1 ## *Dickeya ananae* sp. nov., pectinolytic bacterium isolated from pineapple
2 ## (*Ananas comosus*)

3 Shefali Dobhal[1], Nicole Hugouvieux-Cotte-Pattat[2], Dario Arizala[1], Gamze Boluk Sari[1,3], Shu-
4 Cheng Chuang[1], Anne M. Alvarez[1], and Mohammad Arif [1*]

5 Author affiliations: [1]Department of Plant and Environmental Protection Sciences, University of Hawaii at
6 Manoa, Honolulu, HI, United States. [2]Univ Lyon, CNRS, INSA Lyon, UCBL, UMR 5240 Microbiologie
7 Adaptation et Pathogénie, F-69622 Villeurbanne, France. [3]Current Affiliation: Biological Control
8 Research Institute, Ministry of Agriculture, Adana, Turkey

9

10 *Correspondence: arif@hawaii.edu

11

12 **ABSTRACT**

13 Recently, species clustering within *Dickeya zeae* has been identified as complex, encompassing

14 validly published names, including *D. oryzae* and *D. parazeae*, with some strains potentially

15 delineating new species. In this study, genomes of strains isolated from a bacterial heart rot

16 outbreak in pineapple (*Ananas comosus* var. *comosus*) on Oahu, Hawaii, along with two strains

17 from pineapple in Malaysia, were sequenced. Orthologous average nucleotide identity (ANI) and

18 digital DNA-DNA hybridization (dDDH) values among the sequenced genomes ranged from

19 98.93-99.9% and 91.8-99.9%, respectively, supporting the classification of seven strains within

20 the same species. Comparisons of ANI and dDDH values between these seven strains and type

21 strains of *D. zeae, D. parazeae,* and *D. oryzae* ranged from 94.4-95.9% and 57.2-66.5%,

22 respectively. These values fall below the proposed boundaries for new species designation,

23 supporting the delineation of a novel species. Phylogenetic analyses, including 16S rRNA, *gapA*,

24 multi-locus sequence analysis (MLSA) of 10 housekeeping genes, whole-genome, and

25 pangenome analyses, were concordant and revealed a distinct monophyletic clade, separating

26 these strains from other members of the *D. zeae* complex, with *D. oryzae* as the closest relative.

27 Notably, a nitrogen fixation gene cluster comprising 28 genes, similar to the *Klebsiella* spp.

28 nitrogenase gene cluster, was found in the genome of the seven pineapple strains. Based on

29 polyphasic approaches, including ANI, dDDH, biochemical, physiological, and phylogenomic

30 analyses, we propose the reclassification in a new species of the five pineapple strains from

31 Hawaii A5391, A5410[T], A5611, A6136, and A6137, together with the two pineapple strains

32 from Malaysia CFBP 1272 and CFBP 1278, previously classified as *D. zeae*. We propose the

33 name *Dickeya ananae* sp. nov. for this taxon, represented by the type strain A5410^T (= ICMP

34 25020^T = LMG 33197^T).

Keywords

36 *Dickeya*, plant pathogen, genome, soft-rot *Pectobacteriaceae*, polyphasic taxonomy, virulence

37

Author Notes

39 Supplementary tables and figures are available in the online version of this article. The

40 nucleotide sequence of the 16S rRNA gene was deposited in the NCBI GenBank database under

41 accession number OR271567. The genome of the type strain A5410, from a previous study, is

42 available under accession number CP040816, with the associated BioProject PRJNA544764 and

43 BioSample SAMN11855518. Other genomes sequenced during this study have been deposited

44 under the following accession numbers: CP040816, CP146223, CP146227, CP146228,

45 CP146224, CP146225, and CP146226.

Abbreviations

47 ANI, average nucleotide identity; AP, alignment percentage; dDDH, digital DNA-DNA-
48 hybridization; MLSA, multi-locus sequence analysis; OGRI, overall genome relatedness indices

49

Introduction

51 The genus *Dickeya*, belonging to the order *Enterobacteriales* and the family *Pectobacteriaceae*,

52 comprises an important group of pectinolytic plant pathogenic bacteria. These bacteria affect and

53 have been isolated from a wide range of hosts, including vegetable crops, grains, and ornamental

54 plants and also from environmental sources [1-8]. The pathogens can adapt to a wide variety of

55 hosts and thrive in different climatic regions. These phytopathogens are notorious for causing

56 soft rot, blackleg, and wilt in a broad array of agronomically significant crops, including

57 potatoes, chicory, rice, cabbage, and ornamental plants, leading to substantial losses. These

58 pectinolytic bacteria produce and secrete through type I or II secretion systems a wide array of

59 plant cell wall-degrading enzymes (PCWDEs) [8-11] that cause maceration of the plant tissues,

60   releasing their contents contributing to bacterial growth in the host plant [12,13]. In some

61   pectinolytic bacteria, such as *D. dadantii* and *D. zeae*, type III secretion system has also been

62   reported to contribute to pathogenesis [14-16]. Additionally, other virulence factors, including

63   motility, biofilm formation, lipopolysaccharides, quorum sensing, siderophores, and the

64   CRISPR-Cas system, contribute to the pathogenicity of these bacteria [8, 10, 14, 17-19].

65   The genus *Dickeya* was first established by Samson in 2005, comprising six recognized species:

66   *D. chrysanthemi, D. dianthicola, D. dadantii, D. dieffenbachiae, D. zeae*, and *D. paradisiaca*

67   [20]. Later, *D. dieffenbachiae* was reclassified as a subspecies of *D. dadantii* [21]. In 2014, the

68   novel species *D. solani* was described as a predominant bacterial pathogen on potatoes in many

69   European countries and was also isolated from hyacinths [22-24]. Recent reclassification

70   proposed reassigning *D. paradisiaca* to the genus level and renaming it *Musicola paradisiaca*

71   [25]. The strains gathered into the "*D. zeae* complex" exhibit high diversity, prompting further

72   examination of this species complex [6]. In 2020, a divergent group of strains from *D. zeae* was

73   identified as a new species, *D. oryzae* [26]. Additionally, another species emerged from the *D.*

74   *zeae* complex, classified as *D. parazeae* [5]. However, this complex has not been fully resolved

75   and is still considered a mixture of more than one species [6].

76   To date, the genus *Dickeya* comprises 12 species, with the type strain of each species isolated

77   from infected plants or environmental sources. *D. dianthicola* was isolated from *Dianthus*

78   *caryophyllus*; *D. dadantii* from *Pelargonium capitatum*; *D. chrysanthemi* from *Chrysanthemum*

79   *morifolium*; *D. zeae* from *Zea mays*; *D. solani* from *Solanum tuberosum*; *D. fangzhongdai* from

80   *Pyrus pyrifolia*; *D. poaceiphila* from *Saccharum officinarum*; and *D. oryzae* from rice. Three

81   species —*D. aquatica, D. lacustris*, and *D. undicola*— were isolated from water. Additionally,

82   the type strain of *D. parazeae*, recently separated from *D. zeae*, was also isolated from water [5].

83   Another proposed species, "*D. colocasiae*", has been isolated from taro but has not yet been

84   validly published [27].

85   In this study, we report the description of a new species, including the strain A5410[T] [10], and

86   other strains isolated from pineapples in Hawaii, causing bacterial heart rot of pineapple. This

87   disease was first reported in Malaysia in 1957 [28] and later described in Costa Rica, Brazil, and

88   the Philippines [29, 30]. The first outbreak in Hawaiian pineapple plantations was reported in

89   2003 [31], with symptoms including water-soaked lesions around the apical meristem, brown

3

90 streaks on the lamina and mesophyll tissues, and characteristic blister-like lesions with light

91 brown exudate appearing as leaves began to rot. In advanced infection stages, the pineapple heart

92 and stem can easily separate from the plant, and young plants may be latently infected, leading to

93 fruit collapse at maturity. The bacterium spreads through wind and water splash, with insects like

94 pineapple fruit mites (*Steneotarsonemus ananas*) also disseminating the pathogen from infected

95 to healthy fruits. The strains causing this disease were identified as *Erwinia chrysanthemi*

96 (previous name of *Dickeya*) [31]. In 2008, the strain A5410 was isolated from Hawaiian

97 pineapples and initially classified under *D. zeae* [10]. Due to the heterogeneity of *D. zeae* strains,

98 further resolution of this species is needed [5, 6].

99 A recent phylogenetic study identified two major clades within the *Dickeya* genus: the first

100 includes members of the *D. zeae* complex, and the second, termed the "main *Dickeya* clade,"

101 comprises species like *D. dadantii, D. dianthicola* and *D. solani* [6]. Strain A5410 forms a

102 distinct branch with dDDH similarities of 66.5%, 58.2% and 57.2% with the *type* strains of *D.*

103 *oryzae*, *D. zeae*, and *D. parazeae*, respectively, supporting its unique status in a recent study of

104 the genus *Dickeya* [6]. The genome sequence of strain A5410 was previously reported (GenBank

105 accession number CP040816 with BioProject: PRJNA544764 and BioSample: SAMN11855518)

106 [10] and on the basis of comprehensive whole genome comparative analysis this strain was

107 identified as a potential candidate for a novel species. In the present research, we included six

108 other strains isolated from pineapple and presented compelling evidence supporting the

109 classification as a new species of seven pineapple strains isolated in Hawaii or in Malaysia. We

110 conducted a comprehensive analysis, employing polyphasic phylogenetic analysis utilizing the

111 *gapA* gene, 16S rRNA gene, MLSA, whole-genome sequencing, as well as phenotypic,

112 biochemical and physiological characterization. Additionally, we used overall genome

113 relatedness indices (OGRI) to determine genome sequence relatedness and analyzed the

114 phenotypic characteristics of the strains. These data confirmed that the sequenced pineapple

115 strains should be assigned to a new species of *Dickeya*. Therefore, we propose that a novel

116 species name be established for this group of strains, *Dickeya ananae* sp. nov.

117 **Methods**

118 **Bacterial strains, isolation and ecology**

119   The strains A5391, A5410^T, A5611, A6136, and A6137 were isolated in 2007 on Oahu (21.31 N,

120   158.02 W), Hawaii, from pineapple leaves (*Ananas comosus*) showing symptoms of pineapple

121   heart rot. The bacterium was isolated by culturing on 2,3,5-triphenyltetrazolium chloride (TZC)

122   agar medium, which contains dextrose (5 gL⁻¹), peptone (10 gL⁻¹), 0.001% sterilized TZC, and

123   agar (18 gL⁻¹) as described by Norman & Alvarez [32]. The plates were incubated at 28°C for

124   24 hours. A single colony was streaked onto DPA (dextrose peptone agar: DPA-dextrose 5

125   gL⁻¹, peptone 10 gL⁻¹, and agar 18 gL⁻¹) and incubated at 28°C for 24 hours. Additionally,

126   the colonies were streaked on crystal violet pectate (CVP) medium [33] to observe pit formation.

127   The purified single colonies were stored at -80°C in a 30% glycerol suspension. Two additional

128   strains isolated from pineapple in Malaysia in 1961, CFBP 1278 (NCPPB 1121, A5417) and

129   CFBP 1272 (NCPPB 1125, A5421), were included in the study. All strains were routinely

130   cultured on DPA media at 28°C.

