Supplementary Data for:

Whole-genome sequencing identifies emergence of a quinolone resistance mutation in a case of *Stenotrophomonas maltophilia* bacteremia

Supplementary Methods

Genome sequencing

Sequencing was performed to a depth of coverage of >150x per genome using the P4-C2 sequencing enzyme and chemistry at the manufacturer's specifications on the PacBio RSII platform (Pacific Biosciences, Menlo Park, CA). For ISMMS2 and ISMMS2R, Sanger sequencing was additionally performed on six PCR-amplified regions encompassing the one single nucleotide variant (SNV) and five one-base indels that differentiated the two PacBio assemblies. Conventional PCR amplification was performed with Choice-Taq Blue (Denville Scientific) and included an initial denaturation step of 180s at 95°C, 30 cycles of denaturation, annealing, and extension at 95°C/30s, 60°C/30s, and 72°C/30s respectively, and a final extension step of 300s at 72°C. Primer sequences are as follows: for the SNV, 5'-CAAGGTGCTGACCGAAATGC-3' forward and 5'-ACACGCCATCCTTCACGTAG-3' reverse; and for the five indels, 5'-GCATGGAAGTACCACTGGGT-3' forward + 5'-TTGGAGGGGTGGTAAAACGG-3' reverse, 5'-TGGCCAACCCCTTCTATGTC-3' forward + 5'-CCATGGCCACAGCAAAATGG-3' reverse, 5'-CTGCCTTCGGTCACTTCGT-3' forward + 5'-TGGAAGTCTCGCTGGAAGGT-3' reverse, 5'-GCCCTCTACACCGTCTTTCC-3' forward + 5'-GAACTACCGGACGGCTTTGA-3' reverse, and 5'-AACTTCTTCGTGTCGGTCCC-3' forward + 5'-AGAACTACCGGACGGCTTTG-3' reverse. Sequences on both strands of the amplified products were determined at an external sequencing facility (Macrogen Inc., Rockville, MD) using the standard Sanger dideoxy-terminator method and the same primers.

Sequence analysis

Sequencing data was processed and assembled *de novo* using PacBio's Hierarchical Genome Assembly Process (HGAP, version 3) (1) in the SMRTanalysis toolkit (version 2.3.0) using standard pre-assembly pipeline parameters. Custom scripts were used to circularize the draft assemblies and orient them similarly to reference assemblies K279a, R551-3, D457, and JV3 using the *gyrB* locus as a landmark; these scripts are available at https://github.com/powerpak/pathogendb-pipeline/releases/tag/steno_v1.0 (doi:10.5281/zenodo.17295) within the files `scripts/circularizeContigs.pl` and `scripts/fasta-orient-to-landmark.pl`. To eliminate overhanging sequence at the end of contigs and to increase accuracy, raw reads were re-mapped to the circularized assemblies using Blasr and the final consensus was re-called using Quiver. Initial annotations were created using the RAST server (2) with specific annotation of *sme* genes derived from BLAST queries.

Depth of coverage reported in Table 1 was calculated by SMRTanalysis (version 2.3.0) during re-mapping of reads to the circularized draft assembly.

Accession numbers

Sequences and annotations for reference assemblies of clinical *S. maltophilia* isolates K279a, R551-3, D457, and JV3 were obtained from GenBank/RefSeq at accession numbers AM743169.1, NC_011071.1, NC_017671.1, and NC_015947.1, respectively. These represent

the entirety of assemblies for *S. maltophilia* found in NCBI Assembly with an Assembly Level of "Complete Genome" (http://www.ncbi.nlm.nih.gov/assembly/organism/40324/all/) at the time of submission. K279a and D457 were isolated from human infections, while R551-3 and JV3 were isolated from plants. Previously published sequences for the quinolone-resistance determining region (QRDR) of the *gyrA*, *gyrB*, *parC* and *parE* genes in *S. maltophilia* (3) were obtained from EMBL/European Nucleotide Archive.

