Untitled

Kimberly Hinz

2020/12/08

# Abstract

*Mixta* is a recently described genus within the bacterial family *Erwiniaceae* that is closely related to *Pantoea*, *Erwinia*, and *Tatumella*. Depending on the genes and method used for phylogenetic analysis, the genus with which *Mixta* shares the most recent ancestor differs. This study aimed to determine the cause behind these contentious results, and identified genes that may have resulted from horizontal gene transfer. Ten complete or partial genomes – two *Mixta* species, six species from other *Erwiniaceae* genera, and two outgroups – were retrieved from NCBI and annotated. Homologous genes were extracted yielding a dataset of 799 genes. The genes were aligned With the ClustalW algorithm and MEGA-CC was used for model testing and computing phylogenetic trees for each gene for both nucleotide and amino acid sequences. Distance matrices were then extracted from these trees. Nucleotide and amino acid identity analyses were also done in R programming language. *Pantoea* was the closest relative to the *Mixta* species in most of the analyses; however, results were not consistent across analyses. Some genes were also found to be more similar to other, non-*Pantoea* species. Diligence must be given to the selection of genes for phylogenetic analysis and to the method chosen to prevent xenologous signal from distorting the true relationships. Furthermore, future research should consider that different phylogenetic analyses may give different results.

# Introduction

The bacterial family Erwiniaceae includes diverse Gram-negative [1, 2] organisms including the genera *Mixta*, *Pantoea*, *Erwinia*, *Tatumella*, *Buchnera*, *Wigglesworthia*, *Phaseolibacter*, *Kalamiella* (LPSN, <http://www.bacterio.net/>). Three of these, *Mixta*, *Pantoea*, and *Erwinia*, have been the focus of numerous studies given their relevance to plant and human pathogens. For example, the genus *Erwinia* contains plant pathogens (e.g. E. amylovora) [3–5], epiphytes (e.g. E. tasmaniensis) [6, 7], species that may contribute to plant disease [8], and others that are insect-associated [9–11]. Similarly, *Pantoea*, a genus reported by Gavini et al.[12], includes species that inhabit a variety of environments [12–14], including human and plant pathogens [2, 15–20] as well as species associated with plant growth [21, 22], insects [23], and fungi [24, 25]. *Mixta*, originally *Pantoea*, is a novel genus first suggested by Palmer et al. [26] and then later proposed by the same researchers in 2018 [27]. Two species, M. calida and M. gaviniae (previously species of *Pantoea*), were originally found in infant formula and its production environment [13]. Fritz et al. [28] demonstrated that some strains of M. calida were linked with post-surgical meningitis in humans. Currently, there are only five *Mixta* species validly published (LPSN, <http://www.bacterio.net/>).

Understanding the phylogeny of bacterial species and genes allows researchers to investigate the evolutionary history of the species, their relatedness, and the individual genes that allow the species to survive under various conditions [26, 29–33]. The most common methods in phylogenetic analysis are those including 16S ribosomal RNA (rRNA) sequences [34], multi-locus sequence analysis (MLSA) with several housekeeping genes [34–36], and ribosomal MLSA with 53 structural ribosomal proteins [26, 29, 37]. However, the phylogenetic trees generated from these methods may not be very robust [34, 36, 38] because of the lower phylogenetic signal from the highly conserved housekeeping genes [34, 35, 39–42]. This is particularly seen within Erwiniaceae wherein most of our understanding of the phylogeny and interrelationships are based on the 16S rRNA genes [43–47] because of its ubiquity. However, it has insufficient resolution at the species level [36] and the genera *Pantoea*, *Erwinia*, *Tatumella*, and *Mixta* are often not monophyletic [48]. Another issue that may arise with MLSA is unknowingly including paralogues (i.e. genes originating from an intragenomic duplication event) or xenologues (i.e. genes originating from horizontal gene transfer) [26, 49].

The DNA topoisomerase (gyrase) subunit (gyrB), DNA-directed RNA polymerase subunit (rpoB), ATP synthase subunit (atpD), and initiation translation factor (infB) have been shown to be more reliable than 16S rRNA gene sequences for determining intra- and intergeneric relationships [50–53] and a robust method for the Erwiniaceae family [54]. The genera within Erwiniaceae are monophyletic when concatenated partial gyrB, rpoB, atpD, and infB gene sequences are used [38, 55, 56].

