June 21, 2019
FRIDAY – EEB-482

Using Whole-Genome Sequencing to Examine the Taxonomy of *Yersinia*

ATCC®
Credible leads to Incredible™

Alia Abdelhadi, MPA, Valeria Barisic, MS, Andrew Frank, MS, Marco A. Riojas, PhD, Manzour H. Hazbón, PhD
ATCC, Manassas, Virginia 20110

Abstract

Comprising 19 species/subspecies, *Yersinia* are Gram-negative coccobacilli implicated in a variety of human and zoonotic diseases. Several species of *Yersinia* share high genomic similarity with each other, and the ability to discern these species is vital—particularly for *Y. pestis*, the causative agent of plague, whose genomic composition is closely related to *Y. pseudotuberculosis*. In this study, we aim to revisit the taxonomy of *Yersinia* through whole-genome sequencing (WGS) of the type strains and confirm their taxonomic assignment. Whole-genome distances and phylogenomic analyses confirmed the current taxonomy of 17 species/subspecies. Unsurprisingly, four species that showed a greater degree of relatedness are *Y. pestis*, *Y. pseudotuberculosis*, *Y. similis,* and *Y. wautersii*, which constitute the *Y. pseudotuberculosis* complex. Recent research using multilocus sequence analysis identified *Y. wautersii* as a novel member species of the *Y. pseudotuberculosis* complex. However, based on whole-genome distances, our data shows enough similarity between *Y. wautersii* and *Y. pseudotuberculosis* to be considered the same species but different subspecies. Phylogenomic trees, which place *Y. wautersii* and *Y. pseudotuberculosis* on the same branch, further substantiate this data. We propose the unification of *Y. pseudotuberculosis* and *Y. wautersii* as *Y. pseudotuberculosis* subsp. *pseudotuberculosis* and *Y. pseudotuberculosis* subsp. *wautersii,* respectively.

Introduction

Each species is represented by a type strain and a description of that strain. The type strain are usually the first strain identified, but it is not necessarily the most typical or representative of the species. The type strain is essentially the “definition” of a species. A strain that shares enough of the essential characteristics of a type strain is said to be within the circumscription of that species/type strain, hence it is classified as belonging to the same species of the type strain.

The comparison to type strains is done via phenotypic and genotypic characteristics. Phenotypic comparisons include cell and colony morphology, staining properties, biochemical tests, etc. Genotypic comparisons include DNA sequence analysis of 16S rDNA genes, *hsp65,* and/or *rpoB*. The combination of these characteristics can lead to a good bacterial identification, but in some instances this can fail or yield vague or inaccurate conclusions. Further, some mutations can alter phenotypic features, which can lead to a mistaken identification. The use of the WGS provides a more reliable tool to compare and identify strains.

Y. pestis and *Y. pseudotuberculosis* are known to belong to the same species. However, because of their major differences in virulence to humans it was decided to maintain them as separate species.[1-4] WGS has not been used to reassess the taxonomy of the genus *Yersinia* (it has been done for only some of the species[5]). For this reason, we obtained the WGS from the type strains of each of the 19 species/subspecies that comprise the genus and analyzed their phylogenetic distribution and relatedness.

Materials and Methods

Bacterial Strains and DNA Extraction. The type strains for each species and subspecies of *Yersinia* were obtained from the American Type Culture Collection (ATCC), the Leibnitz Institute Deutsche Sammlung von Mikroorganismen and Zellkulturen (DSMZ), the Riken BRC/Japan Collection of Microorganisms (JCM), and the Belgian Co-ordinated Collections of Microorganisms (BCCM). Strains were grown as recommended by the manufacturers, and gDNA was extracted using the QIAGEN® MagAttract® High-Molecular Weight (HMW) system. Additionally, existing genomes from GenBank were also used in the analysis. Together, the genomes from the strains sequenced and those from GenBank compose the main dataset for the genomic analysis.

Whole-Genome Sequencing (WGS). DNA was prepared using the Nextera® XT Library Preparation Kit (Illumina®) and sequenced using Illumina MiSeq® v3 flow cells (2×300). Resultant paired-end reads underwent contamination detection using the One Codex microbial genomics read-based identification algorithm. Read pairs were adapter trimmed and quality filtered and were then used for *de novo* genome assembly using SPAdes 3.12.0.

Calculation of Genomic Distances. For independent corroboration of the results, two algorithmic approaches were used. Genomic distance based on digital DNA-DNA hybridization (dDDH) was calculated with the Genome-to-Genome Distance Calculator (GGDC) v2.1 using the recommended Formula 2.[6,7] Average nucleotide identity (ANI) was calculated using OrthoANIu.[7] The species delineation thresholds used were ≥70% via dDDH and ≥96% via ANI.[9,10] A dDDH distance of ≥70-79.9% was considered to represent different subspecies of the same species, whereas ≥80% was considered to represent the same subspecies of the same species (or no subspecies in the case of species without multiple subspecies).[8] No subspecies delineation threshold based on ANI values currently exists. The calculated dDDH values were used as the basis for a phylogenetic tree as described previously.[11]