131

132   **16S rRNA and *gapA* gene phylogeny**

133   The bacterial genomic DNA was isolated using the DNeasy Blood and Tissue Kit (Qiagen,

134   Germantown, MA). For strain identification, the 16S rRNA gene was PCR amplified using the

135   universal primers 27F and 1492R and recommended PCR conditions [34]. The PCR mix

136   consisted of 10 µL of 2x GoTaq Green Master Mix, 1 µL of each primer (27F and 1492R; 5

137   µM), 1 µL of genomic DNA template, and 7 µL of nuclease-free water. Five microliters of the

138   amplified product were enzymatically cleaned using 1 µL of Exo (exonuclease I) and 1 µL of

139   SAP (shrimp alkaline phosphatase; GE Healthcare, Little Chalfont, UK), following the

140   manufacturer's instructions. The amplicons were then subjected to bi-directional DNA

141   sequencing (Sanger method) at the GENEWIZ facility (Genewiz, La Jolla, CA). The obtained

142   sense and anti-sense sequences were aligned and manually edited for accuracy using Geneious

143   Prime 2023.2.1 (Biomatters, Inc., Boston, MA). The consensus sequence was compared and

144   mapped to the whole genome of each strain and also with the NCBI GenBank nucleotide and

145   genome databases using the BLASTn tool. The sequences of the partial 16S rRNA from A5410^T

146   and other strains have been deposited in the NCBI under the accession number: [OR271567;

147   Supplementary File 1]. The 16S rRNA phylogenetic tree was later constructed by uploading the

148   whole genome sequences of the seven strains used in this study, along with 37 whole genomes of

149   other *Dickeya* species, to the Type Strain Genome Server (TYGS at https://tygs.dsmz.de). The

5

150   16S rRNA phylogenetic tree was generated using the TYGS and was visualized using the web-

151   based tool Interactive Tree Of Life (iTOL v6, https://itol.embl.de) [35].

152   Amplification of the partial *gapA* gene, commonly employed for strain classification within the

153   genus *Dickeya* [36, 37], was carried out using gapA primers. The amplified product was

154   enzymatically cleaned as previously explained. The amplicons were subjected to bi-directional

155   DNA sequencing at the GENEWIZ facility. Obtained sense and anti-sense sequences were

156   aligned and manually edited for accuracy using Geneious Prime. The consensus sequence was

157   compared and mapped to the reference whole genome of each strain, as well as with the NCBI

158   GenBank nucleotide and genome databases using the BLASTn tool. The *gapA* gene based

159   phylogenetic tree was reconstructed using the Maximum Likelihood method and the Tamura-Nei

160   model in MEGA 11 [38,39]. The stability of the phylogenetic tree's topology was assessed

161   through a bootstrap analysis with 1,000 replications.

162

163   **Genome sequencing, assembly, overall genome relatedness indices, MLSA, 16S rRNA**

164   **phylogeny and phylogenomics**

165   Purified bacterial colonies were used for genomic DNA extraction using the QIAGEN Genomic-

166   tip 100/G (Qiagen, Valencia, CA), following the manufacturer's instructions. The genome of the

167   type strain A5410^T, as reported by Boluk *et al*. [10] is available at the NCBI under the accession

168   number CP040816. In the process of taxonomic reclassification, the genome of four additional

169   Hawaiian strains —A5391, A5611, A6316, and A6317— as well as two strains previously

170   reported from Malaysia, CFBP 1278 and CFBP 1272, were sequenced. The bacterial genomic

171   DNA was extracted using the previously mentioned method, quantified with Qubit 4 using the

172   Qubit dsDNA High Sensitivity (HS) Kit (Invitrogen, Carlsbad, CA), and analyzed on a 1.5%

173   agarose gel. The libraries were generated without shearing to optimize sequencing read length,

174   utilizing the Native Barcoding Kit 24 (SQK-NBD114-24) and following the protocol available

175   on the manufacturer's website (https://nanoporetech.com/resource-centre/protocols). A barcode-

176   indexed library was loaded onto the R10.4.1 flow cell and sequenced using the Oxford Nanopore

177   MinION Mk1C device (Oxford Nanopore Technologies, ONT, Oxford, UK). The obtained

178   FAST5 sequences were used for base calling and demultiplexing using GUPPY v6.3.2 on

179   MANA, a high-performance computing cluster at the University of Hawaii at Manoa. The *De*

6

180   *Novo* Assemble Long Reads tool of CLC Genomics Workbench v22.0 (Qiagen) was used with

181   default parameters to assemble the genomes.

182   Additionally, Illumina sequencing was performed at the DNA Technologies and Expression

183   Analysis Core Laboratory at UC Davis Genome Center (Davis, CA). Illumina DNA libraries

184   were prepared using 10 ng of input DNA per sample with the plexWell LP384 kit (seqWell,

185   Beverly, MA), following the manufacturer's instructions. This kit uses two rounds of transposase

186   reactions followed by PCR amplification to generate a pool of dual-barcoded indexed sequencing

187   libraries. In the first tagmentation reaction, 96 samples were individually barcoded using

188   transposase reagents loaded with 96 different barcoded oligos, adding Illumina i7 indices and

189   fragmenting the DNA. After a 15-minute incubation at 55°C, the reaction was halted with an

190   SDS-containing buffer, followed by a 10-minute incubation at 68°C. The samples were then

191   pooled in groups of 48 and purified using SPRI-bead purification with MAGwise beads. After

192   elution in TRIS buffer, a second tagmentation was performed on the two pools, adding Illumina

193   i5 indices to identify the sample plate. The second reaction followed similar incubation,

194   termination, and elution steps as the first. The samples then underwent a fill-in reaction and PCR

195   amplification, resulting in a pool of 96 sequencing libraries. The fragment length of the library

196   pool was analyzed using a Bioanalyzer 2100 instrument (Agilent, Santa Clara, CA). The library

197   pool was quantified by qPCR using a Kapa Library Quant RT-qPCR kit (Kapa/Roche, Basel,

198   Switzerland) and sequenced on an Illumina NovaSeq 6000 instrument with paired-end 150 bp

199   reads (Illumina, San Diego, CA). The Illumina paired-end short reads, along with the corrected

200   long reads from Oxford Nanopore, were used to generate precise and comprehensive hybrid

201   assemblies utilizing the "Polish with Reads" tool in CLC Genomics Workbench v22.0 (Qiagen)

202   with default parameters. Contamination checks for the genomes were subsequently performed on

203   the BV-BRC 3.26.4 web server (https://www.bv-brc.org/). Genome annotations were conducted

204   using the Prokaryotic Genome Annotation Pipeline (PGAP v4.10) on NCBI [40] and also via the

205   Rapid Annotation using Subsystem Technology (RAST v2.0) web server [41]. The genomes

206   were submitted to NCBI under the reference numbers: CP040816 (strain A5410[T]), CP146223

207   (strain A5391), CP146227 (strain CFBP 1278), CP146228 (CFBP 1272), CP146224 (strain

208   A6136), CP146225 (strain A6137), and CP146226 (strain A5611).

209   The whole genome sequences of the seven new strains were compared to that of the type strains

210   of the 12 recognized *Dickeya* species using overall genome relatedness indices (OGRI),

7

211   including ANI calculated with the OrthoANI algorithm implemented in Orthologous Average

212   Nucleotide Identity Tool (OAT) software [42], alignment percentage (AP) using CLC Genomics

213   Workbench 22.0.2 (Qiagen), and dDDH calculated using the Genome-to-Genome Distance

214   Calculator version 3.0 (http://ggdc.dsmz.de/) [43], with the recommended formula 2. The

215   calculated matrix with values for ANI and AP, as well as ANI and dDDH was generated.

216   Additionally, OGRI was calculated between *Dickeya ananae* sp. nov., and other strains available

217   in the NCBI database under the names *D. zeae, D. oryzae*, and *D. parazeae*.

218   To determine the phylogenetic position of A5410[T], a whole genome-based taxonomic analysis

219   was performed using the Genome BLAST Distance Phylogeny (GBDP) approach by uploading

220   the whole genome sequences to the Type Strain Genome Server (TYGS), a free bioinformatics

221   platform available at https://tygs.dsmz.de [43]. The distinctiveness of these strains within the

222   genus *Dickeya* was corroborated through multi-locus sequence analysis (MLSA) [44-45], using a

223   methodology similar to that outlined for *D. undicola* [26,46]. The sequences of ten housekeeping

224   genes, including *dnaA, dnaX, fusA, gyrB, infB, recA, recN, rplB, atpD*, and *gapA*, were extracted

225   from the whole genomes of these strains and other *Dickeya* species available in the NCBI

226   database. Each sequence was aligned using Geneious Prime, and a phylogenetic tree was

227   constructed using MEGA 11. Evolutionary history was inferred using the Maximum Likelihood

228   method and the Tamura-Nei model, with 1,000 bootstrap replicates performed.

229   **Pan and core genome analyses**

230   Pan- and core-genome analyses were performed using the whole genome sequences of the newly

231   described species and type strains of closely related species within the genus *Dickeya*. Annotated

232   files (GFF3) were generated using Prokka v1.14.6 [47] and used for pan-genome analysis with

233   the Roary v3.13.0 pipeline [48], employing a minimum BLASTp identity of 90%. Core genomes

234   were aligned using PRANK, a probabilistic multiple alignment program [48,49]. Core and

235   unique gene counts for the type strain of each *Dickeya* species were determined from the Roary

236   output (with a BLASTp identity threshold of 80%), and these data were used to create flower

237   plots using R scripts in RStudio. A core gene phylogenetic tree was constructed using the ML

238   tree inference tool RAxML-NG v0.8.0 (Randomized Axelerated Maximum Likelihood - Next

239   Generation) [50]. The DNA substitution model, General Time Reversible (GTR) + GAMMA

240   (G), was applied separately to the core genomes of the *Dickeya* species, with 1,000 bootstraps.

241 The resulting core genome phylogenetic tree was visualized using iTOL. The Roary matrix,

242 which delineates the presence and absence of core and accessory genes, was combined with the

243 core genome maximum likelihood tree, and the results were presented using roary_plots.py [48].

Phenotypic characteristics of the pineapple strains

245 The bacteria were tested for swimming and swarming motilities using a semisolid medium [51].

246 The swimming assay medium contained 10 g/L tryptone, 5 g/L NaCl, and 3 g/L agar, while the

247 swarming medium had 10 g/L tryptone, 5 g/L NaCl, and 4 g/L agar. Five microliters of overnight

248 culture with an OD at 600 nm of 1.0 were spot inoculated at the center of each medium plate,

249 and the diameter of bacterial growth was recorded. Each assay was performed in triplicate. The

250 GEN III MicroPlate (BIOLOG) was used to assess carbon source utilization and chemical

251 sensitivity of strains A5391, A5410^T, A5611, A6316, A6317, CFBP 1278 (A5417), and CFBP

252 1272 (A5421), following the manufacturer's instructions. The bacterial suspension was prepared

253 in inoculating fluid (IF) by suspending pure bacterial colonies from overnight cultures grown on

254 BUG (Biolog Universal Growth) media, adjusted to 96% transmittance using a SPECTRONIC

255 20D+ spectrophotometer (Thermo Fisher Scientific, MA, USA). A homogeneous suspension was

256 prepared, and 100 µl of the suspension was dispensed into each well of the Biolog GEN III

257 MicroPlate. The plates were incubated at 28 °C for 24 hours, read in the Biolog Microstation

258 ELx808BLG Reader (Biotek Instruments), and the results were recorded as utilization or no

259 utilization for carbon sources, and sensitivity or resistance for chemicals.

260 Biochemical test kits API 20E, APIZYM, and API 50CH from bioMérieux (Marcy-l'Étoile,

261 France) were used to investigate carbohydrate fermentation, biochemical reactions, and

262 enzymatic activities. The bacteria were cultured on NAG media, and 24-hour plates were used

263 for the tests. For the API 20E strips, which consist of 21 miniaturized biochemical tests, the

264 bacterial inoculum was prepared by suspending a single isolated colony in API suspension

265 medium (0.85% NaCl). The inoculum was then used immediately to fill the test tubes and

266 cupules, following the manufacturer's instructions. The test was incubated at 28±2°C, and results

267 were recorded after 24 hours. For the API ZYM enzymatic assay, which identifies 19 enzymatic

268 reactions semi-quantitatively, the bacterial inoculum was prepared in API suspension medium

269 with turbidity adjusted to 5 McFarland. Then, 65 µl of the suspension was added to each cupule

270 and incubated at 28±2°C for 4 hours and 30 minutes. The test was developed by adding one drop

9

271   each of ZYM A and ZYM B reagents and observed after 5 minutes for color changes to

272   determine enzyme activity. API 50CH strips, designed to test the utilization of 49 different

273   carbohydrate sources, were used following the *Enterobacteriaceae* instructions [52]. Briefly,

274   colonies from 18-24 hour-old NAG plates were suspended in sterile distilled water, adjusted to 4

275   McFarland, mixed with CHB/E medium, and used to inoculate the strips. Strips were incubated

276   at 28±2ºC and results were observed at 24 and 48 hours. Cytochrome oxidase activity was tested

277   using OxiStrips^{TM} (Hardy diagnostics, Santa Maria, CA), with a color change to dark blue within

278   90 seconds indicating a positive result. Catalase activity was confirmed by the addition of 3%

279   H_2O_2, with bubble formation indicating a positive result.

280   Extracellular enzyme production, including pectate lyase (pel), cellulase (cel), and protease (prt),

281   was assessed using a semi-quantitative method [53]. Assays involved applying 50 µl of a

282   bacterial suspension with OD600 ∼1.2 to wells in media plates and incubating at 28ºC. The pel

283   assay was developed with 4N HCl after 10 hours, the cel activity with 2% Congo red and 5 M

284   NaCl, and the prt activity observed as a halo after 24 hours of incubation.

285   The maceration ability of these strains was evaluated through virulence assays on potato, taro,

286   and onion. The surface of each tuber or bulb was sterilized with 0.06% sodium hypochlorite for

287   10 minutes, followed by three 3-minute washes in sterile distilled water. The tubers and bulbs

288   were air-dried in a BSC II, sliced with a sterile knife, and placed in sterile Petri plates. A 10 µl

289   bacterial suspension of 10^8 CFU/ml was applied to the center of each slice, and plates were

290   incubated at room temperature (26°C). Results were recorded after 24 hours, with two replicates

291   per strain.