Complete genome sequences for ISMMS2, ISMMS2R, and ISMMS3 were deposited in GenBank at accession numbers CP011305, CP011306, and CP011010, respectively. Deposited sequences for ISMMS2 and ISMMS2R incorporate the Sanger corrected regions described above. Sequences for ISMMS4, ISMMS5, ISMMS6, and ISMMS7 were deposited as Whole Genome Shotgun projects at DDBJ/EMBL/GenBank under the accessions JZIU00000000, JZIV00000000, JZIW00000000, and JZTX00000000, respectively, with the versions described in this paper at JZIU01000000, JZIV01000000, JZIW01000000, and JZTX01000000, respectively.

Comparative genomic analysis

Pairwise comparison between strains was performed with the MUMmer 3.23 package (4), firstly using `nucmer` for pairwise genome alignment. The resulting `nucmer` alignments were filtered for quality and uniqueness via the `delta-filter` tool (using the `-1` flag to identify top alignments between the reference and query intervals). To estimate phylogenetic tree distances, high-quality SNP and indel calls were assigned via the `show-SNPs` tool using the `-C` flag to only report SNPs in regions with unambiguous mappings. For ISMMS2 and ISMMS2R, `show-SNPs` was also used without the `-C` flag to verify that no additional SNPs or indels were in ambiguously mapped regions.

Mugsy 2.2 (5) was used to perform multiple sequence alignment of the whole genome sequences in order to find local collinear blocks (LCBs) of conserved sequence. These aligned blocks were used to establish a core genome (of 3.01 Mbp) across all isolates, from which a phylogenetic tree was constructed using RAxML-8.0.2 (6), employing the GTRGAMMA substitution model and performing 20 runs. Whole genome alignments for visualization of recombination events was performed with Mauve 2.4.0 (7), using the progressiveMauve algorithm (8) with a minimum seed weight of 21, seed families enabled, and all other parameters at defaults. Clustal Omega (9) was used for multiple sequence alignment of putative amino acid sequences, which were then rendered with ESPript version 3.0 (10).

Epigenetic motif analysis

For each isolate, initial DNA modification motifs were first predicted by a *de novo* motif discovery pipeline in SMRTportal (RS_Modifications_Motif_Analysis.1). The pipeline searches for kinetic variations in DNA polymerization events recorded during sequencing that correlate with modifications in the template, with different modifications creating distinct kinetic profiles (11, 12). At the coverage depths reported in Table 1, the probability (power) of detecting a modification event at a site at the 0.1 significance threshold, if it is truly modified, exceeds 99.99% (11). Raw predictions, which often have incorrectly- or over-called motifs, were further refined by a re-analysis of the raw data using a single molecule level characterization method (J

Beaulaurier, X-S Zhang, S Zhu et al. 2015, manuscript in preparation). Conceptually, this method was used to check the single molecule level methylation status of each putative motif and its neighboring (more or less specific) motifs and determine the real motif.