Phylogenetic tree construction. To calculate phylogenetic trees by using constituent genomes, we calculated maximum likelihood species trees via UBCG. UBCG automatically extracts 92 pre-defined bacterial core genes from all provided genome assemblies, creates a multiple sequence alignment (MSA) for each gene, and concatenated all MSAs into a single supermatrix. UBCG was run using default settings. The resultant supermatrix was used as input for RAxML maximum likelihood species tree estimation. RAxML was run using the GTRGAMMA model of nucleotide evolution with the option to calculate 100 bootstrap trees, which were used to calculate branch support values. Each species tree was rooted at the outgroup, and branches with support values less than 70 were collapsed to show only statistically supported branches. Comprehensive Genome Analysis (CGA) was performed using the Pathosystems Resource Integration Center (PATRIC).

Results and Conclusions

- As expected, the type strains of *Y. pestis* and *Y. pseudotuberculosis* show high similarity between them, indicating that genomically they both represent a single species. However, the Judicial Commission of the International Committee on Systematic Bacteriology decided to maintain both species separately.[1-4] This decision has been widely accepted by the scientific community because *Y. pestis* has defined markers that make its identification clear and they correlate with marked increased virulence.
- *Y. wautersii* was defined as a new species in 2014.[11] However, in this study the WGS from *Y. wautersii* show similarity values to *Y. pestis* and *Y. pseudotuberculosis* of 78% and 97.5% for dDDH and ANI respectively. This suggests not only that they are the same species but they are distinct enough to be a separate subspecies. For this reason, *Y. wautersii* should be reclassified as *Y. pseudotuberculosis subsp. wautersii*, resulting in the creation of *Y. pseudotuberculosis subsp. pseudotuberculosis*.
- The species belonging to *Yersinia pseudotuberculosis* complex seem to have evolved from a common ancestor. The identification of subspecies suggests that this evolution is still ongoing.
- The overall taxonomy of the *Yersinia* genus was corroborated using WGS. NGS provides a reliable and reproducible tool to confirm bacterial identification and taxonomic classification down to the subspecies level.

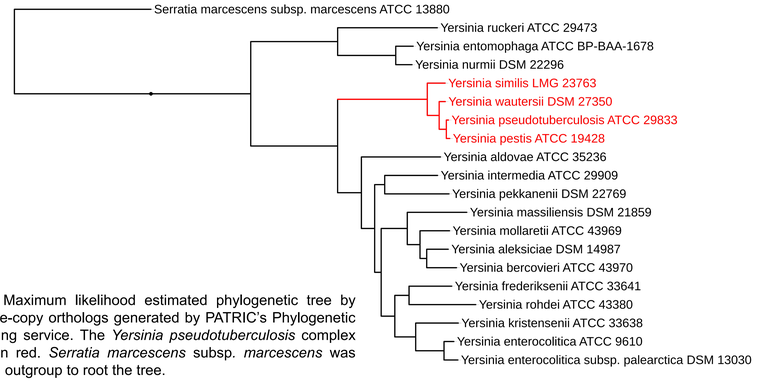

Figure 1. Maximum likelihood estimated phylogenetic tree by using single-copy orthologs generated by PATRIC’s Phylogenetic Tree Building service. The *Yersinia pseudotuberculosis* complex is shown in red. *Serratia marcescens* subsp. *marcescens* was used as an outgroup to root the tree.