Transmission Electron Microscopy (TEM)

293   To examine bacterial morphology, transmission electron microscopy (TEM) was conducted.

294   Samples were prepared following the protocol from the Biological Electron Microscope Facility,

295   University of Hawaii at Manoa. Bacteria were cultured overnight in Tryptic Soy broth at 28°C

296   with shaking at 120 rpm. For TEM, bacteria were mounted onto glow-discharged 200 mesh

297   carbon-coated Formvar-coated copper grids, negatively stained with 1% uranyl acetate, and

298   rinsed thrice with sterile distilled water. Stained grids were viewed on a Hitachi HT7700 TEM at

299   100 kV, and high-resolution images were captured using an AMT BioSprint16M-ActiveVu: 16

300   megapixel cooled 4895 x 3920-pixel CCD camera. Cells were measured in length and diameter

301   with three replicates.

Antibiotic sensitivity assay

303   An antibiotic disc diffusion assay was performed. Briefly, a single colony from a 12-hour grown

304   culture plate was inoculated into 10 ml of nutrient broth supplemented with 0.2% glucose and

305   incubated at 28°C with shaking at 150 rpm for 8 hours. The bacterial inoculum was adjusted to

306   an OD_{600} of 1.0, and 100 µl was evenly spread onto NAG plates. The plates were allowed to dry

307   in the biosafety cabinet. The antibiotic discs were prepared by absorbing discs with the following

308   13 antibiotics: spectinomycin (100 mg/ml), gentamicin (50 mg/ml), kanamycin (50 mg/ml),

309   tetracycline (40 mg/ml), penicillin (50 mg/ml), cefalexin (50 mg/ml), chloramphenicol (50

310   mg/ml), trimethoprim (25 mg/ml), bacitracin (50 mg/ml), hygromycin B (50 mg/ml),

311   vancomycin (50 mg/ml), cefotaxime (50 mg/ml), and carbenicillin (50 mg/ml). Once the discs

312   were dried, one disc was placed at the center of each agar plate by gently pressing down with a

313   sterile tweezer; 2 plates were prepared for each antibiotic tested. For the control, a disc

314   impregnated with sterile distilled water was placed in the center of the culture plate. The plates

315   were inverted and incubated for 24 hours at 28°C. The area of inhibited bacterial growth was

316   measured and recorded after 24 hours.

317

Results and Discussion

Genomic features and genomic relatedness

320   The genome of strain A5410[T] was sequenced by Boluk *et al.* [10] and used for comparative

321   genomic analyses. This study provided evidence that the taxonomic position of this pineapple

322   strain should be revised, as also pointed out by Hugouvieux-Cotte-Pattat *et al.* [6]. In the current

323   study, we included the bacterial strains CFBP 1278 and CFBP 1272, isolated from pineapple in

324   Malaysia, along with strains A5391, A6136, A6137, and A5611, which were isolated in 2006

325   during a bacterial heart rot outbreak in Hawaiian pineapple. The strains isolated from pineapple

326   in Hawaii were collected from the base of leaves from different pineapple samples. The six

327   pineapple strains were sequenced using Illumina NovaSeq and MinION sequencing platforms.

328   The hybrid assembly approach employed yielded high-quality genomes with coverage ranging

329   from 364x to 545x. Genomic features are presented in Table 1. A CheckM-based assessment

330 demonstrated genome completeness within the range of 99.9-100%, with contamination rates

331 ranging from 2.2% to 4%. The genome sizes ranged 4,779,825 bp to 4,837,911 bp, with no

332 plasmids detected. The G+C content ranged from 53.48% to 53.56 mol%. Results from the

333 Bacterial and Viral Bioinformatics Resource Center (BV-BRC version 3.34.11) indicated that the

334 A5410^T genome encodes a total of 4,555 protein-coding sequences, 22 rRNAs, 75 tRNAs, 3

335 CRISPR arrays, and 884 hypothetical proteins. The six other genomes of pineapple strains

336 showed similar values (Table 1).

337 Annotation and comparison of the genome sequences using the RASTtk Model SEED database

338 revealed that 28-29% of the predicted coding sequences were assigned to subsystems. The

339 number of subsystems predicted by the RAST annotation server ranged from 342 to 345

340 (Supplementary Table 1). The subsystems most prominently represented among the seven

341 genomes included "amino acids and derivatives", along with "carbohydrates". Additional

342 common subsystems were associated with "protein metabolism", "cofactors, vitamins, prosthetic

343 groups, pigments", and "respiration".

344 A comprehensive comparative genomics analysis conducted by Boluk *et al.* [10] identified a

345 unique nitrogen fixation cluster in the A5410^T genome, which was absent in the strains of *D.*

346 *oryzae*. The six other genomes sequenced in this study contain a similar cluster, including those

347 from Malaysia. The *nif* gene cluster was absent in all type strains of the genus *Dickeya*. It was

348 also absent in strains within the *D. zeae* species complex but was present in strains CE1 (isolated

349 from *Canna*, China), JZL7 (isolated from *Clivia*, China), and BRIP 64263 (isolated from

350 pineapple, Australia). (Figure 1). The *nif* gene cluster of the A5410^T genome consisted of a total

351 of 20 genes spanning a 25,607 bp region. The arrangement of this gene cluster is similar to that

352 of the nitrogen fixation *nif* cluster of *Klebsiella variicola* and *K. pneumoniae*, comprising the

353 three structural genes: *nifH* (nitrogenase reductase), *nifD* (nitrogenase molybdenum-iron protein

354 α subunit), and *nifK* (nitrogenase molybdenum-iron protein β subunit). It also includes *nifF* and

355 *nifJ* (electron donation to nitrogenase); *nifE, nifN, nifX, nifU, nifS,* and *nifV* (Fe-Mo cofactor

356 biosynthesis); *nifM* (nitrogenase maturation); *nifA* and *nifL* (positive and negative regulators of

357 *nif* gene expression); *nifW* (protection, maturation, and activation of the Fe-Mo protein); and *nifT*

358 of unknown function [54-58]. An additional gene encoding a hypothetical protein is present

359 between *nifJ* and *nifH* in the *nif* gene cluster of the seven pineapple strains (Figure 1).

12

360 Plant cell wall-degrading enzymes, including pectinases, cellulases, and proteases, are among the

361 most important virulence factors of soft rot *Pectobacteriaceae*. The genome of A5410[T] was

362 previously analyzed for virulence-related genes by Boluk *et al.* [10], and in this study, we also

363 evaluated the presence of these genes in the six other pineapple strains. Genes encoding pectate

364 lyase (*pelADE, pelBCZ, pelI, pelL, pelN, pelX*, and *pelW)*; other pectinases (*pemA, , paeY, faeD,*

365 *pehK, pehX, rhiE*); a cellulases (*celZ*); proteases (*prtABCDEFG*); a xylanase (*xynA*); and a

366 phospholipase (*plcA*) were present in the seven genomes. The genes *pnlH, pehN*, and *pemB, and*

367 *avrM* encoding a pectin lyase, a polygalacturonase, a pectin methylesterase, and an avirulence

368 protein respectively, are absent in these genomes, as well as in *D. zeae, D. parazeae* and *D.*

369 *oryzae* [6]. All the genomes of the proposed *D. ananae* sp. nov. members possess the *avrL*

370 avirulence-associated gene, unlike the *D. oryzae* strains ZYY5[T], DZ2Q, and EC1 [5]. The *pnlG,*

371 which encodes *a* pectin lyase, is absent in *D. ananae* members as well as *D. parazeae,* but is

372 present in *D. oryzae* and *D. zeae,* as noted by Hugouvieux-Cotte-Pattat *et al.* [5].

373 Furthermore, the genomes of the proposed *D. ananae* sp. nov. members also contained *cyt* genes

374 encoding Type-1Ba cytolytic delta-endotoxin and Type-2Aa cytolytic delta-endotoxin, which

375 were absent in *D. oryzae* ZYY5[T], DZ2Q, and EC1. The *cyt* genes were reported to be present in

376 some, but not all, strains of *D. oryzae* [5].

Genomic Relatedness

378 Whole genome alignment and 16S rRNA sequence analysis performed using the TYGS genome

379 server identified A5410[T] and the six other pineapple strains as a novel species, all belonging to

380 the same group. The taxonomic position of the pineapple strains was first clarified by comparing

381 the pairwise ANI and dDDH values between the A5410[T] genome and the type strains of each

382 *Dickeya* species (Figure 2). These ANI and dDDH values ranged from 79.6-96% and 23.1-

383 66.8%, respectively, falling below or equal to the thresholds of 95–96% (ANI) and 70% (dDDH)

384 for differentiating bacterial species, confirming the taxonomic novelty of strain A5410[T] [59-62].

385 The ANI values between the A5410[T] genome and the other six genomes sequenced in this study

386 (strains A6136, A6137, A5611, A5391, CFBP 1272, and CFBP 1278) ranged from 99.01-

387 99.99% (Supplementary Table 2). Additionally, the dDDH values between these seven genomes

388 ranged from 91.8-99.9%, strongly supporting that all seven strains belong to the same species.

389 Alignment coverage, represented by AP values between genomes, is one of the criteria evaluated

390 in the description of new species by Parsenen *et al.* [63]. The AP values between the genomes of

391 A5410^T and the six other strains ranged from 97.56-99.42%, strongly supporting their genomic

392 similarity and classification within the same group. In comparison, AP values between the

393 A5410^T genome and those of type strains of each *Dickeya* species range from 26.7 (with *D.*

394 *lacustris*) to 87.2 (with *D. zeae*) (Figure 3).

395 Given the reported diversity among members of the *D. zeae* complex [5,6], the genomes of the

396 seven pineapple strains were further evaluated for ANI and dDDH values in comparison to

397 strains of closely related groups within the *D. zeae* complex, including the species *D. parazeae,*

398 *D. zeae,* and *D. oryzae* (Supplementary Table 2). These genomes share ANI and dDDH values of

399 94.47-94.62% and 57-57.5%, respectively, with the genomes of the three *D. parazeae* strains

400 S31^T, Ech 586 and A586-S18-A17. Comparison with the genomes of ten *D. zeae* strains

401 (NCPPB2538^T, JZL7, PL65, CE1, MS1, MS2014, MS2018, MS2, MK19, NCPPB3532) showed

402 ANI and dDDH values ranging from 94.34-94.79% and 56.7-59%, respectively (Supplementary

403 Table 2). The ANI and dDDH values between the genomes of the seven pineapple strains with

404 the genome of the *D. oryzae* type strain ZYY5^T range from 95.93-96% and 66.4-66.5%,

405 respectively. Additional strains reported to be members of, or close to, *D. oryzae* were also

406 evaluated for ANI and dDDH values with the seven pineapple strains. As described by

407 Hugouvieux-Cotte-Pattat *et al*. [5], the strains ZYY5^T, DZ2Q, ZYU1202, and EC1 were reported

408 to be more closely related to each other than to EC2 (all five strains were isolated from rice). The

409 calculated ANI and dDDH values were 95.97-96% and 66.4-66.9%, respectively, with the

410 genomes of EC1, ZYU1202, or DZ2Q, and reached 96.2-96.4% and of 69.1-69.5%, respectively,

411 with the EC2 genome (Supplementary Table 2). The ANI values were at the boundary for

412 species delineation, ranging from 96.1-96.6% with other strains of *D. oryzae*, such as S20,

413 CSLRW192, and FGV03 (all isolated from water), and NCPPB 3531 (isolated from potato).

414 dDDH values range from 68.2-69.5% with strains S20, CSLRW192, and FGV03, and 66.9-

415 70.3% with NCPPB 3531. These values, at the limit of those recommended for the delimitation

416 of species, show the difficulty of using only ANI and dDDH data for the description of species.