Nevertheless, despite efforts to find a suitable method for identifying the species tree for Erwiniaceae genera, *Mixta*’s position with relation to *Pantoea*, *Erwinia*, and *Tatumella* vastly differs between studies [2, 26, 27, 47, 48, 57–59]. Not only do different gene combinations provide different results, but different algorithms and models for alignments and phylogenetic trees also contribute to the contentious results [45, 60]. For example, Prakash et al. [2], Palmer et al. [26], Rezzonico et al. [48], and Gueule et al. [59] all used gyrB, rpoB, atpD, and infB nucleotide gene sequences in their phylogenetic analyses. Prakash et al. conducted a neighbour-joining statistical method while the others used a maximum-likelihood (ML) method. Three of the studies – Rezzonico et al., Prakash et al., and Gueule et al. – placed the *Mixta* species within the *Pantoea* genera whereas the analysis in Palmer et al. placed *Mixta* outside of the other three genera. Palmer et al. [27], Rezzonico et al. [48], and Brady et al. [58] also ran analyses with the four MLSA genes as amino acid sequences. These three studies show contentious results as Rezzonico et al. found that *Mixta* is related closest with *Tatumella*, Palmer et al. found that *Mixta* is related closest with *Pantoea*, and Brady et al. found *Mixta* within the *Pantoea* genus. In a separate study conducted by Palmer et al. [26], two ML trees were computed, one with 1039 core amino acid sequence and the other with 52 ribosomal MLSA nucleotide genes. Both ML trees contained species from *Mixta*, *Pantoea*, *Erwinia*, and *Tatumella*; however, the analysis with the core genes placed *Pantoea* and *Tatumella* as sister taxa followed by two *Mixta* species whereas the analysis with ribosomal MLSA genes placed *Pantoea* and *Tatumella* as sister taxa followed by *Erwinia* and then by the *Mixta* species.

There are two possible causes behind contentious results in phylogenetic analyses. One is the use of different methods, including the use of a different phylogenetic tree model from another study despite both studies conducting maximum likelihood trees and the same sequence type. Additionally, there is no common phylogenetic analysis model between nucleotide and amino acid sequences; therefore, evolutionary history is estimated differently. The other reason may be horizontal gene transfer (HGT) events within the *Erwiniaceae* or from a non-closely-related species. HGT events occur frequently within bacteria and often the genes provide the bacteria the ability to adapt to new environments and/or survive selective pressures.

The objective of this paper was to determine the cause behind the contentious phylogenetic results for the *Mixta* species. To accomplish this, homologous genes were extracted from ten complete or partial genomes - two for each of *Mixta*, *Pantoea*, *Erwinia*, and *Tatumella*, and two outgroups - that were retrieved from public databases. In this study, both the nucleotide and amino acid sequences for the same genes were separately analyzed. Model testing was done for each gene in the dataset and phylogenetic trees were computed using the recommended model. Distance matrices were exported from the phylogenetic trees and these were used to determine which species were most closely related to *Mixta* for each gene. Nucleotide and amino acid identity analyses were also conducted for each gene in order to compare the results between evolutionary history and sequence similarity. Phylogenetic trees were then done with either one gene or multiple genes to confirm results.

# Materials and Methods

## Genomes and Gene Sets

Publicly available full genomes were retrieved from NCBI comprising of ten type strains from *Mixta*, *Pantoea*, *Erwinia*, *Tatumella*, *Enterobacter*, and *Pseudomonas* (Table 1). *Enterobacter cloacae* is in the same order of *Enterobacterales* as *Mixta*, *Pantoea*, *Erwinia*, and *Tatumella* whereas *Pseudomonas syringiae* is in the same class of *Gammaproteobacteria*. The genomes were annotated with Prokka v1.14.1 [61] using default parameters and homologous genes were extracted using GET\_HOMOLOGUES software package [62] with the bidirectional best-hit search algorithm using default parameters. GET\_HOMOLOGUES extracted 954 homologous genes between the ten species.

