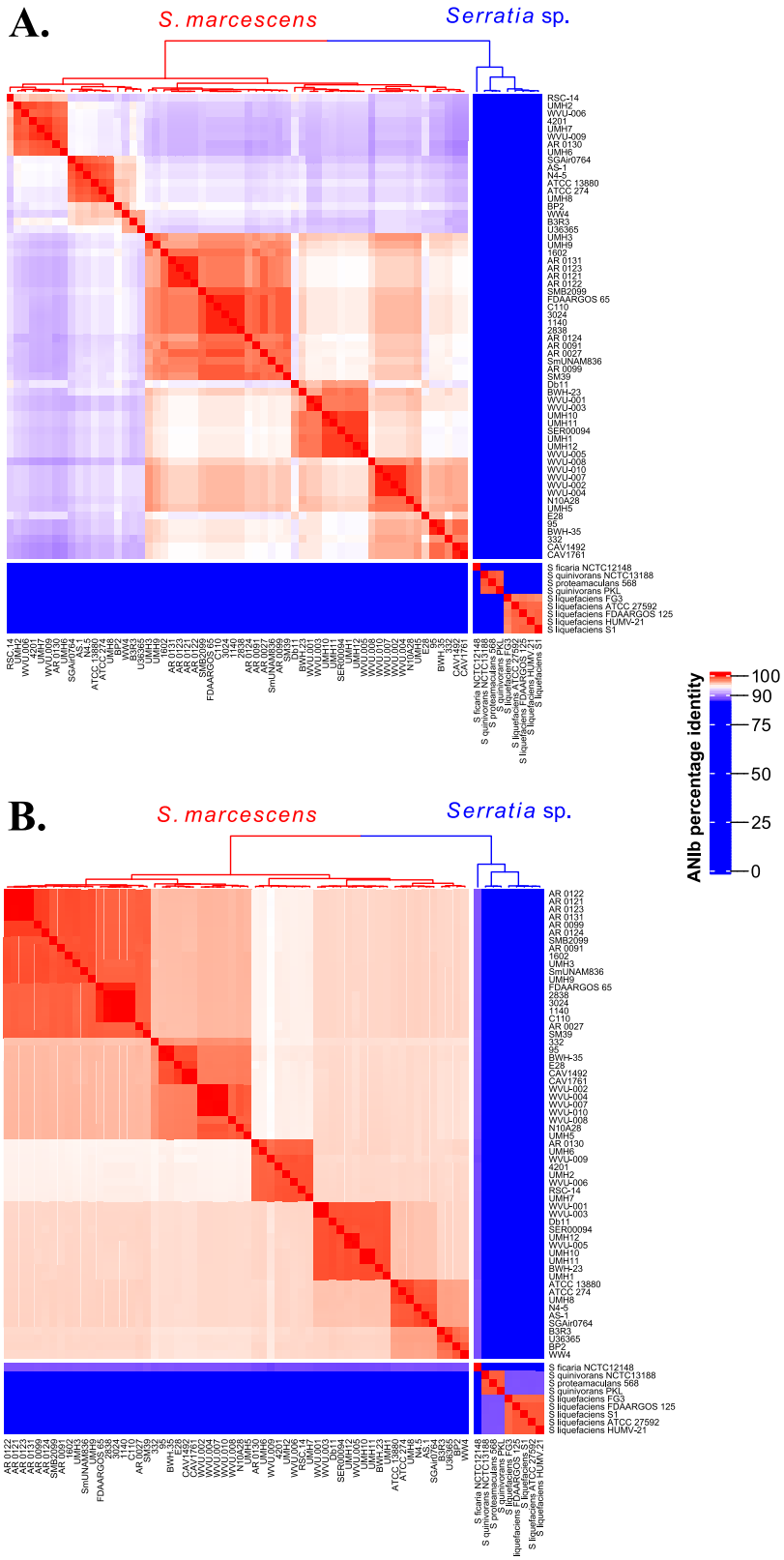

FIG 3 Discrimination of *Serratia* spp. based on the HiSST scheme and whole-genome sequences. The heat-map reports the ANIb score of the three concatenated loci of the HiSST scheme (A) and genome similarity (B). *S. ficaria* (n = 1), *S. quinivorans* (n = 2), *S. proteamaculans* (n = 1) and *S. liquefaciens* (n = 5) were included as the outgroup. The more cells turn from blue to red, the higher the ANIb score, and the more the strains are genotypically similar.