417 The ANI values are in the same range as those reported for the recent description of the species

418 *P. parvum* and *P. quasiaquaticum* [63,64]. However, analysis of the alignment coverage with the

419 genomes of *D. oryzae* strains S20, CSLRW192, FGV03, and NCPPB 3531 showed AP values

420 (85.91-89.56%) lower than those observed with other *Dickeya* species, such as *D. zeae* and *D.*

421 *parazeae* (87-87.2 %) (Supplementary Table 2, Figure 3). In comparison, the genomes of the

422 seven pineapple strains share a high alignment coverage percentage (97.03-99.42%), indicating

423 close relatedness within them. Therefore, the analysis of ANI, dDDH, and AP values is

424 consistent with the proposition of a novel species within the *D. zeae* complex, borderline with

425 some *D. oryzae* strains.

gapA, 16SrRNA, MLSA, whole genome and pan-genome phylogeny

427 The gene *gapA* (encoding glyceraldehyde-3-phosphate dehydrogenase) has been routinely used

428 for the identification and accurate phylogenetic positioning of strains within the genus *Dickeya*

429 [36]. Amplification of the *gapA* gene sequence confirmed that the seven pineapple strains belong

430 to the genus *Dickeya*. Since the strains were closely grouped with members of the *D. zeae*

431 complex, *gapA* sequences of the type strains of *D. zeae, D. parazeae, D. oryzae,* along with other

432 strains available in the NCBI database, were extracted (Supplementary Table 3), aligned, and

433 trimmed to 702 bp of the *gapA* gene region using Geneious Prime. This alignment was used to

434 build a phylogenetic tree using the maximum likelihood method based on the Tamura-Nei model

435 to infer the phylogenetic position. The *gapA* phylogenetic tree illustrates that the seven pineapple

436 strains A5410[T], A5391, CFBP 1278 (A5417), CFBP 1272 (A5421), A6136, A6137, and A5611

437 cluster together to form a distinct monophyletic clade, strongly supported by a high bootstrap

438 percentage (Figure 4). The strains CFBP 1278 and CFBP 1272 (previously submitted as *D.*

439 *zeae*), isolated from pineapple in Malaysia, also grouped within this monophyletic clade,

440 consistent with previous MLSA studies by Marrero *et al*. [65], indicating that the strains isolated

441 from pineapple may constitute a novel genomic species. The corresponding clade is positioned

442 between *D. parazeae* and *D. oryzae*, closely related to *D. oryzae* but clearly distinct from it and

443 other members of the genus [5], supporting the delineation of a new species.

444 The BLAST search in NCBI nearly full length of 16S rRNA gene sequence of strain A5410[T]

445 showed the highest sequence similarity of 98.57% with *D. oryzae* ZYY5[T] followed by 97.80%

446 with S31[T] *D. parazeae* and 97.05% with *D. zeae* NCPPB 2538[T]. The partial 16S rRNA sequence

447 of the proposed type strain A5410[T] was deposited in the GenBank nucleotide database with the

448 accession number OR271567, and the partial sequences of the 16S rRNA amplified using the

449 27F and 1492R primers [34] are provided in Supplementary File 1. The partial amplified 16S

450 rRNA gene sequences of the *D. ananae* sp. nov. strains were further mapped to the whole

451 genome of each sequenced strain, resulting in a 100% match. The 16S rRNA phylogenetic tree,

452 encompassing 50 strains (including type strains within the genus *Dickeya,* other *Dickeya* sp.

453 complex strains, strains listed as *Dickeya* sp., and *P. carotovorum* as an outgroup;

454 Supplementary Table 3), was inferred through the Type (Strain) Genome Server (TYGS), a high-

455 throughput platform for accurate genome-based taxonomy (https://tygs.dsmz.de). Utilizing the

456 GBDP method and full-length 16S rRNA gene sequences without ambiguities, the phylogeny

457 separated the *D. zeae* species complex into two clades: one clade includes the strains of the *D.*

458 *oryzae, D. parazeae*, and *D. zeae* species, and another clade includes the seven *D. ananae* sp.

459 nov. strains which clustered together as a monophyletic clade. This 16S rRNA clustering also

460 supports that the novel species is distinct from the other members of the *D. zeae* complex

461 (Supplementary Figure 1).

462 To further refine the taxonomic position of A5410^T, MLSA based on ten housekeeping genes

463 was conducted, and the phylogenetic positions of the strains were inferred using the Maximum

464 Likelihood method and the Tamura-Nei model with 1,000 bootstraps. The nucleotide sequences

465 of the ten housekeeping genes were retrieved from the NCBI Genome database, encompassing

466 50 strains, including each *Dickeya* species and *P. carotovorum* (used as an outgroup). Each gene

467 sequence was trimmed and assembled to generate a consensus sequence. The ten housekeeping

468 genes comprised *atpD* (1,369 bp), *dnaA* (1,325 bp), *dnaX* (2,058 bp), *fusA* (2,110 bp), *gapA*

469 (1,073 bp), *gyrB* (2,490 bp), *infB* (2,722 bp), *recA* (1,083 bp), *recN* (1,655 bp), and *rplB* (807

470 bp), resulting in a total concatenated sequence length of 16,692 bp. This analysis clearly

471 separated *D. ananae* sp. nov. strains into a distinct clade with a high bootstrap percentage (Figure

472 5). The topology of the ML tree clearly showed that the *D. zeae* species complex is divided into

473 two major clades: the first clade includes members from *D. oryzae* and *D. ananae* sp. nov., and

474 the second clade comprises members of *D. zeae* and *D. parazeae*. The stability of the

475 phylogenetic tree topology was well supported by high bootstrap values, further supporting the

476 clustering of *D. ananae* sp. nov. members in a monophyletic clade.

477 The whole-genome phylogenetic tree was created using the TYGS webserver based on the

478 Genome Blast Distance Phylogeny (GBDP) distance formula d_5 approach, with *P. carotovorum*

16

479   included as an outgroup (Figure 6). This analysis clearly separated the strains of the *D. zeae*

480   complex into two main clades: Clade I, which branched into two subclades corresponding to *D.*

481   *oryzae* strains and *D. ananae* sp. nov. strains, respectively, and Clade II, which branched into

482   three subclades corresponding to *D. zeae, D. parazeae,* and potential other species, with high

483   bootstrap values supporting the tree topology. Thus, on the whole-genome phylogenetic tree,

484   strains of *D. ananae* sp. nov. clearly group separately as a distinct monophyletic clade (Figure 6).

485   The strains of *D. ananae* sp. nov. form a distinct and separate clade, closely related to *D. oryzae*

486   and its members, which are divided into two clades in all phylogenetic analyses performed in this

487   study, consistent with the description of the *D. zeae* complex by Hugouvieux-Cotte-Pattat *et al.*

488   [5].

489   A pan-genome dendrogram of the *D. ananae* sp. nov. strains, along with other *Dickeya* species,

490   was constructed using Roary, based on the prevalence of different gene categories, including

491   core genes (shared by 99–100% of the strains), soft core genes (shared by 95–99% of the

492   strains), shell genes (shared by 15–95% of the strains), and cloud genes (shared by 0–15% of the

493   strains). The strains in genus *Dickeya*, were found to contain 553 core, 73 soft core, 5,218 shell,

494   and 19,068 cloud genes (Figure 7). The pan-genome dendrogram revealed that all seven strains

495   of *D. ananae* were grouped into a single clade with a high bootstrap support value of 100%,

496   showing significant species delineation and consistent placement of the strains, further

497   corroborating previous phylogenetic analyses and supporting the proposal of *D. ananae* as a new

498   species. The Roary matrix, derived from the analysis of 14,912 genes, indicates the diversity

499   within the genus *Dickeya* (Figure 7). Additionally, A5410^T had 312 unique genes when

500   compared to other type strains in the genus (Figure 8). The monophyly of *D. ananae* sp. nov.

501   strains within the *D. zeae* complex are consistent and congruent across the phylogenetic and

502   phylogenomic evaluations, with strong support from high bootstrap values, confirming the

503   assignment of these strains as a novel species within the cluster.

Phenotypic, enzymatic and biochemical characterization

505   Electron microscopic investigations revealed that A5410^T exhibited a rod-shaped morphology.

506   The cells of *D. ananae* sp. nov. measured approximately 2.91 ± 0.02 μm in length and 1 ± 0.07

507   μm in width, with peritrichous flagella (Figure 9), a characteristic trait observed in soft rot

17

508   *Pectobacteriaceae* [66]. Motility assays using swimming and swarming methods showed that all

509   strains exhibited swimming motility. The strains isolated from Hawaii demonstrated similar

510   motility, with radii ranging from 5.35 ± 0.9 to 6.7 cm. However, the strains from Malaysia,

511   CFBP 1278 (A5417) and CFBP 1272 (A5421), exhibited significantly weaker motility, with radii

512   of 0.6 cm and 2.5 ± 0.56 cm, respectively, compared to the strains isolated from pineapples in

513   Oahu. The macerating ability of these strains, previously demonstrated on pineapple, showed

514   that all strains of *D. ananae* sp. nov. were capable of causing maceration of pineapple leaves

515   [31]. All strains, including those isolated from Malaysia, were able to utilize pectate, as

516   evidenced by pit formation on crystal violet pectate medium. Additional maceration tests on

517   potato, taro, and onion slices revealed that all strains could macerate potato and taro, but not

518   onion slices. Notably, all strains displayed stronger macerating activity on potato compared to

519   taro slices (data not shown).

520   Using Biolog GENIII plates, the *D. ananae* sp. nov. members responded positively to the

521   following carbon sources: sucrose, D-raffinose (with A5611, A5391, and CFBP 1272 showing

522   weak positivity), D-melibiose, β-methyl-D-glucoside, D-salicin, N-acetyl-D-glucosamine, D-

523   glucose, D-mannose, D-fructose, D-galactose, D-mannitol, and myo-inositol (except for CFBP

524   1278 and CFBP 1272) (Supplementary Table 4). Other positive responses included glycerol, D-

525   glucose-6-PO$_4$, D-fructose-6-PO$_4$, D-aspartic acid (with CFBP 1278 weakly positive), L-aspartic

526   acid, L-serine, pectin, D-galacturonic acid, L-galactonic acid lactone, D-gluconic acid, D-

527   glucuronic acid, mucic acid, D-saccharic acid, L-lactic acid, citric acid, D-malic acid, L-malic

528   acid, and bromo-succinic acid. All strains tolerated some inhibitory conditions, including pH 6,

529   1% NaCl, 1% sodium lactate, troleandomycin, rifamycin SV, lincomycin, Niaproof 4,

530   vancomycin, tetrazolium violet, and tetrazolium blue. The strains showed inhibition to 4% NaCl,

531   8% NaCl, D-serine, troleandomycin, minocycline, nalidixic acid, potassium tellurite (except

532   A5410^T, which weakly tolerated it), lithium chloride (except A6136 and A5391, which weakly

533   tolerated it), aztreonam, and sodium bromate. (Supplementary Table 4). The type strain A5410^T

534   was unable to utilize the following carbon sources: D-serine, p-hydroxy-phenylacetic acid, D-

535   lactic acid methyl ester, α-hydroxy-butyric acid, β-hydroxy-butyric acid, α-keto-butyric acid, and

536   propionic acid. Some differences were noticed between the type strain A5410^T and the other type

537   strains of the *D. zeae* cluster. In comparison to the strains *D. oryzae* NCPPB 3531^T and *D.*

18

538 *parazeae* S31[T], A5410[T] was able to weakly utilize D-lactose, N-acetyl-D-galactosamine, and

539 inosine (Supplementary Table 4). Moreover, unlike the strains *D. oryzae* NCPPB 3531[T], *D.*

540 *parazeae* S31[T], and *D. zeae* NCPPB 2538[T], *D. ananae* A5410[T] exhibited weak growth at pH 5

541 and was inhibited by troleandomycin (Supplementary Table 4).