Table 1. Genomes used in this study.

|  |  |  |  |
| --- | --- | --- | --- |
| Species | Strain | GenBank.Accession.No. | Level |
| Mixta calida | DSM\_22759 | GCA\_002953215.1 | complete genome |
| Mixta gaviniae | DSM\_22758 | GCA\_002953195.1 | complete genome |
| Pantoea agglomerans | NBRC\_102470 | GCA\_001598475.1 | contigs |
| Pantoea septica | LMG\_5345 | GCA\_002095575.1 | contigs |
| Erwinia amylovora | CFBP\_1232 | GCA\_000367625.2 | contigs |
| Erwinia tasmaniensis | ET1/99 | GCA\_000026185.1 | complete genome |
| Tatumella ptyseos | NCTC\_11468 | GCA\_900478715.1 | complete genome |
| Tatumella saanichensis | NML\_06-3099 | GCA\_000439375.1 | contigs |
| Enterobacter cloacae subsp cloacae | ATCC\_13047 | GCA\_000025565.1 | complete genome |
| Pseudomonas syringae pv syringae | ICMP\_3023 | GCA\_001401075.1 | scaffold |

In R [63], two filter functions were written to pass through gene files of the nucleotide sequences. Since both nucleotide and amino acid sequences were analyzed, the filters were only applied to the nucleotide sequences. A list of the passed nucleotide gene files was used to isolate the corresponding amino acid gene files to be used in later analyses.

The first filter passed through gene files that contained ten, single copy sequences. A minimum relative length was set where all sequences in a gene file had to be at least 90% the length of the longest sequence in the same file. For example, if the longest sequence in a file is 1000 bp, then the rest of the sequences had to be at least 900 bp. Of 954 genes, 38 genes were excluded for having at least extra sequence and an additional 175 genes were excluded because at least one sequence did not meet the length requirements.

However, there may be an evolutionary reason why at least one of the sequences did not meet the length requirements. The second filter took the 175 gene files that were denied by the first filter one at a time and split the sequences into two groups; the first group contained the gene sequences that were over the 90% length cutoff and the second contained the rest that were shorter than the cutoff length. The first requirement for this filter was that all sequences in the file were longer than 80% of the length of the longest genes in the file. The second requirement was that both representative of a genus were in the same group, either above or below the 90% cutoff. The two outgroup species were not required to be in the same group since they are not from the same genus. An additional 58 genes were added to the working dataset, resulting in 799 genes. A list of the 799 gene names was used to grab the corresponding amino acid files.

## Gene Alignment

Gene alignments were in R using a ClustalW algorithm. The gene fasta files were read into R using the readDNAStringSet() function (package: Biostrings) [64]. Alignments were done using the msaClustalW() function (package: msa) [65] using default parameters and setting maximum number of iterations to 100. The aligned functions were then converted into a readble format using msaConvert() (package: msa)[65] and written into a new fasta file using the write.fasta() function (package: seqinr) [66]

## Distance Matrices

After alignments, model testing was done for each gene through MEGA-CC [67] for both nucleotide and amino acid sequences. In a command line terminal, MEGA-CC is called along with a text file listing the file pathways to all 799 genes and a MAO file detailling the analysis preferences. Default paramaters were used for model testing. This resulted in a comma-separated values (CSV) file for each gene. These CSV files were read into R and the model for phylogenetic analysis was extracted based on lowest Bayesian Information Criterion (BIC). A list of the unique models and the number of genes that required each model are given in Table 2 for nucleotide sequences and in Table 3 for amino acid sequences.

Table 2. The number of genes that required each phylogenetic tree model according to model testing and the lowest BIC for nucleotide sequences.

|  |  |  |
| --- | --- | --- |
| Model | Number | Percent |
| GTR\_G | 89 | 11.14 |
| GTR\_G\_I | 42 | 5.26 |
| HKY\_G | 22 | 2.75 |
| HKY\_G\_I | 1 | 0.13 |
| K2 | 1 | 0.13 |
| K2\_G | 94 | 11.76 |
| K2\_G\_I | 1 | 0.13 |
| K2\_I | 2 | 0.25 |
| T92\_G | 313 | 39.17 |
| T92\_G\_I | 33 | 4.13 |
| TN93\_G | 149 | 18.65 |
| TN93\_G\_I | 52 | 6.51 |

Table 3. The number of genes that required each phylogenetic tree model according to model testing and the lowest BIC for amino acid sequences.