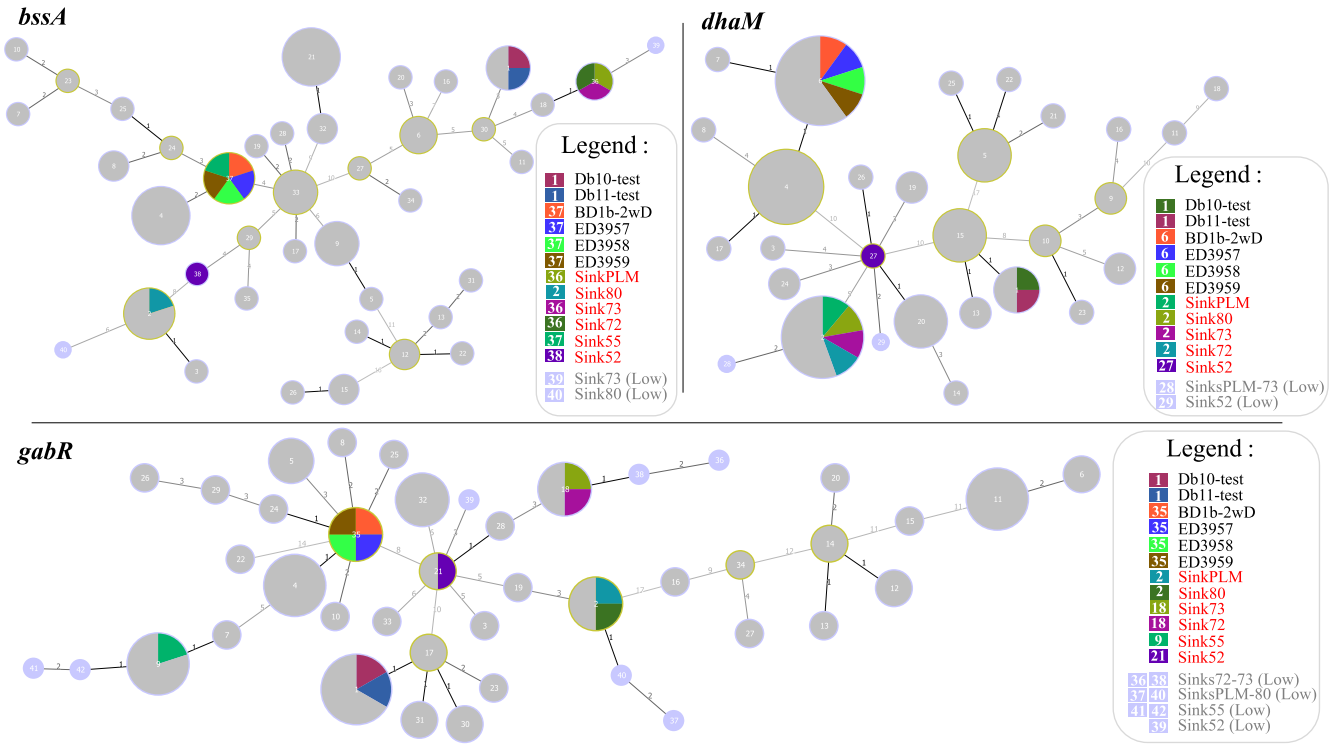

FIG 4 Minimum spanning trees based on SNP analysis of *S. marcescens* and eDNA, using *S. marcescens* Db10 as a reference. The distance labels represent the number of discriminating SNPs between neighboring genotypes. Each pie chart label refers to SST identifier of the corresponding locus. Reference genomes are represented in gray and isolates or sampled sinks are represented by the color legend in pie charts. Dominant SSTs of eDNA are represented with red font in the legend, whereas gray font corresponds to the SST of eDNA in low abundance.

suggesting the colonization by a taxonomically closely related strain. *S. marcescens* 95 and BWH-35 were isolated from sputum in a Boston hospital (Massachusetts, USA) and are most likely variants of the same strain.

A limitation of the method was noticed in sink #55 where no PCR detection of *dhaM* was observed with positive amplification of *gabR* and *bssA* genes. Although this can be explained by the low level of *S. marcescens* in this sink combined with different amplification efficiencies between the three reactions, examination of future genome sequences deposited in public databases will be necessary to confirm the ubiquitous distribution of *dhaM* in *S. marcescens*.

These case studies illustrate the strengths of the HiSST scheme to identify clones and its broad applicability for epidemiological investigations. Beyond the conventional application of the method to genotype isolates, examination of eDNA offers a complementary tool for the source tracking of opportunistic pathogens. This could be done by the monitoring of bacterial succession in the NICU environment and patient samples through HiSST eDNA profiling. Under that framework, a convergence of HiSST profiles along spatial or temporal sampling sequences would provide strong evidence of opportunistic pathogen transfer across different environments.

In contrast to conventional application for isolate identification, HiSST profile analysis from eDNA is less prone to misinterpretation or aborted analysis for samples displaying no signal for certain genes. Indeed, the pairwise comparison of HiSST bacterial profiles can be expressed as a pairwise Jaccard distance computed with presence or absence score for detected or nondetected loci-SST, respectively. Downstream clustering and multivariate analyses offer options to correlate loci-SST distribution patterns with environmental features encompassing sink location, utilization, and microbial diversity (Fig. 6). Although this approach is a gold standard in microbial ecology, the second case study presented in this article is the first culture-independent application of loci-SST profile analysis of opportunistic pathogens for epidemiologic survey.

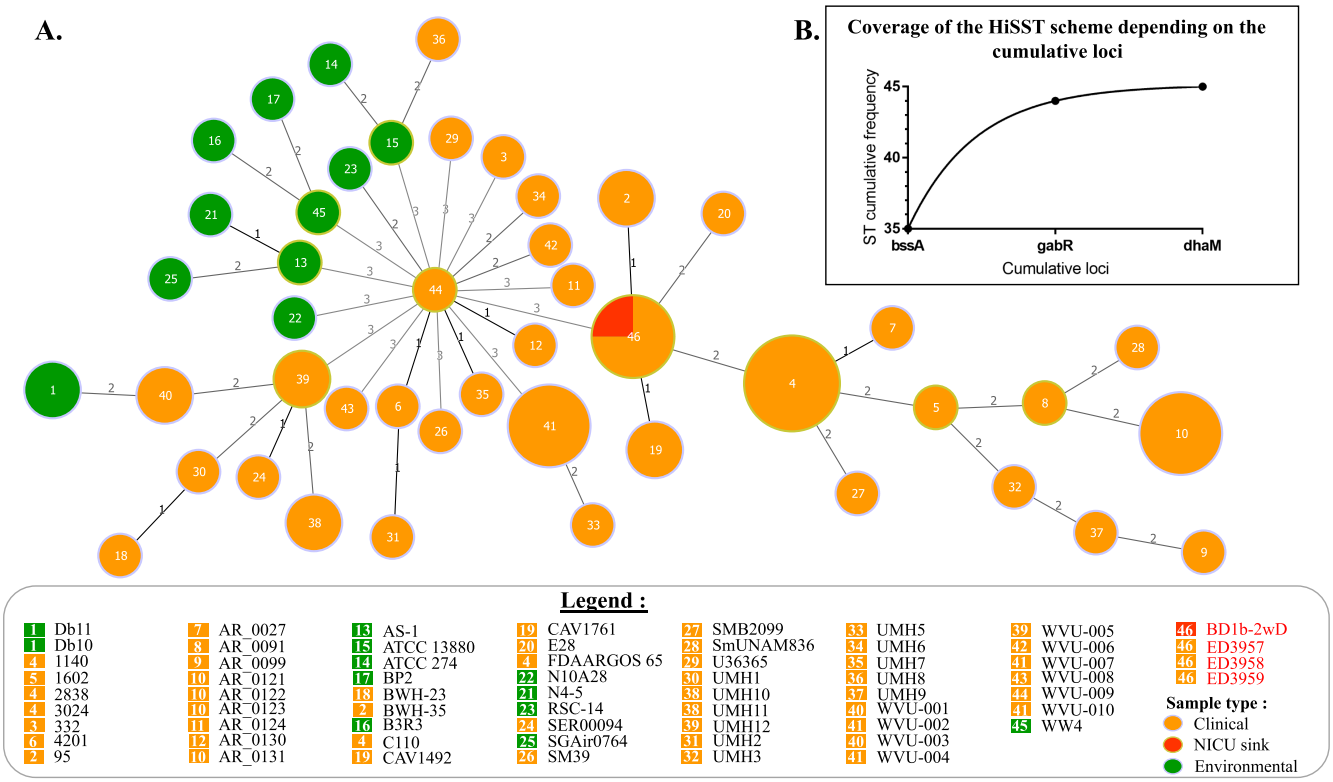

FIG 5 Relationship among the ST profiles of reference strains and isolates and diversity coverage of the HiSST scheme. (A) A minimum spanning tree based on MLST analysis of the HiSST scheme is represented with distance labels corresponding to the number of discriminating alleles and pie chart labels referring to the ST identifier of the HiSST scheme. Orange nodes correspond to clinical isolates, the red node to the isolate from the NICU sink drain, and the green nodes to environmental isolates. In the legend, strains represented by red font correspond to unknown clinical (ED3957, ED3958, and ED3959) and sink drain (BD1b-2wD) isolates from this study. (B) Cumulative frequency of ST depending on the number of loci included in the HiSST scheme.