542 The API 20E kit is specifically designed for identifying *Enterobacteriaceae* and other non-

543 fastidious Gram-negative rods. It includes tests suitable for both genus-level identification and

544 species characterization. The seven *D. ananae* strains were positive for indole production and

545 acetoin production (Voges-Proskauer test), were able to utilize citrate, and exhibited β-

546 galactosidase and gelatinase activity (Table 2). They were also able to ferment D-glucose, D-

547 mannitol, L-rhamnose, D-saccharose, amygdalin, L-arabinose, and D-melibiose (except for

548 A5391, which was unable to ferment D-melibiose). All strains were negative for the other tests,

549 including H_2S production, urease, arginine dihydrolase, lysine decarboxylase, ornithine

550 decarboxylase (ODC), tryptophan deaminase. They were unable to ferment inositol and D-

551 sorbitol. Nitrate reduction to nitrite was observed for all strains. The *D. ananae* strains sequenced

552 in this study were negative for oxidase but were positive for gelatin hydrolysis, in contrast to *D.*

553 *oryzae* strains, as reported by Wang *et al*. (2020) (Table 2).

554 Enzymatic reactions tested with the API ZYM kit revealed that *D. ananae* strains exhibit activity

555 for the enzymes alkaline phosphatase, esterase (C4), leucine arylamidase, naphthol-AS-Bi-

556 phosphohydrolase, β-galactosidase, and β-glucosidase. They were negative for esterase lipase

557 (C8), valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsin, α-galactosidase, β-

558 glucuronidase, α-glucosidase, N-acetyl-β-glucosaminidase, α-mannosidase, and α-fucosidase

559 (Table 2). The API 50 CH kit is designed for studying the metabolism of carbohydrates and

560 carbohydrate derivatives, while the API 50 CHB/E medium is specifically designed for use with

561 *Enterobacteriaceae*. All strains of *D. ananae* sp. nov. were able to produce acids from glycerol,

562 D-arabinose, L-arabinose, D-ribose, D-xylose, D-galactose, D-glucose, D-fructose, D-mannose,

563 L-rhamnose, D-mannitol, N-acetylglucosamine, arbutin, esculin ferric citrate, salicin, D-

564 cellobiose, D-melibiose, D-sucrose. The seven strains showed a weak response for acid

565 production with potassium gluconate. These strains showed variable responses in acid production

566 from inositol (positive or weak), raffinose (positive or weak) and lactose (positive, weak, or

567 negative) (Supplementary Table 5). All strains of *D. ananae* sp. nov were not able to produce

19

568   acids from erythritol, L-xylose, D-adonitol, methyl-ß-D-xylopyranoside, L-sorbose, dulcitol, D-

569   sorbitol, methyl-α-D-mannopyranoside, methyl-α-D-glucopyranoside, amygdalin, D-maltose, D-

570   trehalose, inulin, D-melezitose, starch (amidon), glycogen, xylitol, gentiobiose, D-turanose, D-

571   lyxose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium 2-ketogluconate and

572   potassium 5-ketogluconate. When compared to *D. oryzae* ZYY5^T, differences were observed in

573   the production of acids from D-galactose, D-glucose, D-fructose, D-mannitol, N-

574   acetylglucosamine, and D-sucrose, as ZYY5^T was reported to be unable to produce acids from

575   these carbon sources [26]. In contrast, strains of *D. ananae* sp. nov. were unable to produce acids

576   from L-sorbose, unlike *D. oryzae* ZYY5^T, which has been reported to produce acid from this

577   carbohydrate [26]. Additionally, *D. ananae* sp. nov. could not produce acid from potassium 2-

578   ketogluconate, similar to *D. oryzae* ZYY5^T, but differing from the *D. zeae* type strain, which was

579   capable of acid production from this carbohydrate [26].

580   Pectate lyases, cellulases, and proteases are important extracellular enzymes playing critical role

581   in *Dickeya* pathogenesis. All strains of *D. ananae* sp. nov. possess pectinase activity, with

582   A5410^T, A6136, and A6137 exhibiting the highest levels of activity followed by A5391 and

583   A5611. In contrast, strains CFBP 1278, and CFBP 1272 displayed weaker activity, as determined

584   by semi-quantitative assays. Similarly, for cellulase activity, A5410^T, A6136, and A6137 showed

585   the highest activity, followed by A5391, A5611, CFBP 1272, and CFBP 1278, which exhibited

586   the least activity. Regarding protease activity, *D. ananae* sp. nov. strains A5410^T, A6136,

587   A6137, A5391, and A5611 demonstrated comparable protease activity, while the Malaysian

588   strains CFBP 1272 and CFBP 1278 showed only very weak activity after 30 hours of incubation

589   (Supplementary Table 4).

590   Various degrees of sensitivity were observed across the six major classes of antibiotics tested,

591   which included a total of 13 antibiotics (Supplementary Table 6). For the aminoglycoside class,

592   the seven *D. ananae* strains exhibited similar responses, showing sensitivity to spectinomycin

593   (100 mg/ml), gentamicin (50 mg/ml), and kanamycin (50 mg/ml). All strains showed equal

594   sensitivity to chloramphenicol (50 mg/ml), trimethoprim (25 mg/ml), and carbenicillin (50

595   mg/ml). A5611 showed the highest sensitivity to vancomycin (50 mg/ml) with an inhibition zone

596   of 1.85 cm. Differences in sensitivity to bacitracin (50 mg/ml), hygromycin B (50 mg/ml), and

597   tetracycline (40 mg/ml) were noted among the strains In the β-lactam class, CFBP 1278 showed

598   the highest sensitivity to penicillin (50 mg/ml) with a 2.15 cm inhibition zone, while A5410^T

599   demonstrated the greatest sensitivity to cefalexin (50 mg/ml), with a 2.1 cm inhibition zone. All

600   strains exhibited similar sensitivity to cefotaxime (50 mg/ml), with inhibition zones ranging from

601   2.4 to 3 cm.

602   In conclusion, based on the results of this polyphasic study, strain A5410^T is proposed to

603   represent a novel species in the genus *Dickeya*, named *Dickeya ananae* sp. nov.

604   **Description of *Dickeya ananae* sp. nov. (a.na.nae. L. F. gen. n. *ananas* isolated from**

605   **pineapple)**

606   *Dickeya ananae (*a.na.nae. L. F. gen. n. *ananas* referring to the isolation source of the type strain,

607   pineapple).

608   *Dickeya ananae* is a Gram-negative, rod-shaped, motile bacterium measuring 2.91 ± 0.02 μm in

609   length, with peritrichous flagella. It is fermentative and grows optimally at 28 °C on nutrient

610   agar medium supplemented with 0.4% glucose and on King's B medium, forming milky white

611   colonies within 24 hours. This bacterium forms pits on crystal violet pectate medium within 24-

612   48 hours [33] and causes maceration of potato tubers, taro slices, and pineapple leaves. *Dickeya*

613   *ananae* sp. nov. is catalase positive, positive for nitrate reduction, oxidase negative and

614   pectinolytic. *D. ananae* sp. nov. is able to utilize sucrose, D-raffinose, D. melibiose, β-methyl-D-

615   glucoside, D-salicine, N-acetyl-D-glucosamine, D-glucose, D-mannose, D-fructose, D-galactose,

616   D-sorbitol, D-mannitol, myo-inositol, glycerol, D-glucose-6-PO$_4$, D-fructose-6-PO$_4$, D-aspartic

617   acid, L-aspartic acid, L-glutamic acid, L-serine, pectin, D-galacturonic acid, L-galactonic acid

618   lactone, D-gluconic acid, D-glucuronic acid, methyl pyruvate, L-lactic acid, citric acid, D-malic

619   acid, L-malic acid, bromo-succinic acid, acetic acid and formic acid as a carbon sources. It

620   shows inhibition at pH 5, or in the presence of 8 % NaCl, D-serine, minocycline, nalidixic acid,

621   potassium tellurite, aztreonam and sodium bromate. The type strain is unable to utilize the

622   following carbon sources: dextrin, stachyose, N-acetyl neuraminic acid, D-serine, quinic acid, p-

623   hydroxy-phenylacetic acid, D-lactic acid methyl ester, tween 40, -amino-butryric acid, α-

624   hydroxy-butyric acid, β-hydroxy-D, L-butyric acid, α -keto-butyric acid and propionic acid.

625   Enzymatic reactions tested with the API ZYM kit give positive results for alkaline phosphatase,

626 esterase, leucine arylamidase, acid phosphatase, naphthol-AS-BI-phosphohydrolase, β-
627 galactosidase, and β-glucosidase. The type strain is able to utilize citrate, produces indole and
628 acetoin and able to hydrolyze esculin; it is negative for urease. The API 50CH kit gave positive
629 results for the utilization of glycerol, D-arabinose, L-arabinose, D-ribose, D-xylose, D-galactose,
630 D-glucose, D-fructose, D-mannose, L-rhamnose, D-mannitol, *N*-acetylglucosamine, arbutin,
631 aesculin ferric citrate, salicin, lactose, melibiose, sucrose, raffinose, and weakly positive for
632 inositol and potassium gluconate. Assay for antibiotic susceptibility showed that *D. ananae*
633 A5410^T is sensitive to 13 antibiotics including trimethoprim, kanamycin, bacitracin, cephalexin,
634 hygromycin B, vancomycin, cefotaxime, tetracycline, spectinomycin, chloramphenicol,
635 carbenecillin, penicillin and gentamicin. The highest sensitivity was observed for the cefotaxime,
636 carbenicillin and cefalexins. The *D. ananae* type strain is A5410^T (=ICMP 25020^T =LMG
637 33197^T) isolated from Pineapple in Hawaii. Its genome size of 4.77 Mbp with a G+C content of
638 53.5 mol% based on the genome sequence. Other isolates of this species were also obtained from
639 pineapple in Hawaii (A5391, A6136, A6137, and A5611), and two strains, CFBP 1272 and
640 CFBP 1278, were isolated from pineapple in Malaysia.

22

654   Genome Center DNA Technologies and Expression Analysis Core, supported by NIH Shared

655   Instrumentation Grant 1S10OD010786-01.

Conflicts of interest

657   The authors declare that there are no conflicts of interest.

658

659

660

Legends for Tables

662   **Table 1**: Genome features and Statistics for the six strains of *Dickeya ananae* sp. nov. other
663   strains sequenced in this study. The genomic features as annotated were assessed by The
664   Bacterial and Viral Bioinformatics Resource Center (BV-BRC version 3.34.11).

665   **Table 2:** Biochemical and enzymatic characterization of tested strains in *D. ananae* sp. nov.
666   using API 20E and API ZYM tests strips.

Legends for supplementary Figures and Tables

668   **Supplementary Figure 1.** Maximum-likelihood phylogenetic tree based on 16S rRNA
669   sequences comparison showing the position of *D. ananae* sp. nov. A5410^{T} within the genus
670   *Dickeya.* The *Pectobacterium carotovorum* NCPPB 312 was taken as an outgroup**.** The
671   phylogenomic tree based on 16S rRNA genome sequences in the TYGS tree inferred with
672   FastME 2.1.6.1 (Meier-Kolthof and Göker 2019) from the Genome BLAST Distance Phylogeny
673   approach (GBDP); distances were calculated from genome sequences. The phylogenetic trees
674   with bootstrapping value and colored annotations were created using web-based tool Interactive
675   Tree Of Life (iTOL v6). The branch lengths are scaled in terms of GBDP distance formula d5.
676   The numbers above the branches are GBDP pseudo-bootstrap support values>60% from 100
677   replications. The tree was rooted at the midpoint.