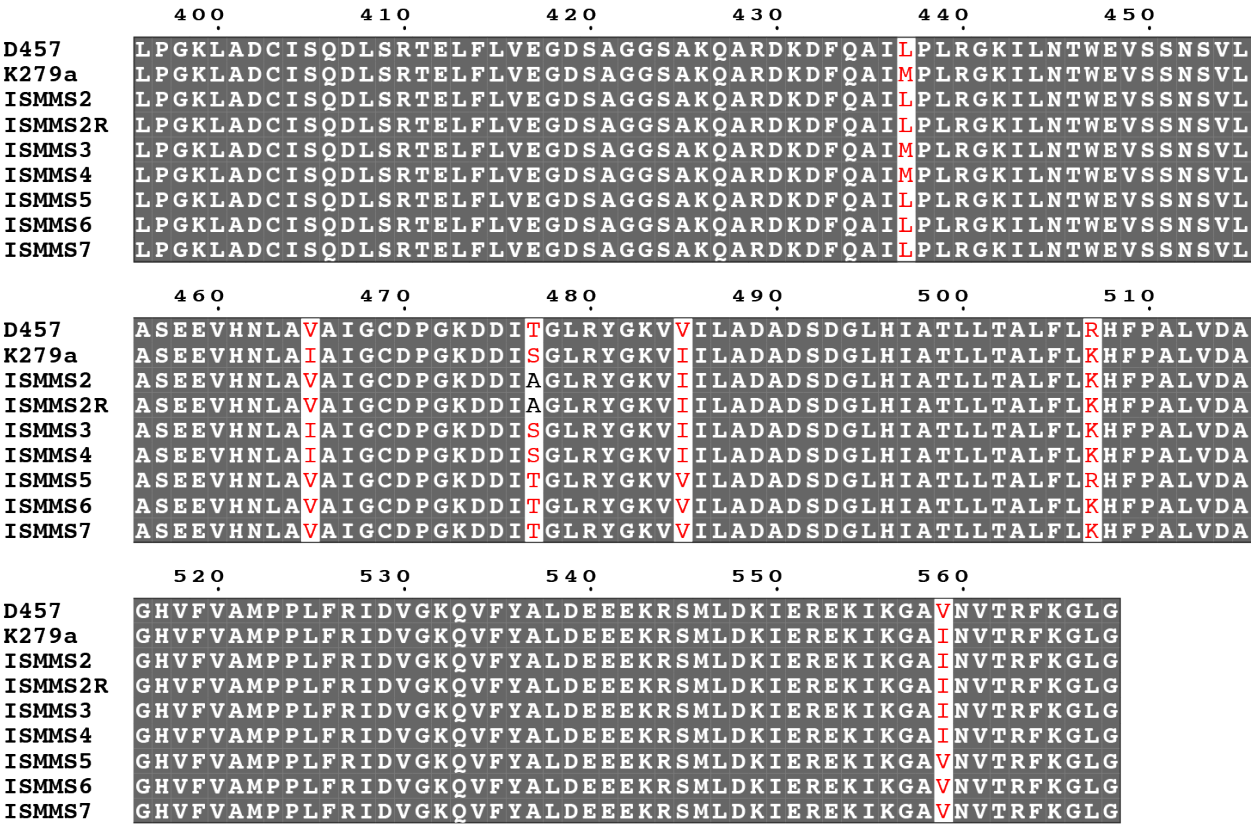
Supplementary Tables

Supplementary Table 1. Diverse epigenetic motifs, representing putative target sequences for each strain's DNA methyltransferase enzymes, discovered for clinical isolates of *S. maltophilia*. Isolates are named as in Table 1. The underlined A's correspond to putative 6-methyladenine residues, which was the only modification type found in this study.

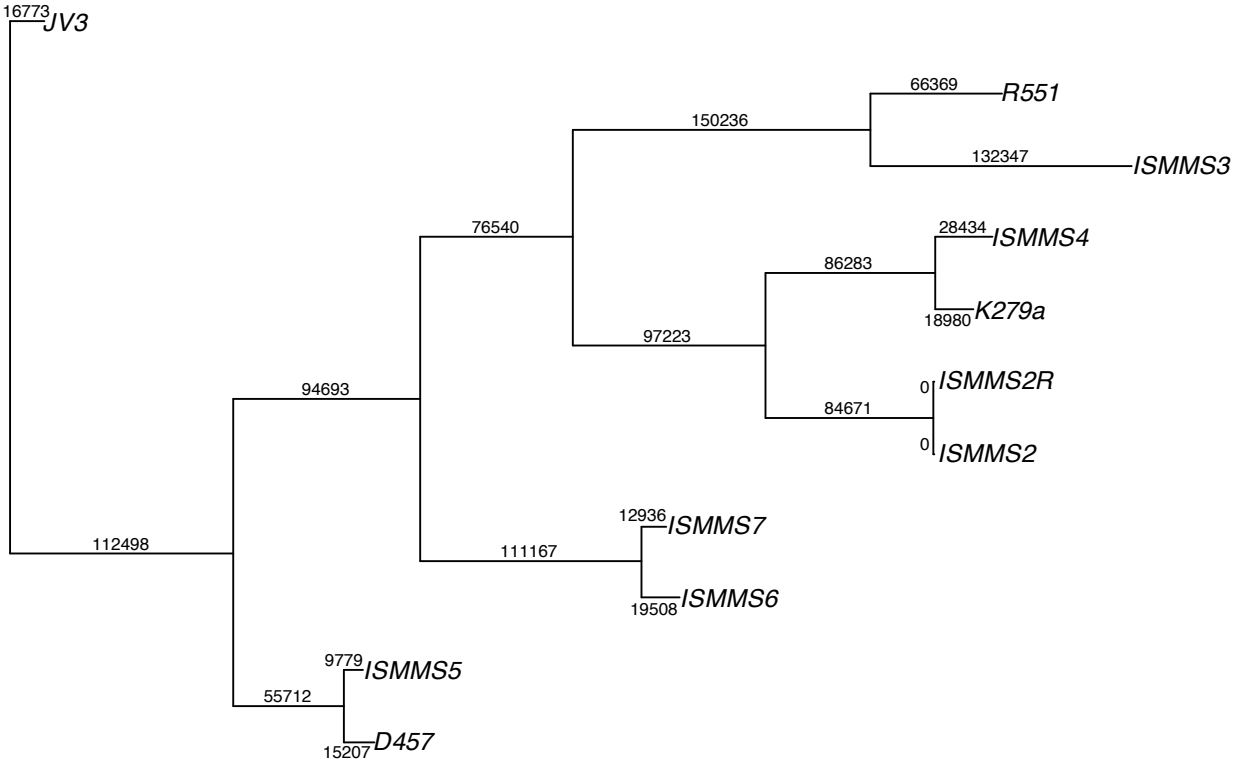
Isolate name	Epigenetic motifs
ISMMS2	AGTACT
ISMMS2R	AGTACT
ISMMS3	None
ISMMS4	CAGAG
ISMMS5	CTGGAC, CACANAG
ISMMS6	CAACAC, CTGATG, CAACGAC
ISMMS7	CAGAG