In conclusion, a combination of *in silico* analyses led to the development of a powerful HiSST assay to identify isolates of *S. marcescens* strains. The approach relying on pangenome examination rather than selection of conventional housekeeping genes contributed to the method specificity. For instance, conventional MLST schemes for *P. aeruginosa* and *S. maltophilia* are less specific than the HiSST method developed here for *S. marcescens*. Application of the procedure presented in this article to these other opportunistic pathogens of environmental origin led to more robust HiSST schemes (T. Bourdin, E. Déziel, and P. Constant, unpublished data). Despite the precision of the method presented here, specificity and coverage of the HiSST scheme will require regular validation and updates with the addition of new genome sequences in public databases. The bioinformatic pipeline implemented here or alternative methods (64–66) will facilitate regular updates of the HiSST scheme. This fact holds true for any molecular classification tool. Even though comparison of whole genomes appears to be the most robust method (12), public genome databases contain contaminations that may introduce biases for the identification of highly similar strains (67). In addition, the high proportion of similar or identical genes in whole genomes hides some dissimilarities between isolates, while HiSST highlights the most discriminating alleles. On the other hand, HiSST and other molecular typing methods (based on PCR products) cannot achieve the accuracy of whole-genome sequencing. Thus, a combination of whole-genome sequencing and a high discriminatory molecular typing method is recommended for culture-dependent epidemiological investigation (68). Beside isolate identification, the HiSST method proved efficient for loci-SST comparison and source tracking purposes of *S. marcescens* in eDNA samples without the need for culture.

Based on these results, the following epidemiological interpretations for molecular typing of isolates when using the HiSST scheme are proposed: (i) isolates that are

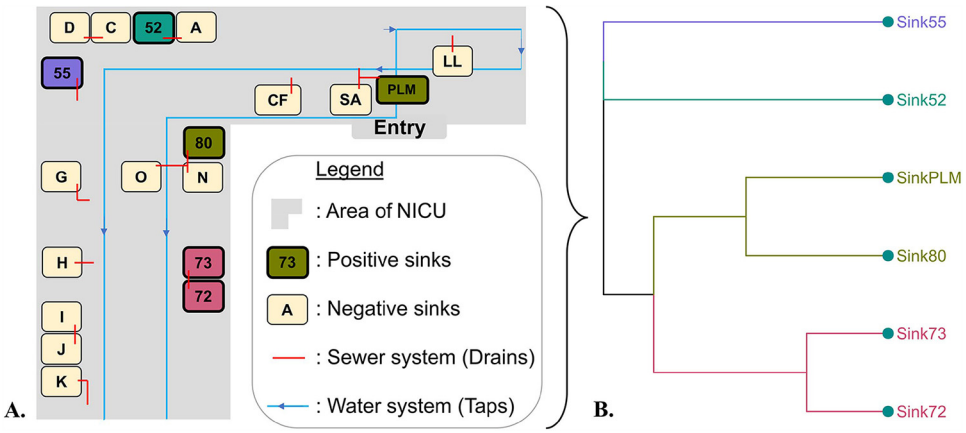

FIG 6 Survey of *Serratia marcescens* in the sink drains of a NICU. A scheme of the surveyed NICU is depicted (A) along with an UPGMA dendrogram (B) based on the Jaccard distance computed with the HiSST profiles of *gabR*, *bssA* and *dhaM* loci among sink drains that showed positive PCR amplifications. Sink drains sampled in this study are symbolized by a rectangle with their identification in black letters. Each sink color corresponds to a HiSST genotypic profile, as seen in the dendrogram (B). If the color is identical between two sinks, then there is a higher probability that these sinks are colonized with the same strain(s), which indicates a potential relationship between these sink drains.

identified by at least 2 of 3 loci-SST are confirmed as *S. marcescens*, (ii) isolates with an identical ST (i.e., identical *gabR*-SST, *bssA*-SST and *dhaM*-SST) are most likely clones and belong to the same genotype, and (iii) isolates that differ by 2 or 3 loci-SST are mostly unrelated and do not belong to the same genotype. For an epidemiological survey on eDNA samples when using the HiSST scheme described here, the following interpretations are proposed: (i) eDNA samples with loci-SST corresponding to the HiSST scheme indicate the presence of *S. marcescens*, (ii) eDNA samples with several loci-SST of one locus indicate the presence of several *S. marcescens* strains, and (iii) samples with an identical ST (i.e., identical *gabR*-SST, *bssA*-SST and *dhaM*-SST) contain very closely related strains and the sampling environment is most likely linked.