678   **Supplementary Table 1.** Subsystem Annotation summary of 27 categories of subsystem
679   features as predicted by RAST web server for new sp. A5410^{T} and the other genomes used in
680   this study.

681   **Supplementary Table 2**. Combined Average nucleotide identity (ANI) (above diagonal)
682   calculated using the OrthoANI algorithm implemented in the software OAT and digital DNA-
683   DNA hybridization (dDDH) determined using Genome-to-Genome Distance Calculator version
684   3.0 (https://ggdc.dsmz.de/) (below diagonal) for all strains grouping in *D. ananae* sp. nov and
685   other strains from NCBI database belonging to the *D. zeae* complex. Combined Average
686   nucleotide identity (ANI) (above diagonal) and alignment percentage (AP) (below diagonal)

calculated using CLC Genomics workbench 22.0.2 for all strains grouping in *D. ananae* sp. nov and other strains from NCBI database belonging to the *D. zeae* species complex. Da, *D. ananae*; Dor, *D. oryzae*; Dz, *D. zeae;* Dpz, *D. parazeae*.

Supplementary Table 3. List of *Dickeya* strains of and *Pectobacterium carotovorum* (outgroup) used in the *gapA,* 16S rRNA, MLSA, and whole genome phylogenetic analyses.

Supplementary Table 4. BIOLOG GEN III and other phenotypic profiles of *D. ananae* sp. nov. compared to *D. zeae*, *D. oryzae* and *D. parazeae*.

Supplementary Table 5. Fermentation of different carbon sugar by *D. ananae* sp. nov. determined using API 50 CH strips.

Supplementary Table 6. Antibiotic sensitivity assays of *D. ananae* sp. nov. against a panel of 13 different antibiotics.

References

1. Ma B, Hibbing ME, Kim HS, Reedy RM, Yedidia I, *et al*. Host range and molecular phylogenies of the soft rot enterobacterial genera *Pectobacterium* and *Dickeya*. *Phytopathol* 2007; 97:1150-1163.

2. Charkowski A, Blanco C, Condemine G, Expert D, Franza T, *et al*. The role of secretion systems and small molecules in soft-rot *Enterobacteriaceae* pathogenicity. *Annu Rev Phytopathol*. 2012; 50:425-49. doi: 10.1146/annurev-phyto-081211-173013

3. Parkinson N, DeVos P, Pirhonen M, Elphinstone J. *Dickeya aquatica* sp. nov., isolated from waterways. *Int J Syst Evol Microbiol* 2014;64: 2264–2266.

4. Hugouvieux-Cotte-Pattat N, Jacot-des-Combes C, Briolay J. *Dickeya lacustris* sp. nov., a water-living pectinolytic bacterium isolated from lakes in France. *Int J Syst Evol Microbiol* 2019;69:721-726.

5. Hugouvieux-Cotte-Pattat N, Van Gijsegem F. Diversity within the *Dickeya zeae* complex, identification of *Dickeya zeae* and *Dickeya oryzae* members, proposal of the novel species *Dickeya parazeae* sp. nov. *Int J Syst Evol Microbiol* 2021;71(11). doi: 10.1099/ijsem.0.005059.

6. Hugouvieux-Cotte-Pattat N, Pédron J and Van Gijsegem F. Insight into biodiversity of the recently rearranged genus *Dickeya*. *Front Plant Sci* 2023;14:1168480. doi: 10.3389/fpls.2023.1168480.

7. Oulghazi S, Pédron J, Cigna J, Lau YY, Moumni M, *et al. Dickeya undicola* sp. nov., a novel species for pectinolytic isolates from surface waters in Europe and Asia. *Int J Syst Evol Microbiol* 2019;69:2440-2444.

8. Arif M, Czajkowski R, Chapman TA. Editorial: Genome-Wide Analyses of *Pectobacterium* and *Dickeya* Species. *Front. Plant Sci* 2022;13:855262.

9. Hugouvieux-Cotte-Pattat, N, Condemine, G, Shevchik VE. Bacterial pectate lyases, structural and functional diversity. *Environ Microbiol Rep* 2014;6:427–440.

10. Boluk G, Arizala D, Dobhal S, Zhang J, Hu J, *et al*. Genomic and phenotypic biology of novel strains of *Dickeya zeae* isolated from pineapple and taro in Hawaii: insights into

genome plasticity, pathogenicity, and virulence determinants. *Front Plant Sci* 2021;12:663851. doi: 10.3389/fpls.2021.663851

11. Reverchon S, Muskhelisvili G, Nasser W. Virulence program of a bacterial plant pathogen: the *Dickeya* model. *Prog Mol Biol Transl Sci* 2016;142:51–92. doi: 10.1016/bs.pmbts.2016.05.005

12. Van Gijsegem F, Hugouvieux-Cotte-Pattat N, Kraepiel Y, Lojkowska E, Moleleki LN, Gorshkov V, *et al*. "Molecular interactions of *Pectobacterium* and *Dickeya* with plants," in Plant diseases caused by *Dickeya and Pectobacterium species*. (eds) F. Van Gijsegem, J. M. van der Wolf and I. K. Toth (Cham, Switzerland: Springer International Publishing), 2021; pp 85–147.

13. Pedron J, van der Wolf JM, Portier P, Caullireau E, Van Gijsegem F. The broad host range plant pathogen *Dickeya dianthicola* shows a high genetic diversity. *Microorganisms* 2022;10:1024. doi: 10.3390/microorganisms10051024.

14. Yang C□H, Gavilanes□Ruiz M, Okinaka Y, Vedel R, Berthuy I, Boccara, M, *et al*. *hrp* genes of *Erwinia chrysanthemi* 3937 are important virulence factors. *Mol Plant– Microbe Interact* 2002; 15:472–480.

15. Li Y, Yamazaki A, Zou L, Biddle E, Zeng Q, Wang Y, *et al*. ClpXP protease regulates the Type III secretion system of *Dickeya dadantii* 3937 and is essential for the bacterial virulence. *Mol Plant-Microbe Interact* 2010;23(7):871–878.

16. Tan X-J, Zhang Z-W, Xiao J-J, Wang W, He F, Gao X, *et al*. Genomic and phenotypic biology of a novel *Dickeya zeae* WH1 isolated from rice in China: Insights into pathogenicity and virulence factors. *Front Microbiol* 2022;13:997486. doi: 10.3389/fmicb.2022.99748.

17. Yap MN, Yang CH, Barak JD, Jahn CE, Charkowski AO. The *Erwinia chrysanthemi* type III secretion system is required for multicellular behavior. *J Bacteriol* 2005;187:639–648. doi: 10.1128/JB.187.2.639-648.2005

18. Moleleki LN, Pretorius RG, Tanui CK, Mosina G, Theron J. A quorum sensing-defective mutant of *Pectobacterium carotovorum* ssp. *brasiliense* 1692 is attenuated in virulence and unable to occlude xylem tissue of susceptible potato plant stems. *Mol Plant Pathol* 2017;18:32–44.

19. Ge T, Jiang H, Tan EH, Johnson SB, Larkin RP, Charkowski AO, *et al*. Pangenomic Analysis of *Dickeya dianthicola* Strains Related to the Outbreak of Blackleg and Soft Rot of Potato in the United States. *Plant Dis* 2021;105(12):3946–3955.

20. Samson R, Legendre JB, Christen R, Fischer-Le Saux M, Achouak W. Transfer of *Pectobacterium chrysanthemi* (Burkholder *et al.* 1953) Brenner *et al.* 1973 and *Brenneria paradisiaca* to the genus *Dickeya* gen. nov. as *Dickeya chrysanthemi* comb. nov. and *Dickeya paradisiaca* comb. nov. and delineation of four novel species, *Dickeya dadantii* sp. nov., *Dickeya dianthicola* sp. nov., *Dickeya dieffenbachiae* sp. nov. and *Dickeya zeae* sp. nov. *Int J Syst Evol Microbiol* 2005;55:1415–1427.

21. Brady CL, Cleenwerck I, Denman S, Venter SN, Rodrıguez-Palenzuela P, *et al*. Proposal to reclassify *Brenneria quercina* (Hildebrand & Schroth 1967) Hauben *et al.* 1999 into a novel genus, *Lonsdalea* gen. nov., as *Lonsdalea quercina* comb. nov., descriptions of *Lonsdalea quercina* subsp. *quercina* comb. nov., *Lonsdalea quercina* subsp. *Iberica*

subsp. nov., and *Lonsdalea quercina* subsp. *britannica* subsp. nov., emendation of the description of the genus *Brenneria*, reclassification of *Dickeya dieffenbachiae* as *Dickeya dadantii* subsp. *dieffenbachiae* comb. nov., and emendation of the description of *Dickeya dadantii*. *Int J Syst Evol Microbiol* 2012;62:1592–1602.

22. Slawiak M, van Beckhoven J, Speksnijder A, Czajkowski R, Grabe G, *et al*. Biochemical and genetical analysis reveal a new clade of biovar 3 *Dickeya* spp. strains isolated from potato in Europe. *Eur J Plant Pathol* 2009;125:245–261.

23. Toth IK, van der Wolf JM, Saddler G, Lojkowska E, Hé lias V, Pirhonen M, *et al*. *Dickeya* species: an emerging problem for potato production in Europe. *Plant Pathol* 2011;60:385–399. doi: 10.1111/j.1365-3059.2011.02427.x

24. van der Wolf JM, Nijhuis EH, Kowalewska MJ, Saddler GS, Parkinson N, *et al*. *Dickeya solani* sp. nov., a pectinolytic plantpathogenic bacterium isolated from potato (*Solanum tuberosum*). *Int J Syst Evol Microbiol* 2014;64:768–774.

25. Hugouvieux-Cotte-Pattat N, des-Combes CJ, Briolay J, Pritchard L. Proposal for the creation of a new genus *Musicola* gen. nov., reclassification of *Dickeya paradisiaca* (Samson *et al.* 2005) as *Musicola paradisiaca* comb. nov. and description of a new species *Musicola keenii* sp. nov. *Int J Syst Evol Microbiol* 2021;71:005037. doi: 10.1099/ijsem.0.005037.

26. Wang X, He SW, Guo HB, Han JG, Thin KK, Gao JS, *et al*. *Dickeya oryzae* sp. nov., isolated from the roots of rice. *Int J Syst Evol Microbiol* 2020;70:4171–4178. doi: 10.1099/ijsem.0.004265

27. Boluk, G., Dobhal, S., Arizala, D., Alvarez, A. M., Arif, M. *Dickeya colocasiae* sp. nov. isolated from wetland taro, *Colocasia esculentum*. bioRxiv 2022;doi: 10.1101/2022.01.14.476417.

28. Johnston, A. Bacterial heart rot of the pineapple. *Malay. Agric. J* 1957;40:2–8.

29. Chinchilla CM, Gonzales LC, Morales F. Purdrición bacteriana delcogollo de la piña en Costa Rica. *Agron Co-starric* 1979;3:183–185.

30. Rohrbach KG, Johnson, MW. Pests, diseases and weeds. (eds). The Pineapple: Botany, Production and Uses. D. P. Bartholomew, R. E. Paull, and K. G.Rohrbach, 2003;203-251 in: CABI Publishing, Wallingford, UK.

31. Kaneshiro WS, Burger M, Vine BG, de Silva AS, Alvarez AM. Characterization of *Erwinia chrysanthemi* from a Bacterial Heart Rot of Pineapple Outbreak in Hawaii. *Plant Dis* 2008;92(10)1444–1450.

32. Norman, D, Alvarez A. A rapid method for presumptive identification of *Xanthomonas campestris* pv. *dieffenbachiae* and other xanthomonads. *Plant Dis* 1989;73(8):654–658.

33. Helias V, Hamon P, Huchet E, Wolf JVD, Andrivon D. Two new effective semiselective crystal violet pectate media for isolation of *Pectobacterium* and *Dickeya*. *Plant Pathol* 2012;61:339–345.

34. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci* 1985;82:6955–6959.

35. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;49:W293–W296. doi: 10.1093/nar/gkab301.

36. Cigna J, Dewaegeneire P, Beury A, Gobert V, Faure D. A gapA PCR sequencing assay for identifying the *Dickeya* and *Pectobacterium* potato pathogens. *Plant Dis* 2017;101:1278–1282.

37. Hugouvieux-Cotte-Pattat N, Brochier-Armanet C, Flandrois JP, Reverchon S. *Dickeya poaceaphila* sp. nov., a plant-pathogenic bacterium isolated from sugar cane (*Saccharum officinarum*). *Int. Syst Evol Microbiol* 2020;70:4508–4514.

38. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *MBE* 1993;10:512–526.

39. Tamura K, Stecher G, Kumar S. MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *MBE* 2021. https://doi.org/10.1093/molbev/msab120.

40. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, *et al*. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016;44:6614–6624.

41. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, *et al*. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 2015;10:8365.

42. Lee I, Ouk Kim Y, Park S-C, Chun J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* 2016;66:1100–1103.

43. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res* 2022;50:D801–D807. doi: 10.1093/nar/gkab902.

44. Glaeser SP, Kämpfer P. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst Appl Microbiol* 2015;38:237–245.

45. De Vos P. Multilocus Sequence Determination and Analysis. *Methods in Microbiology Academic Press* 2011;385–407.

46. Oulghazi S, Pedron J, Cigna J, Lau YY, Moumni M, Van Gijsegem F, *et al*. *Dickeya undicola* sp. nov., a novel species for pectinolytic isolates from surface waters in Europe and Asia. *Int J Syst Evo. Microbiol* 2019;69:2440–2444.

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

48. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG., *et al*. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.

49. Löytynoja A. Phylogeny-aware alignment with PRANK. *Methods Mol Biol* 2014;1079:155–170.

50. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 2019;35:4453–4455.

51. Chen Y, Lv M, Liao L, Gu Y, Liang Z, Shi Z, *et al*. Genetic modulation of c-di-GMP turnover affects multiple virulence traits and bacterial virulence in rice pathogen *Dickeya zeae*. *PLoS One* 2016;11:e0165979. 10.1371/journal.pone.0165979.

27

853 52. Rath AC, Carr CJ, Graham BR. Characterization of *Metarhizium anisopliae* strains by
854   carbohydrate utilization (API50CH). *J Invertebr Pathol* 1995;65:152–161. doi:
855   10.1006/jipa.1995.1023.

856 53. Chatterjee A, Cui Y, Liu Y, Dumenyo CK, Chatterjee AK. Inactivation of rsmA leads to
857   overproduction of extracellular pectinases, cellulases, and proteases in *Erwinia*
858   *carotovora* subsp. carotovora in the absence of the starvation/cell density-sensing signal,
859   N-(3-oxohexanoyl)-L-homoserine lactone. *Appl Environ Microbiol* 1995;(61)1959–1967.

860 54. Jacobson MR, Brigle KE, Bennett LT, *et al*. Physical and genetic map of the major *nif*
861   gene cluster from *Azotobacter vinelandii*. *J Bacteriol* 1989;171(2):1017–1027.

862 55. Oh CJ, Kim HB, An CS. Molecular cloning and complementation analysis of *nifV* gene
863   from *Frankia* EuIK1 strain. *Mol Cell* 2003;15:27–33.

864 56. Rubio LM, Ludden PW. Biosynthesis of the iron-molybdenum cofactor of nitrogenase.
865   *Annu Rev Microbiol* 2008;62:93-111.

866 57. Poudel S, Colman DR, Fixen KR, Ledbetter RN, Zheng Y, Pence N, *et al*. Electron
867   transfer to Nitrogenase in different genomic and metabolic backgrounds. *J Bacteriol*
868   2018;200(10):e00757-e00717.

869 58. Nonaka A, Yamamoto H, Kamiya N, Kotani H, Yamakawa H, Tsujimoto R, Fujita Y.
870   Accessory Proteins of the Nitrogenase Assembly, *NifW*, *NifX/NafY*, and *NifZ*, Are
871   Essential for diazotrophic growth in the nonheterocystous Cyanobacterium *Leptolyngbya*
872   *boryana*. *Front Microbiol* 2019; 10:495. doi: 10.3389/fmicb.2019.00495.

873 59. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-
874   DNA hybridization values and their relationship to whole-genome sequence similarities.
875   *Int J Syst Evol Microbiol* 2007;57(1):81–91.

876 60. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic
877   species definition. *Proc Natl Acad Sci* 2009;19126–19131.

878 61. Meier-Kolthoff JP, Auch AF, Klenk H-P, Goeker M. Genome sequence-based species
879   delimitation with confidence intervals and improved distance functions. *BMC Bioinform*
880   2013;14(1):60.

881 62. Chun J,   Oren A, Ventosa A, Christensen H, Arahal DR, *et   al*. Proposed   minimal
882   standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol*
883   *Microbiol* 2018;68:461–466.

884 63. Pasanen M, Waleron M, Schott T, Cleenwerck I, Misztak A, Waleron K, *et*
885   *al*. *Pectobacterium parvum* sp. nov., having a *Salmonella* SPI-1-like Type III secretion
886   system and low virulence. *Int J Syst Evol Microbiol* 2020;70:2440–2448.

887 64. Ben Moussa H, Pédron J, Bertrand C, Hecquet A, Barny MA. *Pectobacterium*
888   *quasiaquaticum* sp. nov., isolated from waterways. *Int J Syst Evol Microbiol*
889   2021;71(10).

890 65. Marrero G, Schneider KL, Jenkins DM, Alvarez AM. Phylogeny and classification of
891   *Dickeya* based on multilocus sequence analysis. *Int J Syst Evol Microbiol*
892   2013;63(9):3524–3539.

893 66. Rungnapha K, Yu SH, Xie GL. Bacterial stem rot of poinsettia caused by a *Dickeya* sp.
894   (*Pectobacterium chrysanthemi*) in China. *Plant Dis* 2008;92:1135.

895

896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913

Table 1: Genome features and Statistics for the six strains of *Dickeya ananae* sp. nov other strains sequenced in this study. The genomic features as annotated were assessed by The Bacterial and Viral Bioinformatics Resource Center (BV-BRC version 3.34.11).

Genome Statistics	Strains						
	A5410[T*]	A5391	CFBP 1278 (A5417)	CFBP 1272 (A5421)	A6136	A6137	A5611
GenBank accession number	CP040816.1	CP146223	CP146227	CP146228	CP146224	CP146225	CP146226
Biosample	SAMN11855518	SAMN39956632	SAMN39957828	SAMN39957829	SAMN39956949	SAMN39956997	SAMN39956998
Bioproject	PRJNA544764	PRJNA1076913	PRJNA1076913	PRJNA1076913	PRJNA1076913	PRJNA1076913	PRJNA1076913
Assembly level	complete	complete	complete	complete	complete	complete	complete
Genome Sequencing Platform	PacBio RSII	Hybrid (MinIon+ illumina Novaseq)	Hybrid (MinIon+ illumina Novaseq)	Hybrid (MinIon+ illumina Novaseq)	Hybrid (MinIon+ illumina Novaseq)	Hybrid (MinIon+ illumina Novaseq)	Hybrid (MinIon+ illumina Novaseq)
Genome coverage (x)	558.0x	364.20x	476.99x	336.798x	469.897x	459.07x	545.335x
Contigs	1	1	1	1	1	1	1
Genome Length (bp)	4779199	4834333	4805951	4797088	4779825	4779844	4837911
GC Content (mol%)	53.481453	53.492363	53.52932	53.5207	53.48164	53.48093	53.56957
Contig L50	1	1	1	1	1	1	1
Contig N50 length (bp)	4779199	4834333	4805951	4797088	4779825	4779844	4837911
Annotation Statistics							
No. of transfer RNAs (tRNAs)	75	75	76	76	75	75	75
No. of ribosomal RNAs (rRNAs)	22	22	22	22	22	22	22
CDS	4555	4728	4658	4629	4692	4708	4756
CDS Ratio	0.9530886	0.97800463	0.969215	0.96496	0.981626	0.984969	0.983069
Hypothetical CDS	884	911	876	877	898	917	907
Hypothetical CDS Ratio	0.29330406	0.29483926	0.287033	0.289263	0.290494	0.293968	0.291421
PLFAM CDS	4096	4219	4194	4164	4214	4225	4229
PLFAM CDS Ratio	0.8992316	0.89234346	0.900386	0.899546	0.898124	0.897409	0.889193
CRISPR_array	3	3	2	2	3	3	3
Genome Quality							

Coarse Consistency	99.6	99.5	99.3	99.6	99.4	99.6	99.5
Fine Consistency	99.1	96.6	97	97.1	96.8	96.8	96.4
CheckM Completeness	100	100	100	99.9	99.9	100	100
CheckM Contamination	0.3	3.1	4	2.2	2.5	4	3.6
Genome Quality	Good	Good	Good	Good	Good	Good	Good

*Previously sequenced and reported in Boluk et al., 2021

Table 2: Biochemical and enzymatic characterization of tested strains in *D. ananae* sp. nov. using API 20E and API ZYM tests strips.

Assays kit	Test	Strains						
		A5410 (Type Strain)	A6136	A6137	A5611	A5391	A5417 (CFBP1278)	A5421 (CFBP1272)
API 20 E	β-Galactosidase (Ortho nitrophenyl-βD-galactopyranosidase, ONPG)	Yellow (+)	+	+	+	+	+	+
	Arginine dihydrolase (ADH)	Yellow (-)	-	-	-	-	-	-
	Lysine decarboxylase (LDC)	Yellow (-)	-	-	-	-	-	-
	Ornithine decarboxylase (ODC)	Yellow (-)	-	-	-	-	-	-
	Citrate utilization (CIT)	Blue (+)	+	+	+	+	+	+
	H_2S production (H_2S)	colorless (-)	-	-	-	-	-	-
	Urease (URE)	Yellow (-)	-	-	-	-	-	-
	Tryptophan deaminase (TDA)	Brownish yellow (-)	-	-	-	-	-	-
	Indole production (IND)	Pink (+)	+	+	+	+	+	+
	Acetoin production (Voges Proskauer, VP)	Pink (+)	+	+	+	+	+	+
	Gelatinase (GEL)	Diffusion of Black pigment (+)	+	+	+	+	+	+
	D-Glucose (GLU)	yellowish green (+)	+	+	+	+	+	+
	D-Mannitol (MAN)	Yellow (+)	+	+	+	+	+	+
	Inositol (INO)	Blue (-)	-	-	-	-	-	-
	D-Sorbitol (SOR)	Blue (-)	-	-	-	-	-	-
	L-Rhamnose (RHA)	Yellow (+)	Blue green (-)	+	+	+	+	+
	D-Saccharose (sucrose; SAC)	Yellow (+)	+	+	+	+	+	+
	D-Melibiose (MEL)	Yellow (+)	+	+	+	Blue green (-)	+	+
	Amygdalin (AMY)	Yellow (+)	+	+	+	+	+	+
	L-Arabinose (ARA)	Yellow (+)	+	+	+	+	+	+
	Nitrate Reduction (NIT) Reduction of Nitrates to No2	Red (+)	+	+	+	+	+	+
API ZYM	Alkaline phosphatase	Violet (+)	+	+	+	+	+	+
	Esterase (C 4)	Violet (+)	+	+	+	+	+	+

	Esterase (C 8)	Colorless (-)	-	-	-	-	-	-
	Lipase (C14)	Colorless (-)	-	-	-	-	-	-
	Leucine arylamidase	Orange (+)	+	+	+	+	+	+
	Valine arylamidase	Colorless (-)	-	-	-	-	-	-
	Cystine arylamidase	Colorless (-)	-	-	-	-	-	-
	Trypsin	Colorless (-)	-	-	-	-	-	-
	α-chymotrypsin	Colorless (-)	-	-	-	-	-	-
	Acid phosphatase	Violet (+)	+	+	+	+	+	+
	Naphthol-AS-BI-phosphohydrolase	Blue (+)	+	+	+	+	+	+
	α-galactosidase	Colorless (-)	-	-	-	-	-	-
	β-galactosidase	Violet (+)	+	+	+	+	+	+
	β-glucuronidase	Colorless (-)	-	-	-	-	-	-
	α-glucosidase	Colorless (-)	-	-	-	-	-	-
	β-glucosidase	Violet (+)	+	+	+	+	+	+
	N-acetyl-β-glucosaminidase	Colorless (-)	-	-	-	-	-	-
	α-mannosidase	Colorless (-)	-	-	-	-	-	-
	α-fucosidase	Colorless (-)	-	-	-	-	-	-
Other tests	CVP (pit formation)	+	+	+	+	+	+	+
	Oxidase (cytochrome-oxidase)	-	-	-	-	-	-	-
	Catalase	+	+	+	+	+	+	+