Species/Strain	dDDH / ANI	ATCC 3523^T	DSM 14987^T	ATCC 43970^T	ATCC 9610^T	DSM 13030^T	ATCC BAA-1678^T	ATCC 33641^T	ATCC 29909^T	ATCC 33638^T	DSM 21859^T	ATCC 43969^T	DSM 22296^T	DSM 22769^T	ATCC 43380^T	ATCC 29473^T	LMG 23763^T	ATCC 19428^T	ATCC 29833^T	DSM 27350^T	ATCC 13880^T
Y. aldovae	ATCC 3523^T	100	26.2	26.3	28.2	28.5	22.1	25.9	27.2	27.1	24.8	26.1	22.1	27.1	25.7	21.9	25.6	25.9	25.7	25.9	20.4
Y. aleksiciae	DSM 14987^T	82.63	100	42.6	27.2	26.9	22	27.2	27.9	27.2	28.8	38.9	22	28.9	26.7	22.2	25.8	26	25.7	25.8	20.5
Y. bercovieri	ATCC 43970^T	82.47	90.78	100	27.2	26.8	22	27.1	27.9	27.2	28.6	38.1	21.8	28.7	26.7	22	25.7	26.2	25.7	25.9	20.5
Y. enterocolitica subsp. *enterocolitica*	ATCC 9610^T	84.04	83.4	83.24	100	72.6	21.9	28.4	27.2	34.1	25.4	27	21.9	27.7	27.8	21.8	25.6	26	25.6	25.4	20.1
Y. enterocolitica subsp. *palearctica*	DSM 13030^T	84.06	83.25	83.15	96.72	100	22	28.2	27.1	33.6	21.5	26.9	21.8	27.5	27.6	21.7	25.5	26.3	25.6	25.6	20.1
Y. entomophaga	ATCC BAA-1678^T	77.95	77.64	77.81	77.58	77.51	100	21.5	22.1	21.8	21.6	22.1	59.3	22	22.1	25.6	22.4	23.3	22.5	22.6	20.5
Y. frederiksenii	ATCC 33641^T	82.29	83.46	83.48	84.19	84.14	77.39	100	26.8	28.2	25.3	27.1	21.4	27.6	29.5	22.1	25.5	25.9	25.7	25.6	19.9
Y. intermedia	ATCC 29909^T	83.36	83.94	83.81	83.52	83.18	77.71	83.14	100	27.1	25.8	27.6	22	29.9	26.5	22.5	25.8	26.1	25.9	26	20.3
Y. kristensenii	ATCC 33638^T	82.94	83.22	83.42	87.47	87.36	77.41	84.11	83.45	100	25.2	27.1	21.8	28.1	27.8	21.8	25.5	25.7	25.7	25.6	20.1
Y. massiliensis	DSM 21859^T	81.03	84.78	84.55	81.67	81.83	77.03	81.9	81.95	81.86	100	28.4	21.4	25.9	25.1	21.8	24.6	25	24.7	24.7	19.9
Y. mollaretii	ATCC 43969^T	82.51	89.47	89.16	83.35	83.34	77.67	83.47	83.66	83.43	82.9	100	21.8	28.7	26.7	22.1	25.5	25.8	25.6	25.7	20.2
Y. nurmii	DSM 22296^T	77.58	77.5	77.54	77.34	77.28	94.91	77.49	77.56	77.39	76.85	77.57	100	21.8	22	25.4	22.5	22.8	22.5	22.4	20.5
Y. pekkanenii	DSM 22769^T	83.34	84.34	84.37	83.72	83.6	77.7	83.77	85.07	83.92	82.16	84.38	77.78	100	26.9	22	26.7	26.5	26.2	26.5	19.9
Y. rohdei	ATCC 43380^T	81.86	82.94	82.93	83.63	83.48	77.38	85.06	82.73	83.56	81.41	82.9	77.47	82.96	100	22.4	25.1	25.5	25.2	25.1	20.3
Y. ruckeri	ATCC 29473^T	77.9	77.87	77.89	77.72	77.69	82.42	77.72	77.9	77.41	77.2	77.82	82.24	77.73	77.55	100	22.5	22.9	22.7	22.7	19.9
Y. similis	LMG 23763^T	81.04	81.25	81.26	81.19	80.92	77.21	81.31	81.58	81.01	80.25	81.2	77.32	81.88	80.53	77.56	100	57.9	57.8	59.3	20.1
Y. pestis	ATCC 19428^T	81.29	81.31	81.33	81.32	81.49	77.68	81.15	81.48	81.12	80.39	81.22	77.4	81.71	80.84	77.64	94.57	100	92.7	78	20.7
Y. pseudotuberculosis	ATCC 29833^T	81.16	81.39	81.26	81.09	80.98	77.5	81.29	81.51	80.98	80.24	81.29	77.39	81.55	80.58	77.66	94.46	99.08	100	78.1	20.3
Y. wautersii	DSM 27350^T	81.3	81.35	81.38	81.16	81.03	77.37	84.15	81.49	80.86	80.37	81.44	77.34	81.72	80.6	77.6	94.76	97.51	97.6	100	20.2
Serratia marcescens subsp. *marcescens*	ATCC 13880^T	74.69	75	75.27	74.39	74.37	75.59	74	74.35	74.42	74.21	75.11	75.37	74.7	74.18	75.18	74.09	74.38	74.3	74.23	100

dDDH

ANI

Table 1. Genomic distances between the *Yersinia* type strains examined in this work. The type strain of *Serratia marcescens* subsp. *marcescens* was used as an outgroup. dDDH values are shown above the self-comparison diagonal; ANI values are shown below the diagonal.

ANI	Interpretation	dDDH
98.0-100	Same species	80-100
96.5-97.9	Same species, different subspecies	70-80
<96.5	Different species	<70

Subspecies

References

1. Int J Syst Evol Microbiol 34: 268-269, 1984
2. Int J Syst Evol Microbiol 35: 540-540, 1985
3. Int J Syst Evol Microbiol 36: 357-358, 1986
4. Stand Genomic Sci 2(1): 117-34, 2010
5. Int J Syst Evol Microbiol 66(12): 5575-5599, 2016
6. BMC Bioinformatics 14: 60, 2013
7. Antonie van Leeuwenhoek 110(10): 1281-1286, 2017
8. Stand Genomic Sci 9: 2, 2014
9. Proc Natl Acad Sci U S A 106(45): 19126-31, 2009
10. Int J Syst Evol Microbiol 68(1): 324-332, 2018
11. Int J Syst Evol Microbiol 304: 452-463, 304

ATCC 10801 University Boulevard, Manassas, Virginia 20110-2209 **Phone:** 703.263.8277 **Email:** SalesRep@atcc.org **Web:** www.atcc.org