Supplementary Figures

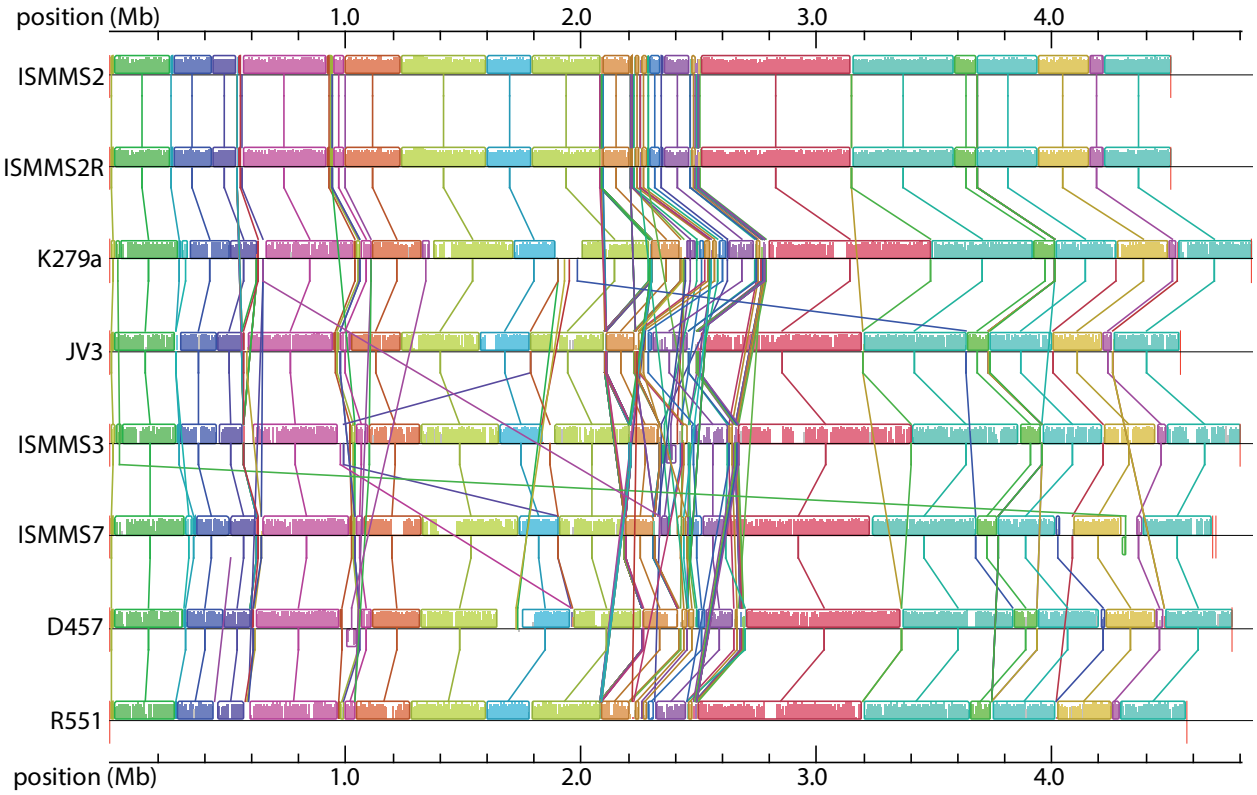
Supplementary Figure 1. Amino-acid sequence alignment for the quinolone-resistance determining region (QRDR) of the *parE* gene for seven *S. maltophilia* clinical isolates (ISMMS2 through 7 and ISMMS2R) and two reference assemblies of clinical isolates obtained from GenBank.



