

# Bioinformatic basis to define the species formation within *Pectobacterium* and *Dickeya* bacterial genera

Peter Evseev<sup>1</sup>; Alexander Ignatov<sup>2</sup>; Fevzi Dzhaliilov<sup>2</sup>; Konstantin Miroshnikov<sup>1</sup>

<sup>1</sup>Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, <sup>2</sup>Research Center "PhytoEngineering" Ltd.

## Abstract

Soft rot *Pectobacteriaceae* (SRP) causing soft rot, aerial stem rot, and black leg diseases of potatoes and vegetables pose serious economic threat to agriculture World-wide. In this work we have studied phylogenetic relations within the genera *Pectobacterium* and *Dickeya* applying a full-genome comparison approach.

Phylogenetic trees were constructed using 16S rRNA sequences and different sets of housekeeping genes. Pan-genomic studies and proteomics was used to assess a suggested possibility of horizontal gene transfer. The analysis of NCBI environmental and metagenome databases have revealed possible SRP sequences belonging to the organisms not yet described taxonomically.

## Introduction

SRP are important pathogens of many crops and ornamental plants worldwide (Figure 1). In addition to potatoes, the bacteria comprising the genera *Pectobacterium* and *Dickeya* can infect other cultivated monocotyledonous and dicotyledonous plants – root crops (carrots, beets), fruits and berries, onions, solanaceous, cucurbits, brassicas, as well as ornamental plants [1, 2]. Members of these bacterial genera are quite similar both in their biochemical properties and in the manifested symptoms of plant diseases [3]. The difference between them is observed at the genomic level mainly [1].

Since 1990s, phylogenetic analysis of 16S and 23S rRNA gene sequences encoding the small and large ribosome subunits has been used to identify the evolutionary relationships of bacteria. This method was also applied to SRP [4]. However, such differentiation in the case of *Pectobacteriaceae* (and other groups of *Enterobacteriales*) is often unreliable or insufficient [2, 5].

Cluster analysis following the parameters characterizing the genome similarity cannot convey the evolutionary history and be a substitute for phylogeny reliably, despite the fact that the parameters of the genome similarity certainly belong to the data defining the taxonomic classification. At the moment, the upper threshold value of the average nucleotide identity, adopted as one of the criteria of the species, is 95-96%

In this research we used a complex approach comprising full-genome analysis, and phylogeny with the use of both 16S rRNA and single gene sequences, as well as different sets of concatenated sequences. We also analyzed available metagenome sequence databases to find more relatives of already described organisms. The results of this research can be used for the SRP taxonomical studies and related areas, e. g. for the SRP diagnostics.

## Methods

Bacterial genomes were downloaded from NCBI Genbank (<ftp://ftp.ncbi.nlm.nih.gov/genbank>). Search for bacterial core genes was conducted using UBCG (up-to-date bacterial core gene) pipeline [6] and BPGA (bacterial pan-genome analysis) pipeline [7]. The search for homologous sequences was conducted using a BLAST search with E-value < 1e-5. Custom BLAST databases were mounted with the BLAST tool (<https://blast.ncbi.nlm.nih.gov/>). Target genes were extracted from Genbank annotations. For some unannotated sequences, ORFs were found by Glimmer [8]. ORFs were validated and corrected by comparison with known homologous genes.

Alignments were made with MAFFT (L-INS-i algorithm) [9] and MUSCLE with auto settings [10]. The alignments were trimmed with trimAL [11] with gappyout settings. Trees were constructed using the maximum likelihood (ML) method with an RAxML program [12] with CAT I and GTR+G+I nucleotide model and GAMMA I BLOSSUM62 protein model, and the robustness of the trees was assessed by bootstrapping and with MrBayes [13] (rate matrix poisson, rate variation gamma, chain length 1,100,000, subsampling frequency 200 and unconstrained branch length settings). ANI (average nucleotide identity) calculations, a genome similarity matrix, and UPGMA (unweighted pair group method with arithmetic mean) clustering of *Pectobacteriaceae* taxa were done using a server (<http://enve-omics.ce.gatech.edu>).



Fig. 1. Soft rot disease of plants.

## Contacts

Peter Evseev  
Alexander Ignatov  
Konstantin Miroshnikov

Email: [petevseev@gmail.com](mailto:petevseev@gmail.com)  
Email: [an.ignatov@gmail.com](mailto:an.ignatov@gmail.com)  
Email: [kmi@bk.ru](mailto:kmi@bk.ru)



## D. Pan-genome analysis

Pan-genome analysis, conducted with BPGA employing 213 *Dickeya* and *Pectobacterium* genomes, defined 997 core proteins present in all genomes, 2422 to 3914 accessory proteins and 0 to 967 unique proteins. KEGG distribution demonstrates that the most common core proteins belong to genetic information processing and metabolism proteins. The greatest number of unique proteins was expectably found in deeply rooted and early diverged species.