|  |  |  |
| --- | --- | --- |
| Model | Number | Percent |
| cpREV | 3 | 0.38 |
| cpREV\_G | 6 | 0.75 |
| Dayhoff | 2 | 0.25 |
| Dayhoff\_G | 11 | 1.38 |
| Dayhoff\_I | 2 | 0.25 |
| JTT | 1 | 0.13 |
| JTT\_G | 25 | 3.13 |
| JTT\_G\_F | 1 | 0.13 |
| JTT\_G\_I | 2 | 0.25 |
| JTT\_I | 1 | 0.13 |
| LG | 19 | 2.38 |
| LG\_G | 563 | 70.46 |
| LG\_G\_F | 27 | 3.38 |
| LG\_G\_I | 3 | 0.38 |
| LG\_I | 5 | 0.63 |
| mtREV24 | 1 | 0.13 |
| mtREV24\_G\_I | 1 | 0.13 |
| rtREV\_G | 3 | 0.38 |
| WAG | 1 | 0.13 |
| WAG\_G | 118 | 14.77 |
| WAG\_G\_F | 2 | 0.25 |
| WAG\_G\_I | 1 | 0.13 |
| WAG\_I | 1 | 0.13 |

The gene files were separated into groups according to the suggested model and text files were created with their file pathways. These text files were called in turn along with a corresponding MAO file through MEGA-CC to compute the distance matrices for each gene. Variance was estimated with 500 bootstraps, transitions and transversions were allowed, and gaps and missing data were treated with pairwise deletion.

The genetic distances and standard errors relative to M. calida and M. gaviniae were extracted to create plots in R, categorize the closest relatives, run generalized additive models to determine potential non-linear relationships between parameters and genetic distances, and identify genes for further analysis using phylogenetic trees. The ML phylogenetic trees were created using MEGAX and with a model suggested through model testing.

# Results and Discussion

# Conclusion

# Literature Cited

1. **Winslow C-EA, Broadhurst J, Buchanan RE, Krumwiede C, Rogers LA *et al.*** The families and genera of the Bacteria. Preliminary report of the Committee of the Society of American Bacteriologists on characterization and classiﬁcation of bacterial type. *Journal of Bacteriology* 1917;2:505.

2. **Prakash O, Nimonkar Y, Vaishampayan A, Mishra M, Kumbhare S *et al.*** Pantoea intestinalis sp. nov., isolated from the human gut. *International Journal of Systematic and Evolutionary Microbiology* 2015;65:3352–3358.

3. **Hauben L, Swings J**. *Genus XII. Erwinia*. 2nd ed. New York: Springer; 2005.

4. **Lopez MM, Rosello M, Llop P, Ferrer S, Christen R *et al.*** Erwinia piriflorinigrans sp. nov., a novel pathogen that causes necrosis of pear blossoms. *International Journal of Systematic and Evolutionary Microbiology* 2011;61:561–567.

5. **Matsuura T, Mizuno A, Tsukamoto T, Shimizu Y, Saito N *et al.*** Erwinia uzenensis sp. nov., a novel pathogen that affects European pear trees (Pyrus communis L.). *International Journal of Systematic and Evolutionary Microbiology* 2012;62:1799–1803.

6. **Geider K, Auling G, Du Z, Jakovljevic V, Jock S *et al.*** Erwinia tasmaniensis sp. nov., a non-phytopathogenic bacterium from apple and pear trees. *International Journal of Systematic and Evolutionary Microbiology* 2006;56:2937–2943.

7. **Mergaert J, Hauben L, Cnockaert MC, Swings J**. Reclassification of non-pigmented Erwinia herbicola strains from trees as Erwinia billingiae sp. nov. *International Journal of Systematic and Evolutionary Microbiology* 1999;49:377–383.

8. **Buonaurio R, Moretti C, Silva DP da, Cortese C, Ramos C *et al.*** The olive knot disease as a model to study the role of interspecies bacterial communities in plant disease. *Frontiers in plant science* 2015;6:434.

9. **Campillo T, Luna E, Portier P, Fischer-Le Saux M, Lapitan N *et al.*** Erwinia iniecta sp. nov., isolated from Russian wheat aphid (Diuraphis noxia). *International Journal of Systematic and Evolutionary Microbiology* 2015;65:3625–3633.

10. **Harada H, Oyaizu H, Kosako Y, Ishikawa H**. Erwinia aphidicola, a new species isolated from pea aphid, Acyrthosiphon pisum. *The Journal of General and Applied Microbiology* 1997;43:349–354.