MATERIALS AND METHODS

Development of the HiSST scheme. A pan-genome allele database was assembled from 60 complete genomes of *S. marcescens* retrieved from the NCBI GenBank database (last updated in July 2020) with the *Build_PGAdb* module available on the online tool PGAdb-builder (31). Conserved genes showing the highest number of alleles were selected as the most variable and discriminant. Alleles of each selected genes (n = 32) were aligned, nonoverlapping ends were removed and the sequence identity matrix was computed with the software BioEdit (32). Gene fragments (<350 bp) displaying the highest variability were chosen and aligned against the NCBI database with the Basic Local Alignment Search Tool (BLAST) to assess specificity. The 15 loci (i.e., nucleotide sequences of internal fragments of the previously selected genes) showing the highest variability and with the most specific nonoverlapping ends were selected as candidates for the HiSST scheme. A trade-off between the number of different loci and specificity of the assay was achieved by qualitative visual comparison and tree topology tests of concatenated HiSST loci and genome similarity (see below). Three successive steps were necessary to implement the approach relying on the 60 complete genomes of *S. marcescens* (Table S1) and 9 other strains of non-*marcescens Serratia* spp. available in GenBank. First, ANIb (33) analyses were performed on the 69 complete genomes with the BLAST+ alignment tool (https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) with the Python package *pyani* (https://github.com/widdowquinn/pyani) (34). Second, a stepwise approach was implemented to assemble concatenated loci alignments. A backward selection procedure was applied on the 15 most discriminant loci until only three loci remained. This led to multiple concatenated alignments comprising either 15, 7, 4 or 3 gene fragments. ANIb scores were calculated on the concatenated alignments of loci. Third, the discriminating power of the four different HiSST schemes was validated by topology data analysis of concatenated loci and complete genome trees. The topology of the different complete linkage dendrograms, based on maximum distance (i.e., supremum norm), were compared with R version 4.0.4 (35) using the packages *pvclust* (36), *dendextend* (37), *vegan* (38) and *tidyverse* (39). Topological distance was measured by Robinson and Foulds distance, and then tree topology tests were performed by calculating cophenetic correlation and Goodman-Kruskal-gamma index. A contingency table of the two topologies was created using k-mean partitioning approach, repeated 1,000 times. Thus, a Pearson's Chi-squared test was performed to verify the relationship between the two dendrograms, as described in Borcard et al. (40).

TABLE 1 HiSST locus specific primers sequences and PCR cycle conditions[a]

Locus	Primer sequence (5′-3′)	PCR amplicon length (bp)	PCR cycle conditions
gabR	Forward: GAGCATCTGCGYAATATGCG Reverse: CAGCGCGYTGAACACCTG	318	Initial denaturation at 95°C for 5 min, followed by 40 cycles at 95°C for 20 s, 58°C for 40 s, 72°C for 30 s and a final extension period of 5 min at 72°C.
bssA	Forward: CGCAGTTTCTCAACGCYATCG Reverse: CGAATGGCCGTTGGATTCGATC	242	Initial denaturation at 95°C for 5 min, followed by 35 cycles at 95°C for 20 s, 58°C for 40 s, 72°C for 30 s and a final extension period of 5 min at 72°C.
dhaM	Forward: GGCGTCCAGCATYGCCTT Reverse: GACGTGCGCGACATGCTG	279	Initial denaturation at 95°C for 5 min, followed by 35 cycles at 95°C for 20 s, 60°C for 40 s, 72°C for 30 s and a final extension period of 5 min at 72°C.

[a]Illumina linker sequences were added at each 3′-end sequence of primers: 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′ for forward and 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3′ for reverse primers.

Further, ANIb values of the selected HiSST loci and complete genomes were compared using the packages *circlize* (41) and *ComplexHeatmap* (42), which help to visualize the difference of the discriminatory power between the HiSST scheme and the whole genome of *S. marcescens*. The R scripts we developed are available on GitHub (https://github.com/TBourd/R_scripts_for_HiSST_scheme). Following these analyses, the three gene fragments of *gabR* (HTH-type transcriptional regulatory protein), *bssA* (benzylsuccinate synthase alpha subunit) and *dhaM* (PTS-dependent dihydroxyacetone kinase, phosphotransferase subunit) were retained for the further development of the HiSST scheme, as described below.

Primer design and PCR amplification of *gabR*, *bssA* and *dhaM* internal loci. Oligonucleotides comprising 18- to 22-mers with either a single or no substituted base were designed to target discriminant internal loci in the *gabR*, *bssA* and *dhaM* genes (Table 1). *In silico* tests of primers were performed with the software tool "Primer-BLAST" (43), using the RefSeq nonredundant proteins database to assess specificity and the *Serratia marcescens* subset RefSeq database to verify the coverage of primers for this species. The reaction was carried out in 25 μl of master mix containing 2.5 U/μl Fast-*Taq* DNA polymerase (Bio Basic Inc., Markham, Ontario, Canada), 1× of Fast-*Taq* buffer (Bio Basic Inc., Markham, Ontario, Canada), 200 μM dNTPs, 0.4 mg/ml bovine serum albumin (BSA), 0.4 μM each primer, and 2 ng/μl extracted DNA. A solution of 0.5× band sharpener (Bio Basic Inc., Markham, Canada) was included for the *gabR* mixture only. PCR conditions were optimized for each primer set with genomic DNA of *S. marcescens* strains as the template (Table 1).

Validation of the HiSST scheme with reference strains. Validation of primers was done with 28 reference strains of various origins (Table 2). Selected strains comprised *S. marcescens* (n = 15), *Serratia rubidaea* (n =1), *Serratia liquefaciens* (n = 1), *Serratia plymuthica* (n = 1), *Pseudomonas aeruginosa* (n = 3), *Klebsiella pneumoniae* (n = 1), *Stenotrophomonas maltophilia* (n = 4), *Stenotrophomonas acidaminiphila* (n = 1) and *Stenotrophomonas nitritireducens* (n = 1). The strains were purified on Trypticase soy broth (TSB) (Difco Laboratories, Sparks, MD, USA) with agar (15 g/liter) (Alpha Biosciences, Inc., Baltimore, MD, USA) at 30°C for 48 h. A single colony of each strain was inoculated in 2 ml TSB and grown for 48 h at 30°C for subsequent genomic DNA extraction.