We found that some contradictions of the obtained phylogeny with different genes or sets of genes can be caused by a significant role of horizontal gene transfer within the SRP population. Therefore, as different SRP are infecting the same host plant simultaneously, and they can be found together even inside vector insects, they have multiple chances for gene exchange via bacteriophage transduction or Type 4 Secretion System-mediated conjugation.

## E. Analysis of environmental sequences

In search of possible relatives of described SRP's we conducted a BLAST search of the SRP  $\beta$ -subunit of ATP synthase consensus sequence on NCBI databases. Phylogenetic tree with environmental and metagenome sequences is shown in Figure 7. The sequences belong to human gut and wasp gut microbiome. It was suggested earlier [21], that in the human intestine-associated bacterium *Clostridium butyricum-beijerinckii* contributes to the digestion of ingested pectic substance, producing exo-pectate lyase, endo-pectate lyase, and pectin methylesterase. The presence of SRP bacteria in the wasp gut is rather expected, since we know the examples of symbiotic *Pectobacterium* species [17] as well as a *Dickeya* pathogen of Pea Aphid [18].

## Conclusions

Taxonomic classification of these SRP bacteria is problematic because of low difference between rRNA sequences, which hampers the use of traditional 16S-rRNA phylogeny. On the other hand, exhaustive horizontal gene transfers create difficulties in the use of multilocus sequence comparisons. The search for homologs of SRP genes on the phage database suggests the important role of phages in the speciation of this taxonomic bacterial groups. The results of the present research illustrate the need in using combined full-genome data with the concatenated gene phylogeny.

Our studies revealed two groups of strains *Pectobacterium brasiliense* that could form two close species. The analysis of NCBI environmental and metagenome database sequences from human and wasp gut microbiome can testify of the existence of organisms, related to *Pectobacterium* and needs in focused research.