11. **Skrodenyte-Arbaciauskiene V, Radziute S, Stunzenas V, Buda V**. Erwinia typographi sp. nov., isolated from bark beetle (Ips typographus) gut. *International Journal of Systematic and Evolutionary Microbiology* 2012;62:942–948.

12. **Gavini F, Mergaert J, Beji A, Mielcarek C, Izard D *et al.*** Transfer of Enterobacter agglomerans (Beijerinck 1888) Ewing and Fife 1972 to Pantoea gen. nov. as Pantoea agglomerans comb. nov. and Description of Pantoea dispersa sp. nov. *International Journal of Systematic and Evolutionary Microbiology* 1989;39:337–345.

13. **Popp A, Cleenwerck I, Iversen C, De Vos P, Stephan R**. Pantoea gaviniae sp. nov. and Pantoea calida sp. nov., isolated from infant formula and an infant formula production environment. *International Journal of Systematic and Evolutionary Microbiology* 2010;60:2786–2792.

14. **Liu Y, Wang S, Zhang D, Wei S, Zhao S *et al.*** Pantoea beijingensis sp. nov., isolated from the fruiting body of Pleurotus eryngii. *Antonie van Leeuwenhoek* 2013;104:1039–1047.

15. **De Baere T, Verhelst R, Labit C, Verschraegen G, Wauters G *et al.*** Bacteremic Infection with Pantoea ananatis. *Journal of Clinical Microbiology* 2004;42:4393.

16. **Fullerton D, Lwin AA, Lal S**. Pantoea agglomerans liver abscess presenting with a painful thigh. *European journal of gastroenterology & hepatology* 2007;19:433–435.

17. **Kratz A, Greenberg D, Barki Y, Cohen E, Lifshitz M**. Pantoea agglomerans as a cause of septic arthritis after palm tree thorn injury; case report and literature review. *Archives of disease in childhood* 2003;88:542–544.

18. **Lim P, Chen S-L, Tsai C-Y, Pai M-A**. Pantoea peritonitis in a patient receiving chronic ambulatory peritoneal dialysis (Case Report). *Nephrology (Carlton, Vic)* 2006;11:97–99.

19. **Schmid H, Schubert S, Weber C, Bogner J**. Isolation of a Pantoea dispersa-Like Strain from a 71-Year-Old Woman with Acute Myeloid Leukemia and Multiple Myeloma. *Infection* 2003;31:66–67.

20. **Rostenberghe H van, Noraida R, Wan Ibrahim WP, Hasan H, Mohamed Z *et al.*** The clinical picture of neonatal infection with Pantoea species. *Japanese journal of infectious diseases* 2006;59:120–121.

21. **Smits THM, Rezzonico F, Kamber T, Blom J, Goesmann A *et al.*** Metabolic versatility and antibacterial metabolite biosynthesis are distinguishing genomic features of the fire blight antagonist Pantoea vagans C9-1. *PLoS ONE* 2011;6:e22247–e22247.

22. **Kim HJ, Lee JH, Kang BR, Rong X, McSpadden Gardener BB *et al.*** Draft genome sequence of Pantoea ananatis B1-9, a nonpathogenic plant growth-promoting bacterium. *Journal of Bacteriology* 2012;194:729.

23. **Palmer M, Maayer P de, Poulsen M, Steenkamp ET, Zyl E van *et al.*** Draft genome sequences of Pantoea agglomerans and Pantoea vagans isolates associated with termites. *Standards in genomic sciences* 2016;11:23.

24. **Ma Y, Yin Y, Rong C, Chen S, Liu Y *et al.*** Pantoea pleuroti sp. nov., Isolated from the Fruiting Bodies of Pleurotus eryngii. *Current Microbiology* 2016;72:207–212.

25. **Rong C, Ma Y, Wang S, Liu Y, Chen S *et al.*** Pantoea hericii sp. nov., Isolated from the Fruiting Bodies of Hericium erinaceus. *Current Microbiology* 2016;72:738–743.

26. **Palmer M, Steenkamp ET, Coetzee MPA, Chan W-Y, Zyl E van *et al.*** Phylogenomic resolution of the bacterial genus Pantoea and its relationship with Erwinia and Tatumella. *Antonie van Leeuwenhoek* 2017;110:1287–1309.