Validation of the HiSST scheme with environmental DNA. Biofilm and 50 ml of water from 10 different sink drains were sampled in April 2019 during an outbreak of *S. marcescens* in a neonatal intensive care unit (NICU) in a Montreal Hospital (Québec, Canada). The same day, samples were inoculated on a semiselective DNase test agar (44) supplemented with ampicillin (5 μg/ml), colistin (5 μg/ml), cephalothin (10 μg/ml), and amphotericin B (2.5 μg/ml) incubated for 48 h at 30°C. Colonies were purified on TSB with agar at 30°C for 48 h. During a second sampling campaign, biofilm and water (50 ml) samples from sink drains were collected twice from 19 sinks in January 2020. Samples were kept on ice during their transportation to the laboratory. Genomic DNA from isolated strains and environmental samples was extracted by a procedure combining mechanical and chemical lysis, using bead beater and ammonium acetate treatment, as previously described (45), prior to PCR amplicon sequencing. The two successive PCR amplifications necessary for the preparation of *gabR*, *bssA* and *dhaM* sequencing libraries were conducted with the AccuPrime *Taq* DNA polymerase system, high fidelity (Invitrogen Ltd., Carlsbad, CA, USA). PCR conditions and reaction mixtures were adapted following manufacturer instructions (Table 1). The first PCR was performed using modified *gabR*, *bssA* and *dhaM* primers, including Illumina linker sequences (Table 1) and 2 ng/μl of template DNA. PCR products were purified with AMPure XP beads (Beckman Coulter Inc., Brea, CA, USA). Purified PCR products were subjected to a second PCR performed for library preparation using barcoded primers (Table S2) supplied by Integrated DNA Technologies Inc. (Mississauga, Ontario, Canada). Purified PCR amplicons were quantified using the Quant-iT PicoGreen dsDNA assay kit (Invitrogen Ltd., Carlsbad, CA, USA), diluted and pooled into 75 μl comprising a final concentration of 1.5 ng/μl of DNA before shipping for sequencing. PCR amplicons were sequenced with the Illumina MiSeq PE-250 platform at the Centre d'Expertise et de Services Génome Québec (Montréal, Québec, Canada). Raw sequencing read processing included primer sequence removal with the software Cutadapt v2.10 (46), followed by quality control, paired ends merging and chimera check using the default parameters specified in the package dada2 v1.8.0 (47) that include packages ShortRead v1.48.0 (48) and Biostrings v2.58.0 (49). Reads containing a mismatch in the primer region were deleted (R script available at https://github.com/TBourd/R_scripts_for_HiSST_scheme). Filtered sequences were clustered into amplicon sequence variants (ASVs) displaying 100% identity. ST assignation of chimera-free ASV

TABLE 2 Reference strains utilized as a positive or negative control for HiSST scheme validation[h]

Species and strains (no. of strains or additional designation)	Lab collection no.	Isolation origin (country)	Provided by
Serratia marcescens (*n* = 15)			
PCI 1107 (ATCC 14756, LMG 13576)	ED3691	Fort Detrick, Maryland (USA)	BCCM-LMG[c]
BD1b-2wD	ED4305	Drain water from a NICU (Canada)	Lab strain
Db11	ED3837	Insect isolate, *Drosophila melanogaster* (France)	A. Brassinga[a] ; J. J. Ewbank[b]
Db10	ED3838	Insect isolate, *Drosophila melanogaster* (France)	A. Brassinga[a] ; J. J. Ewbank[b]
BS 303 (ATCC 13880, LMG 2792)	ED3696	Pond water (Czech Republic)	BCCM-LMG[c]
L00128734	ED3957	Human clinical specimen (Canada)	LSPQ[d]
L00128736	ED3958	Human clinical specimen (Canada)	LSPQ[d]
L00128737	ED3959	Human clinical specimen (Canada)	LSPQ[d]
L00128966	ED3960	Human clinical specimen (Canada)	LSPQ[d]
L00128967	ED3961	Human clinical specimen (Canada)	LSPQ[d]
L00129585	ED3962	Human clinical specimen (Canada)	LSPQ[d]
L00130169	ED3963	Human clinical specimen (Canada)	LSPQ[d]
L00133794	ED3964	Human clinical specimen (Canada)	LSPQ[d]
L00134617	ED3965	Human clinical specimen (Canada)	LSPQ[d]
L00085643	ED3966	Environmental (Canada)	LSPQ[d]
Serratia rubidaea (*n* = 1)			
FB299	ED3693	Environmental (USA)	Roger Laurent Bernier[i]
Serratia liquefaciens (*n* = 1)			
ID150497	ED3967	Human clinical specimen (Canada)	LSPQ[d]
Serratia plymuthica (*n* = 1)			
ID157970	ED3968	Human clinical specimen (Canada)	LSPQ[d]
Stenotrophomonas maltophilia (*n* = 4)			
560 (ATCC 13636, LMG 961, NCTC 10258)	ED3699	Human, cerebrospinal fluid (USA)	BCCM-LMG[c]
L00083595	ED3969	Human clinical specimen (Canada)	LSPQ[d]
L00092250	ED3970	Human clinical specimen (Canada)	LSPQ[d]
L00124341	ED3971	Human clinical specimen (Canada)	LSPQ[d]
Stenotrophomonas acidaminiphila (*n* = 1)			
L00129488	ED3979	Human clinical specimen (Canada)	LSPQ[d]
Stenotrophomonas nitritireducens (*n* = 1)			
ATCC BAA-12 (LMG 22074, DSM 12575)	ED3701	Laboratory scale biofilter (Germany)	BCCM-LMG[c]
Pseudomonas aeruginosa (*n* = 3)			
UCBPP-PA14	ED1	Human clinical specimen (USA)	69
PAO1	ED956	Human, wound (Australia)	Sylvie Chevalier[e]
FKS4A	ED0129	Human, cystic fibrosis (USA)	Luke Hoffman[f]
Klebsiella pneumoniae (*n* = 1)			
ATCC 4352 (LMG 3128)	ED3692	Cow's milk	ATCC[g]

[a]Ann Brassinga, University of Manitoba, Winnipeg, MB, Canada.
[b]Jonathan J. Ewbank, University of Aix-Marseille, Marseille, France.
[c]Belgian Coordinated Collections of Microorganisms, University of Ghent, Ghent, Belgium.
[d]Laboratoire de Santé Publique du Québec, Sainte-Anne-de-Bellevue, Québec City, Québec, Canada.
[e]Sylvie Chevalier, University of Rouen-Normandie, Rouen, France.
[f]Luke Hoffman, Seattle Children's Hospital, Seattle, WA, USA.
[g]American Type Culture Collection, Rockville, MD, USA.
[h]Ten strains of *Serratia* spp. were included to verify the specificity for *S. marcescens* of each selected locus, with *S. ficaria* (*n* = 1), *S. liquefaciens* (*n* = 5), *S. quinivorans* (*n* = 2), and *S. proteamaculans* (*n* = 2) also downloaded from the NCBI GenBank database. These *Serratia* spp. have the most similar nucleotide sequences of the selected locus according to the results of the BLAST run.
[i]Gift of Roger Laurent Bernier (Burlington, CA, USA).

was done using *gabR*, *bssA* and *dhaM* reference databases with a 100% identity cutoff (Table S1). The proportion of reads remaining after each step of the bioinformatics pipeline is provided in Table S2.