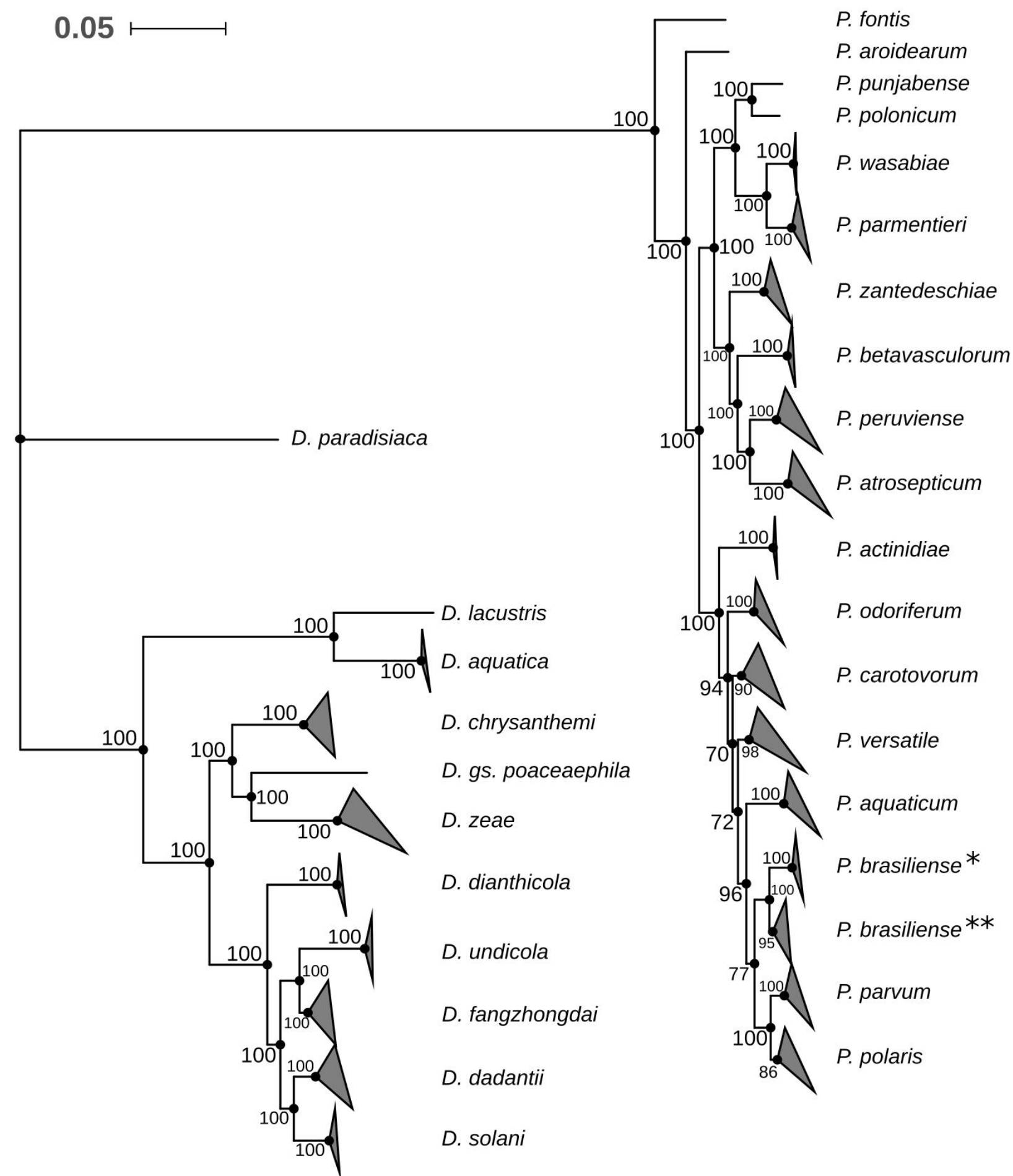


Fig. 4. RAxML tree based on 997 core protein concatenated sequences extracted from 213 *Dickeya* and *Pectobacterium* genomes.