27. **Palmer M, Steenkamp ET, Coetzee MPA, Avontuur JR, Chan WY *et al.*** Mixta gen. Nov., a new genus in the Erwiniaceae. *International Journal of Systematic and Evolutionary Microbiology* 2018;68:1396–1407.

28. **Fritz S, Cassir N, Noudel R, De La Rosa S, Roche P-H *et al.*** Postsurgical Pantoea calida meningitis: a case report. *Journal of Medical Case Reports* 2014;8:195.

29. **Bennett JS, Jolley KA, Earle SG, Corton C, Bentley SD *et al.*** A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus Neisseria. *Microbiology* 2012;158:1570–1580.

30. **Prasanna A, Mehra S**. Comparative Phylogenomics of Pathogenic and Non-Pathogenic Mycobacterium. *PLoS ONE* 2013;8:e71248.

31. **Angus A, Agapakis C, Fong S, Yerrapragada S, Estrada-de los Santos P *et al.*** Plant-Associated Symbiotic Burkholderia Species Lack Hallmark Strategies Required in Mammalian Pathogenesis. *PloS one* 2014;9:e83779.

32. **De Maayer P, Chan WY, Rubagotti E, Venter SN, Toth IK *et al.*** Analysis of the Pantoea ananatis pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts. *BMC Genomics* 2014;15:404.

33. **Fouts D, Matthias M, Adhikarla H, Adler B, Santos L *et al.*** What Makes a Bacterial Species Pathogenic?:Comparative Genomic Analysis of the Genus Leptospira. *PLoS Neglected Tropical Diseases* 2016;10:e0004403.

34. **Konstantinidis KT, Tiedje JM**. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Current Opinion in Microbiology* 2007;10:504–509.

35. **Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T *et al.*** Re-evaluating prokaryotic species. *Nature Reviews Microbiology* 2005;3:733–739.

36. **Glaeser SP, Kämpfer P**. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Systematic and Applied Microbiology* 2015;38:237–245.

37. **Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C *et al.*** Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* 2012;158:1005–1015.

38. **Brady C, Cleenwerck I, Venter S, Vancanneyt M, Swings J *et al.*** Phylogeny and identification of Pantoea species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). *Systematic and Applied Microbiology* 2008;31:447–460.

39. **Fox GE, Wisotzkey JD, Jurtshuk P**. How Close Is Close: 16S rRNA Sequence Identity May Not Be Sufficient To Guarantee Species Identity. *International Journal of Systematic Bacteriology* 1992;42:166–170.

40. **Konstantinidis KT, Tiedje JM**. Towards a genome-based taxonomy for prokaryotes. *Journal of bacteriology* 2005;187:6258–6264.

41. **Staley JT**. The bacterial species dilemma and the genomic-phylogenetic species concept. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2006;361:1899–1909.

42. **Richter M, Rosselló-Móra R**. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America* 2009;106:19126–19131.

43. **Francino M, Santos S, Ochman H**. *Phylogenetic relationships of bacteria with special reference to endosymbionts and enteric species*. New York: Springer; 2006.

44. **Hauben L, Moore ERB, Vauterin L, Steenackers M, Mergaert J *et al.*** Phylogenetic Position of Phytopathogens within the Enterobacteriaceae. *Systematic and Applied Microbiology* 1998;21:384–397.

45. **Naum M, Brown EW, Mason-Gamer RJ**. Is 16S rDNA a Reliable Phylogenetic Marker to Characterize Relationships Below the Family Level in the Enterobacteriaceae? *Journal of Molecular Evolution* 2008;66:630–642.

46. **Spröer C, Mendrock U, Swiderski J, Lang E, Stackebrandt E**. The phylogenetic position of Serratia, Buttiauxella and some other genera of the family Enterobacteriaceae. *International Journal of Systematic Bacteriology* 1999;49 Pt 4:1433–1438.

47. **Adeolu M, Alnajar S, Naushad S, Gupta RS**. Genome-based phylogeny and taxonomy of the ‘Enterobacteriales’: Proposal for enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morgane. *International Journal of Systematic and Evolutionary Microbiology* 2016;66:5575–5599.

48. **Rezzonico F, Smits THM, Born Y, Blom J, Frey JE *et al.*** Erwinia gerundensis sp. nov., a cosmopolitan epiphyte originally isolated from pome fruit trees. *International Journal of Systematic and Evolutionary Microbiology* 2016;66:1583–1592.