SNP and HiSST profile analyses. SNPs of each locus were analyzed from all unique nucleotide sequences for references strains and environmental DNA (Table S3). The SNP matrix and HiSST profile were analyzed by the geoBURST algorithm using the PHYLOViZ platform v2.0a (50), creating minimum spanning trees using default software settings. The coverage of the HiSST scheme was visualized in a chart representing the cumulative frequency of ST depending on the number of cumulative loci, based on 60 reference strains of *S. marcescens* (Table S1).

Validation of molecular typing by whole-genome sequencing (WGS). Environmental strain BD1b-2wD and clinical strains ED3957, ED3958, and ED3959 were subjected to whole-genome sequencing (WGS) with the Illumina NextSeq 550 platform at the Microbial Genome Sequencing Center (Pittsburgh, PA, USA). A quality control of the WGS data was checked with FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc). Illumina adapter clipping and quality trimming were performed with Trimmomatic v0.39 (51), by specifying an average quality required greater than 30. Then, genomes were

assembled from a FASTQ file of paired-end reads using SPAdes *de novo* assembler (52) and Bandage (53) to visualize SPAdes graphical output. Contigs obtained from SPAdes output were aligned, ordered and oriented with the closest reference genome (*S. marcescens* AR_0122) to create a contiguated genome by using the tool ABACAS (54). Pairwise comparison of contiguated genomes of DB1b-2wD, ED3957, ED3958 and ED3959 was performed using ANIm scores (calculation of ANI based on the MUMmer algorithm, which is more adapted to comparing genomes with a high degree of similarity) (55, 56).

HiSST nomenclature and assignation. The following nomenclature is used to identify HiSST loci and HiSST profiles: a "locus short sequence type" number (locus-SST) is assigned for each ASV of the individual loci, and the combination of multilocus SST of the overall HiSST profile is defined by a "sequence type" (ST) number. The HiSST identification of an isolate is performed with an R script, and new HiSST profiles for unknown isolates are added to the HiSST database using the same script. The HiSST scheme database and the R script called "HiSST-Assignation" are available on GitHub at https://github.com/TBourd/R_scripts_for_HiSST_scheme.

Data availability. Raw sequencing reads have been deposited in the NCBI Sequence Read Archive under BioProject no. PRJNA729113. Raw sequencing reads of isolates are deposited under BioSample no. SAMN19117128 to SAMN19117139, and eDNA raw sequencing reads are under BioSample no. SAMN19110658 to SAMN19110711. Assembled genomes of the environmental strain BD1b-2wD and clinical strains ED3957, ED3958, and ED3959 have been deposited under the same BioProject accession no., PRJNA729113, and under BioSample accession no. SAMN19232018 to SAMN19232021.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 1.3 MB.

SUPPLEMENTAL FILE 2, XLSX file, 2.5 MB.

SUPPLEMENTAL FILE 3, XLSX file, 0.03 MB.

SUPPLEMENTAL FILE 4, XLSX file, 0.1 MB.

SUPPLEMENTAL FILE 5, XLSX file, 0.04 MB.

ACKNOWLEDGMENTS

We thank the hospital staff for help in sampling, Ann Brassinga (Department of Microbiology, University of Manitoba), Jonathan J. Ewbank (Centre d'Immunologie de Marseille-Luminy, Aix-Marseille University, Marseille, France), Sabine Favre-Bonté (Université Lyon 1, UMR, CNRS 5557, Ecologie Microbienne, Lyon, France), and the Laboratoire de Santé Publique du Québec for providing reference strains.

This work was supported by NSERC and CIHR through the IRC Industrial Chair on Drinking Water and by funding from the Collaborative Health Research Program (CHRP 523790-18).

REFERENCES

1. Regev-Yochay G, Smollan G, Tal I, Pinas Zade N, Haviv Y, Nudelman V, Gal-Mor O, Jaber H, Zimlichman E, Keller N, Rahav G. 2018. Sink traps as the source of transmission of OXA-48-producing *Serratia marcescens* in an intensive care unit. Infect Control Hosp Epidemiol 39:1307–1315. https://doi.org/10.1017/ice.2018.235.
2. Bédard E, Laferrière C, Charron D, Lalancette C, Renaud C, Desmarais N, Déziel E, Prévost M. 2015. Post-outbreak investigation of *Pseudomonas aeruginosa* faucet contamination by quantitative polymerase chain reaction and environmental factors affecting positivity. Infect Control Hosp Epidemiol 36:1337–1343. https://doi.org/10.1017/ice.2015.168.
3. Lalancette C, Charron D, Laferrière C, Dolcé P, Déziel E, Prévost M, Bédard E. 2017. Hospital drains as reservoirs of *Pseudomonas aeruginosa*: multiple-locus variable-number of tandem repeats analysis genotypes recovered from faucets, sink surfaces and patients. Pathogens 6:36. https://doi.org/10.3390/pathogens6030036.
4. Inouye M, Conway TC, Zobel J, Holt KE. 2012. Short read sequence typing (SRST): multi-locus sequence types from short reads. BMC Genomics 13:338. https://doi.org/10.1186/1471-2164-13-338.
5. Boers SA, van der Reijden WA, Jansen R. 2012. High-throughput multilocus sequence typing: bringing molecular typing to the next level. PLoS One 7:e39630. https://doi.org/10.1371/journal.pone.0039630.
6. Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijl JM, Laurent F, Grundmann H, Friedrich AW, on behalf of the ESCMID Study Group of Epidemiological Markers (ESGEM). 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill 18:20380. https://doi.org/10.2807/ese.18.04.20380-en.
7. Basset P, Blanc DS. 2014. Fast and simple epidemiological typing of *Pseudomonas aeruginosa* using the double-locus sequence typing (DLST) method. Eur J Clin Microbiol Infect Dis 6:927–932. https://doi.org/10.1007/s10096-013-2028-0.
8. de Been M, Pinholt M, Top J, Bletz S, Mellmann A, van Schaik W, Brouwer E, Rogers M, Kraat Y, Bonten M, Corander J, Westh H, Harmsen D, Willems RJL. 2015. Core genome multilocus sequence typing scheme for high-resolution typing of *Enterococcus faecium*. J Clin Microbiol 53:3788–3797. https://doi.org/10.1128/JCM.01946-15.
9. Chen Y, Frazzitta AE, Litvintseva AP, Fang C, Mitchell TG, Springer DJ, Ding Y, Yuan G, Perfect JR. 2015. Next generation multilocus sequence typing (NGMLST) and the analytical software program MLSTEZ enable efficient, cost-effective, high-throughput, multilocus sequencing typing. Fungal Genet Biol 75:64–71. https://doi.org/10.1016/j.fgb.2015.01.005.
10. Tewolde R, Dallman T, Schaefer U, Sheppard CL, Ashton P, Pichon B, Ellington M, Swift C, Green J, Underwood A. 2016. MOST: a modified MLST typing tool based on short read sequencing. PeerJ 4:e2308. https://doi.org/10.7717/peerj.2308.
11. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95:3140–3145. https://doi.org/10.1073/pnas.95.6.3140.
12. Maiden MCJ, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND. 2013. MLST revisited: the gene-by-gene approach to