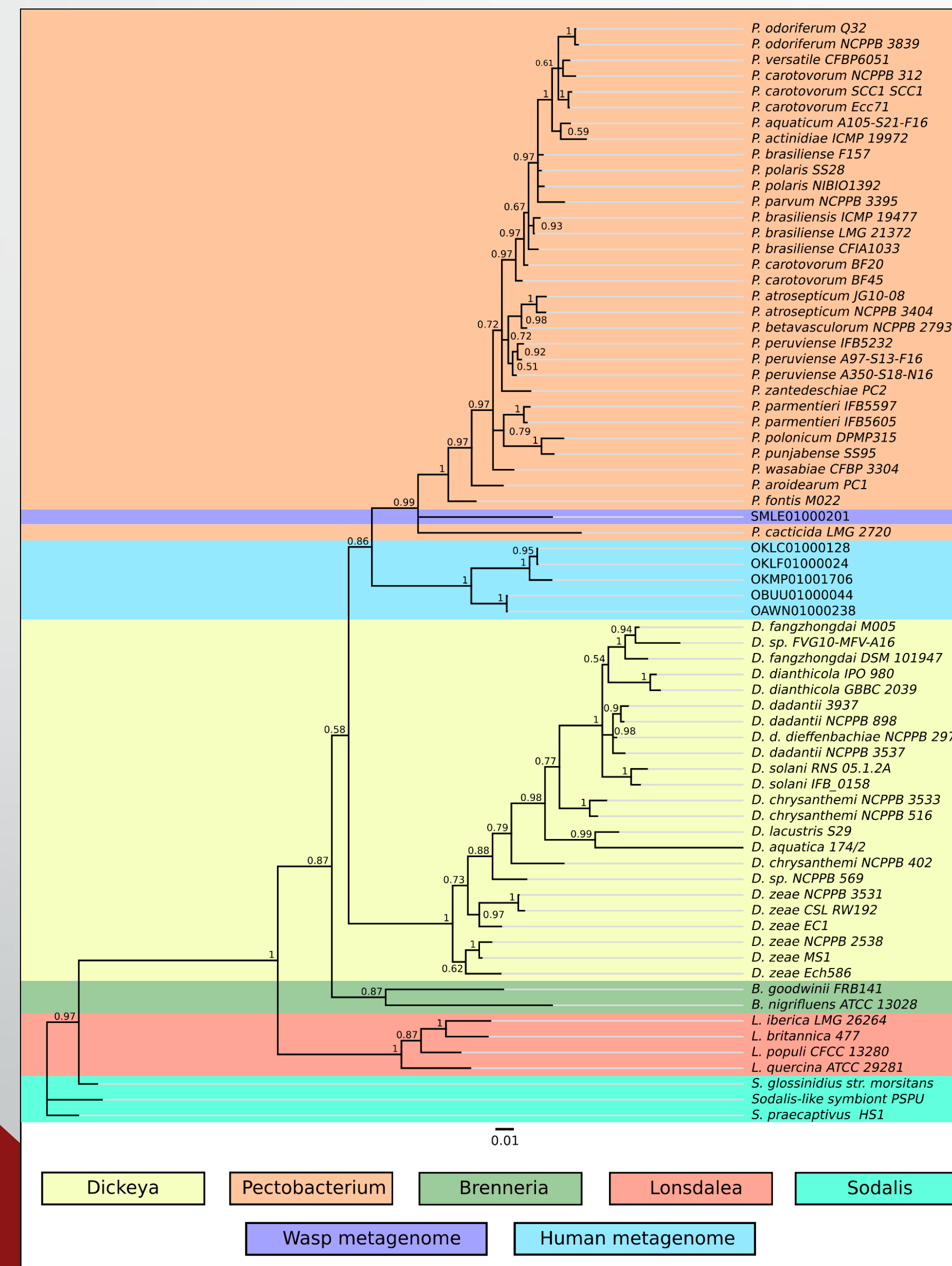


Fig. 5. MrBayes phylogenetic tree based on  $\beta$ -subunit of ATP synthase sequences.

## References

- [1] B. Ma et al., "Host range and molecular phylogenies of the soft rot enterobacterial genera *Pectobacterium* and *Dickeya*," *Phytopathology*, vol. 97, pp. 1150–1163, 2007.
- [2] M. Adeolu et al., "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., Morganellaceae fam. nov., and Budviciaceae fam. nov.," *Int. J. Syst. Evol. Microbiol.*, vol. 66, pp. 5575–5599, 2016.
- [3] M. C. M. Pérombelon, "Potato diseases caused by soft rot erwinias: An overview of pathogenesis," *Plant Pathol.*, vol. 51, pp. 1–12, 2002.
- [4] A. Nassar et al., "Characterization of *Erwinia chrysanthemi* by pectinolytic isozyme polymorphism and restriction fragment length polymorphism analysis of PCR-amplified fragments of *pel* genes," *Appl. Environ. Microbiol.*, vol. 62, pp. 2228–2235, 1996.
- [5] M. Naum et al., "Is 16S rDNA a Reliable Phylogenetic Marker to Characterize Relationships Below the Family Level in the Enterobacteriaceae?," *J. Mol. Evol.*, vol. 66, pp. 630–642, 2008.
- [6] S. I. Na et al., "UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction," *J. Microbiol.*, vol. 56, pp. 280–285, 2018.
- [7] N. Chaudhari et al., "BPGA – an ultra-fast pan-genome analysis pipeline," *Sci Rep*, vol. 6, 24373, 2016.
- [8] A. L. Delcher et al., "Improved microbial gene identification with GLIMMER," *Nucleic Acids Research*, vol. 27, pp. 4636–4641, 1999.
- [9] M. Katoh et al., "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform," *Nucleic Acids Res.*, vol. 30, pp. 3059–3066, 2002.
- [10] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, pp. 1792–1797, 2004.
- [11] S. Capella-Gutierrez et al., "trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses," *Bioinformatics*, vol. 25, pp. 1972–1973, 2009.
- [12] A. Stamatakis, "RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models," *Bioinformatics*, vol. 22, pp. 2688–2690, 2006.
- [13] J. P. Huelsenbeck et al., "MRBAYES: Bayesian inference of phylogeny," *Bioinformatics*, vol. 17, pp. 754–755, 2001.
- [14] A. Lukianova et al., "Morphologically Different *Pectobacterium brasiliense* Bacteriophages PP9g and PP10i: Deacetylation of O-Polysaccharide by the Tail Spike Protein of Phage PP9g Accompanies the Infection," *Front. Microbiol.*, vol. 10, 3147, 2020.
- [15] M. Pasanen et al., "*Pectobacterium parvum* sp. nov., having a Salmonella SPI-1-like Type III secretion system and low virulence," *Int. J. Syst. Evol. Microbiol.*, vol. 70, 2020.
- [16] S. Gadagkar et al., "Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree," *J. Exp. Zool. B: Mol. Dev. Evol.*, vol. 304, pp. 64–74, 2005.
- [17] N. Nakajima et al., "Degradation of pectic substances by two pectate lyases from a human intestinal bacterium, *Clostridium butyricum-beijerinckii* group," *J. Biosci. Bioeng.*, vol. 88, pp. 331–333, 1999.
- [18] A. Michalik et al., "Symbiosis in the green leafhopper, *Cicadella viridis* (Hemiptera, Cicadellidae)," *Arthropod Structure & Development*, vol. 43, pp. 579–587, 2014.

The research was supported by Russian Science Foundation grant #16-16-00073.