49. **Koonin EV**. Orthologs, Paralogs, and Evolutionary Genomics. *Annual Review of Genetics* 2005;39:309–338.

50. **Hedegaard J, Steffensen S, Nørskov-Lauritsen N, Mortensen K, Sperling-Petersen H**. Identification of Enterobacteriaceae by partial sequencing of the gene encoding translation initiation factor 2. *International Journal of Systematic Bacteriology* 1999;49 Pt 4:1531–1538.

51. **Dauga C**. Evolution of the gyrB gene and the molecular phylogeny of Enterobacteriaceae: A model molecule for molecular systematic studies. *International Journal of Systematic and Evolutionary Microbiology* 2002;52:531–547.

52. **Mollet C, Drancourt M, Raoult D**. rpoB sequence analysis as a novel basis for bacterial identification. *Molecular Microbiology* 1997;26:1005–1011.

53. **Paradis S, Boissinot M, Paquette N, Bélanger S, Martel E *et al.*** Phylogeny of the Enterobacteriaceae based on genes encoding elongation factor Tu and F-ATPase -subunit. *International journal of systematic and evolutionary microbiology* 2005;55:2013–2025.

54. **Brady CL, Venter SN, Cleenwerck I, Vandemeulebroecke K, De Vos P *et al.*** Transfer of Pantoea citrea, Pantoea punctata and Pantoea terrea to the genus Tatumella emend. as Tatumella citrea comb. nov., Tatumella punctata comb. nov. and Tatumella terrea comb. nov. and description of Tatumella morbirosei sp. nov. *International Journal of Systematic and Evolutionary Microbiology* 2010;60:484–494.

55. **Brady C, Venter S, Cleenwerck I, Engelbeen K, Vancanneyt M *et al.*** Pantoea vagans sp nov., Pantoea eucalypti sp nov., Pantoea deleyi sp nov and Pantoea anthophila sp nov. *International Journal of Systematic and Evolutionary Microbiology* 2009;59:2339–2345.

56. **Moretti C, Hosni T, Vandemeulebroecke K, Brady C, De Vos P *et al.*** Erwinia oleae sp. nov., isolated from olive knots caused by Pseudomonas savastanoi pv. savastanoi. *International Journal of Systematic and Evolutionary Microbiology* 2011;61:2745–2752.

57. **Brady CL, Goszczynska T, Venter SN, Cleenwerck I, Vos P de *et al.*** Pantoea allii sp. nov., isolated from onion plants and seed. *International Journal of Systematic and Evolutionary Microbiology* 2011;61:932–937.

58. **Brady CL, Cleenwerck I, Westhuizen L van der, Venter SN, Coutinho TA *et al.*** Pantoea rodasii sp. nov., Pantoea rwandensis sp. nov. and Pantoea wallisii sp. nov., isolated from Eucalyptus. *International Journal of Systematic and Evolutionary Microbiology* 2012;62:1457–1464.

59. **Gueule D, Fourny G, Ageron E, Le Flèche-Matéos A, Vandenbogaert M *et al.*** Pantoea coffeiphila sp. nov., cause of the ‘potato taste’ of Arabica coffee from the African great lakes region. *International Journal of Systematic and Evolutionary Microbiology* 2015;65:23–29.

60. **Octavia S, Lan R**. The Family Enterobacteriaceae. *The Prokaryotes: Gammaproteobacteria* 2014;225–286.

61. **Seemann T**. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.

62. **Contreras-Moreira B, Vinuesa P**. GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Applied and Environmental Microbiology* 2013;79:7696–7701.

63. **R Core Team**. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. <https://www.r-project.org/> (2019).

64. **Pages H, Aboyoun P, Gentleman R, Debroy S**. Biostrings: Efficient manipulation of biological strings.

65. **Bodenhofer U, Bonatesta E, Horejs-Kainrath C, Hochreiter S**. msa: an R package for multiple sequence alignment. *Bioinformatics* 2015;31:3997–3999.

66. **Charif D, Lobry JR**. Seqin{R} 1.0-2: a contributed package to the {R} project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Roman HE, Vendruscolo M (editors). *Structural approaches to sequence evolution: Molecules, networks, populations*. New York: Springer Verlag; 2007. pp. 207–232.

67. **Kumar S, Stecher G, Peterson D, Tamura K**. MEGA-CC: Computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics* 2012;28:2685–2686.

# Appendix