bacterial genomics. Nat Rev Microbiol 11:728–736. https://doi.org/10.1038/nrmicro3093.
13. Bleidorn C, Gerth M. 2018. A critical re-evaluation of multilocus sequence typing (MLST) efforts in *Wolbachia*. FEMS Microbiol Ecol 94 https://doi.org/10.1093/femsec/fix163.
14. Castillo-Ramírez S, Graña-Miraglia L. 2019. Inaccurate multilocus sequence typing of *Acinetobacter baumannii*. Emerg Infect Dis 25:186–187. https://doi.org/10.3201/eid2501.180374.
15. Grimont PAD, Grimont F. 1978. The genus *Serratia*. Annu Rev Microbiol 32:221–248. https://doi.org/10.1146/annurev.mi.32.100178.001253.
16. Hejazi A, Falkiner FR. 1997. *Serratia marcescens*. J Med Microbiol 46:903–912. https://doi.org/10.1099/00222615-46-11-903.
17. Villari P, Crispino M, Salvadori A, Scarcella A. 2001. Molecular epidemiology of an outbreak of *Serratia marcescens* in a neonatal intensive care unit. Infect Control Hosp Epidemiol 22:630–634. https://doi.org/10.1086/501834.
18. Assadian O, Berger A, Aspöck C, Mustafa S, Kohlhauser C, Hirschl AM. 2002. Nosocomial outbreak of *Serratia marcescens* in a neonatal intensive care unit. Infect Control Hosp Epidemiol 23:457–461. https://doi.org/10.1086/502085.
19. Milisavljevic V, Wu F, Larson E, Rubenstein D, Ross B, Drusin LM, Della-Latta P, Saiman L. 2004. Molecular epidemiology of *Serratia marcescens* outbreaks in two neonatal intensive care units. Infect Control Hosp Epidemiol 25:719–722. https://doi.org/10.1086/502466.
20. Maragakis LL, Winkler A, Tucker MG, Cosgrove SE, Ross T, Lawson E, Carroll KC, Perl TM. 2008. Outbreak of multidrug-resistant *Serratia marcescens* infection in a neonatal intensive care unit. Infect Control Hosp Epidemiol 29:418–423. https://doi.org/10.1086/587969.
21. Zingg W, Soulake I, Baud D, Huttner B, Pfister R, Renzi G, Pittet D, Schrenzel J, Francois P. 2017. Management and investigation of a *Serratia marcescens* outbreak in a neonatal unit in Switzerland – the role of hand hygiene and whole genome sequencing. Antimicrob Resist Infect Control 6:125. https://doi.org/10.1186/s13756-017-0285-x.
22. Åttman E, Korhonen P, Tammela O, Vuento R, Aittoniemi J, Syrjänen J, Mattila E, Österblad M, Huttunen R. 2018. A *Serratia marcescens* outbreak in a neonatal intensive care unit was successfully managed by rapid hospital hygiene interventions and screening. Acta Paediatr 107:425–429. https://doi.org/10.1111/apa.14132.
23. Martineau C, Li X, Lalancette C, Perreault T, Fournier E, Tremblay J, Gonzales M, Yergeau É, Quach C. 2018. *Serratia marcescens* outbreak in a neonatal intensive care unit: new insights from next-generation sequencing applications. J Clin Microbiol 56:235.
24. Moles L, Gómez M, Moroder E, Jiménez E, Escuder D, Bustos G, Melgar A, Villa J, del Campo R, Chaves F, Rodríguez JM. 2019. *Serratia marcescens* colonization in preterm neonates during their neonatal intensive care unit stay. Antimicrob Resist Infect Control 8:135. https://doi.org/10.1186/s13756-019-0584-5.
25. Cristina ML, Sartini M, Spagnolo AM. 2019. *Serratia marcescens* infections in neonatal intensive care units (NICUs). Int J Environ Res Public Health 16:610. https://doi.org/10.3390/ijerph16040610.
26. Varsha G, Shiwani S, Kritika P, Poonam G, Deepak A, Jagdish C. 2021. *Serratia* no longer an opportunistic uncommon pathogen – case series & review of literature. Infect Disord Drug Targets 21:1–1.
27. Johnson J, Quach C. 2017. Outbreaks in the neonatal ICU: a review of the literature. Curr Opin Infect Dis 30:395–403. https://doi.org/10.1097/QCO.0000000000000383.
28. Friedman ND, Kotsanas D, Brett J, Billah B, Korman TM. 2008. Investigation of an outbreak of *Serratia marcescens* in a neonatal unit via a case-control study and molecular typing. Am J Infect Control 36:22–28. https://doi.org/10.1016/j.ajic.2006.12.012.
29. Rossen JWA, Dombrecht J, Vanfleteren D, Bruyne KD, Belkum A van, Rosema S, Lokate M, Bathoorn E, Reuter S, Grundmann H, Ertel J, Higgins PG, Seifert H. 2019. Epidemiological typing of *Serratia marcescens* isolates by whole-genome multilocus sequence typing. J Clin Microbiol 57:e01652. https://doi.org/10.1128/JCM.01652-18.
30. Abreo E, Altier N. 2019. Pangenome of *Serratia marcescens* strains from nosocomial and environmental origins reveals different populations and the links between them. Sci Rep 9:46. https://doi.org/10.1038/s41598-018-37118-0.
31. Liu Y-Y, Chiou C-S, Chen C-C. 2016. PGAdb-builder: a web service tool for creating pan-genome allele database for molecular fine typing. Sci Rep 6:36213. https://doi.org/10.1038/srep36213.
32. Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser (Oxf): 95–98.
33. Figueras MJ, Beaz-Hidalgo R, Hossain MJ, Liles MR. 2014. Taxonomic affiliation of new genomes should be verified using average nucleotide identity and multilocus phylogenetic analysis. Genome Announc 2:e00927. https://doi.org/10.1128/genomeA.00927-14.
34. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal Methods 8:12–24. https://doi.org/10.1039/C5AY02550H.
35. R Core Team. 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
36. Suzuki R, Shimodaira H. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22:1540–1542. https://doi.org/10.1093/bioinformatics/btl117.
37. Galili T. 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics 31:3718–3720. https://doi.org/10.1093/bioinformatics/btv428.
38. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2020. vegan: community ecology package.
39. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the tidyverse. JOSS 4:1686. https://doi.org/10.21105/joss.01686.
40. Borcard D, Gillet F, Legendre P. 2018. Numerical ecology with R. Springer, New York City, New York. https://doi.org/10.1007/978-3-319-71404-2.
41. Gu Z, Gu L, Eils R, Schlesner M, Brors B. 2014. circlize implements and enhances circular visualization in R. Bioinformatics 30:2811–2812. https://doi.org/10.1093/bioinformatics/btu393.
42. Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32:2847–2849. https://doi.org/10.1093/bioinformatics/btw313.
43. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134. https://doi.org/10.1186/1471-2105-13-134.
44. Berkowitz DM, Lee WS. 1973. A selective medium for isolation and identification of *Serratia marcescens*. Abstr. Annu Meet Am Soc Microbiol 1973:105.
45. Durand A-A, Bergeron A, Constant P, Buffet J-P, Déziel E, Guertin C. 2015. Surveying the endomicrobiome and ectomicrobiome of bark beetles: the case of *Dendroctonus simplex*. Sci Rep 5:17190. https://doi.org/10.1038/srep17190.
46. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. 1. Embnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.
47. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869.
48. Morgan M, Lawrence M, Anders S. 2021. ShortRead: FASTQ input and manipulation. Bioconductor version: release (3.12).
49. Pagès H, Aboyoun P, Gentleman R, DebRoy S. 2021. Biostrings: efficient manipulation of biological strings. Bioconductor version: release (3.12).
50. Nascimento M, Sousa A, Ramirez M, Francisco AP, Carriço JA, Vaz C. 2017. PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. Bioinformatics 33:128–129. https://doi.org/10.1093/bioinformatics/btw582.
51. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
52. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes de novo assembler. Curr Protoc Bioinformatics 70:e102. https://doi.org/10.1002/cpbi.102.
53. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352. https://doi.org/10.1093/bioinformatics/btv383.
54. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics 25:1968–1969. https://doi.org/10.1093/bioinformatics/btp347.

55. Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106:19126–19131. https://doi.org/10.1073/pnas.0906412106.
56. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12. https://doi.org/10.1186/gb-2004-5-2-r12.
57. Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics 172:2665–2681. https://doi.org/10.1534/genetics.105.048975.
58. Diggle SP, Whiteley M. 2020. Microbe Profile: *Pseudomonas aeruginosa*: opportunistic pathogen and lab rat. Microbiology (Reading) 166:30–33. https://doi.org/10.1099/mic.0.000860.
59. Freschi L, Vincent AT, Jeukens J, Emond-Rheault J-G, Kukavica-Ibrulj I, Dupont M-J, Charette SJ, Boyle B, Levesque RC. 2019. The *Pseudomonas aeruginosa* pan-genome provides new insights on its population structure, horizontal gene transfer, and pathogenicity. Genome Biol Evol 11:109–120. https://doi.org/10.1093/gbe/evy259.
60. Iguchi A, Nagaya Y, Pradel E, Ooka T, Ogura Y, Katsura K, Kurokawa K, Oshima K, Hattori M, Parkhill J, Sebaihia M, Coulthurst SJ, Gotoh N, Thomson NR, Ewbank JJ, Hayashi T. 2014. Genome evolution and plasticity of *Serratia marcescens*, an important multidrug-resistant nosocomial pathogen. Genome Biol Evol 6:2096–2110. https://doi.org/10.1093/gbe/evu160.
61. Ewing B, Green P. 1998. Base-calling of automated sequencer traces using Phred. II. Error Probabilities. Genome Res 8:186–194. https://doi.org/10.1101/gr.8.3.186.
62. Franco LC, Tanner W, Ganim C, Davy T, Edwards J, Donlan R. 2020. A microbiological survey of handwashing sinks in the hospital built environment reveals differences in patient room and healthcare personnel sinks. Sci Rep 10:8234. https://doi.org/10.1038/s41598-020-65052-7.
63. Wingender J. 2011. Hygienically Relevant Microorganisms in Biofilms of Man-Made Water Systems, p 189–238. *In* Flemming H-C, Wingender J, Szewzyk U (ed.), Biofilm highlights. Springer, New York City, New York.
64. Gaiarsa S, Batisti Biffignandi G, Esposito EP, Castelli M, Jolley KA, Brisse S, Sassera D, Zarrilli R. 2019. Comparative analysis of the two *Acinetobacter baumannii* multilocus sequence typing (MLST) schemes. Front Microbiol 10:930. https://doi.org/10.3389/fmicb.2019.00930.
65. Graña-Miraglia L, Arreguín-Pérez C, López-Leal G, Muñoz A, Pérez-Oseguera A, Miranda-Miranda E, Cossío-Bayúgar R, Castillo-Ramírez S. 2018. Phylogenomics picks out the par excellence markers for species phylogeny in the genus *Staphylococcus*. PeerJ 6:e5839. https://doi.org/10.7717/peerj.5839.
66. Gupta RS, Son J, Oren A. 2019. A phylogenomic and molecular markers based taxonomic framework for members of the order Entomoplasmatales: proposal for an emended order Mycoplasmatales containing the family *Spiroplasmataceae* and emended family *Mycoplasmataceae* comprised of six genera. Antonie Van Leeuwenhoek 112:561–588. https://doi.org/10.1007/s10482-018-1188-4.
67. Steinegger M, Salzberg SL. 2020. Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. Genome Biol 21:115. https://doi.org/10.1186/s13059-020-02023-1.
68. Magalhães B, Valot B, Abdelbary MMH, Prod'hom G, Greub G, Senn L, Blanc DS. 2020. Combining standard molecular typing and whole genome sequencing to investigate *Pseudomonas aeruginosa* epidemiology in intensive care units. Front Public Health 8:3. https://doi.org/10.3389/fpubh.2020.00003.
69. Rahme LG, Stevens EJ, Wolfort SF, Shao J, Tompkins RG, Ausubel FM. 1995. Common virulence factors for bacterial pathogenicity in plants and animals. Science 268:1899–1902. https://doi.org/10.1126/science.7604262.