Supplementary Figure 2. Phylogeny of seven *S. maltophilia* clinical isolates (ISMMS2 through 7 and ISMMS2R) and four reference assemblies obtained from GenBank. Trees were constructed by inferring ancestral states using RAxML-8.0.2 (6); branch lengths correspond to single-nucleotide polymorphism (SNP) distances from branch points, and are drawn using R version 3.0.3 and the APE library version 3.1-1 (13). The core genome did not contain the *smeT* locus; therefore, the SNV differentiating ISSMS2 and ISMMSR is not observed in this tree.



Supplementary Figure 3. Genome-scale comparison of four fully assembled *S. maltophilia* clinical isolates and four reference assemblies obtained from GenBank. Mauve 2.4.0 (7, 8) was used to plot locally collinear blocks (LCBs; conserved segments that appear to be internally free from genome rearrangements) as colored rectangles, with gaps representing non-homologous regions. Vertical bars inside each LCB rectangle show the average level of conservation at that region of the genomic sequence. Colored lines connect homologous LCBs among the genomes, and LCBs plotted below the centerline are in the reverse complement orientation relative to the ISMMS2 sequence. At top, sequences for the isolates from before and after development of quinolone resistance (ISSMS2 and ISSMS2R) in the case patient have identical structures.



References

1. **Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J**. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods **10**:563–9.
2. **Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R**. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res **42**:D206–14.
3. **Valdezate S, Vindel A, Echeita A, Baquero F, Cantón R**. 2002. Topoisomerase II and IV quinolone resistance-determining regions in Stenotrophomonas maltophilia clinical isolates with different levels of quinolone susceptibility. Antimicrob Agents Chemother **46**:665–671.
4. **Delcher AL, Salzberg SL, Phillippy AM**. 2003. Using MUMmer to identify similar regions in large sequence sets. Curr Protoc Bioinformatics **Chapter 10**:Unit 10.3.
5. **Angiuoli S V, Salzberg SL**. 2011. Mugsy: fast multiple alignment of closely related whole genomes. Bioinformatics **27**:334–42.
6. **Stamatakis A**. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**:1312–3.
7. **Darling ACE, Mau B, Blattner FR, Perna NT**. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14**:1394–403.
8. **Darling AE, Mau B, Perna NT**. 2010. Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. PLoS One **5**.
9. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG**. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol **7**.
10. **Robert X, Gouet P**. 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res **42**:320–324.
11. **Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, Feng Z, Losic B, Mahajan MC, Jabado OJ, Deikus G, Clark TA, Luong K, Murray IA, Davis BM, Keren-Paz A, Chess A, Roberts RJ, Korlach J, Turner SW, Kumar V, Waldor MK, Schadt EE**. 2012. Genome-wide mapping of methylated adenine residues in pathogenic Escherichia coli using single-molecule real-time sequencing. Nat Biotechnol **30**:1232–1239.
12. **Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, Fomenkov A, Roberts RJ, Korlach J**. 2012. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. Nucleic Acids Res **40**.
13. **Paradis E, Claude J, Strimmer K**. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics **20**:289–290.