+, positive; - negative
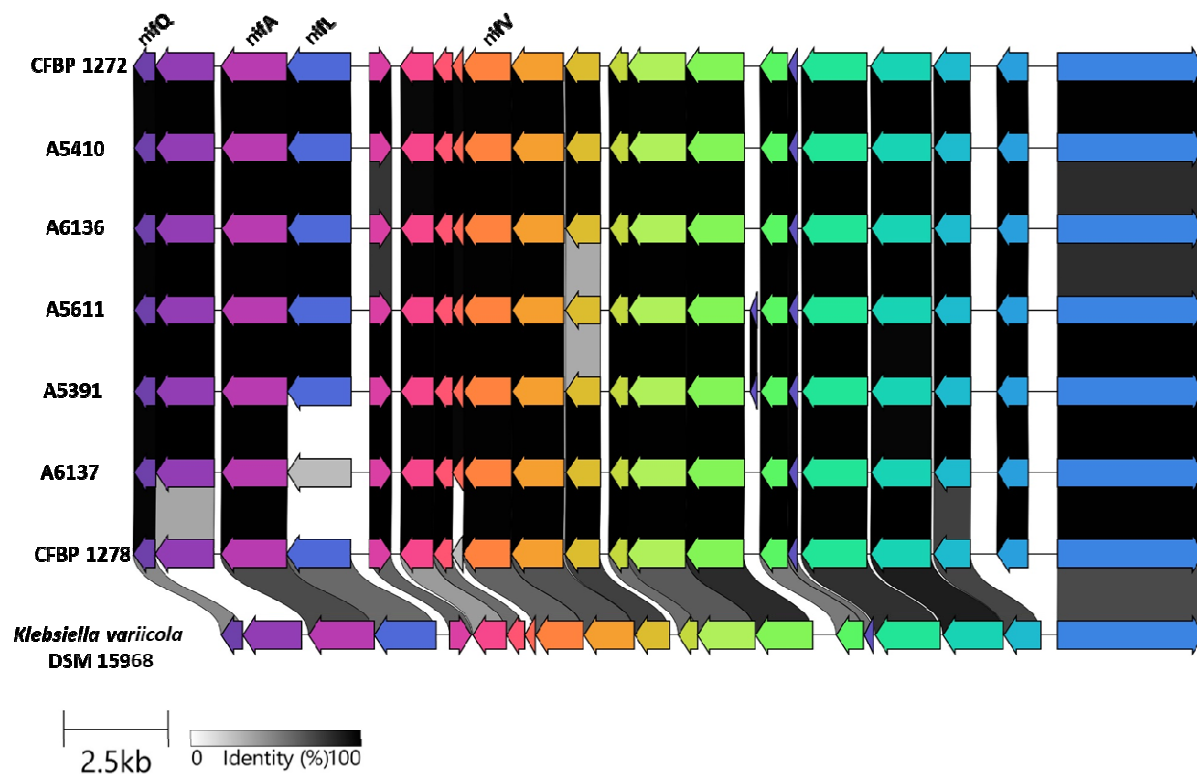
Figure 1. Comparison of the genetic organization of the nitrogen fixation (*nif*) gene cluster among *Dickeya ananae* sp. nov. strains and *Klebsiella variicola* DSM 15968. The arrow positions indicating forward or reverse gene orientation. Arrow colors represent the specific gene composition within the *nif* gene cluster. Gene names are displayed at the top and bottom of the linear graph. Pairwise alignment of the sequences was conducted using the BLAST algorithm with cut-off values of 70%-100%, with regions of higher nucleotide identity shaded in grey and black. Expanded legend entries acronyms are provided: *nif Q* (Fe-Mo cofactor synthesis), *nif B* (Fe-Mo cofactor synthesis), *nifA* (transcriptional activator), *nifL* (transcriptional regulator), *nifF* (electron donation to nitrogenase), *nifM* (nitrogenase maturation), *nifZ* (activation and maturation), *nifW* (protection of the Fe-Mo protein; maturation and activation, *nifV* (Fe-Mo cofactor synthesis), *nifS* (cysteine desulfurase; maturation and activation), *nifU* (Fe-Mo cofactor biosynthesis), *nifX* (Fe-Mo cofactor biosynthesis), *nifN* (Fe-Mo cofactor biosynthesis), *nifE* (Fe-Mo cofactor biosynthesis), *nifY* (Fe-Mo cofactor synthesis), nifT (unknown function), nifK (nitrogenase molybdenum-iron protein β subunit), nifD (nitrogenase molybdenum-iron protein α subunit), *nifH* (nitrogenase reductase), HP (hypothetical protein), *nifJ* (Fe electron transport).
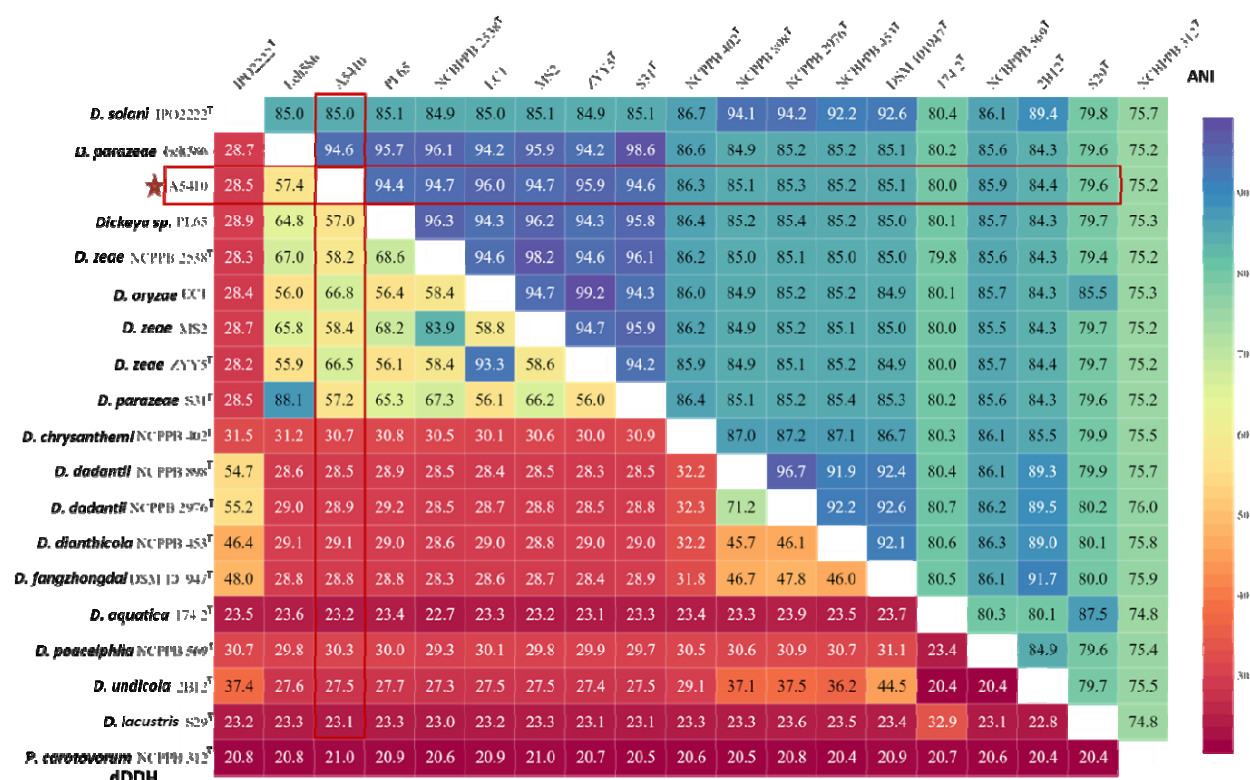
Figure 2. Heatmap displaying combined ANI and dDDH values. ANI values (above diagonal) were calculated using the OrthoANI algorithm in OAT software; dDDH values (below diagonal) were determined using the Genome-to-Genome Distance Calculator (GGDC) version 3.0 (https://ggdc.dsmz.de/). The analysis includes the type strains of each *Dickeya* species and *D. ananae* sp. nov. strain A5410[T].
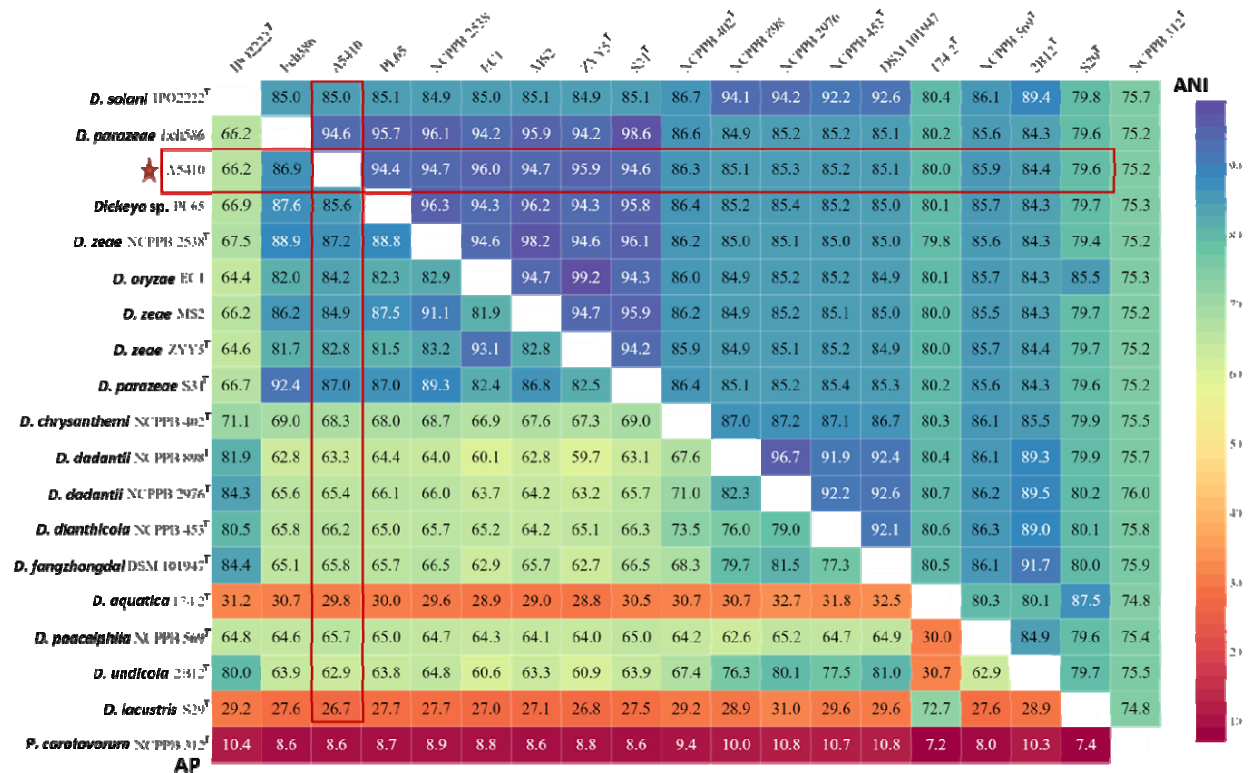
Figure 3. Heatmap showing the combined ANI and AP values. ANI values (above diagonal) and AP values (below diagonal) were calculated using CLC Genomics workbench 22.0.2 for all type strains of the genus *Dickeya* and *D. ananae* sp. nov. A5410^T.
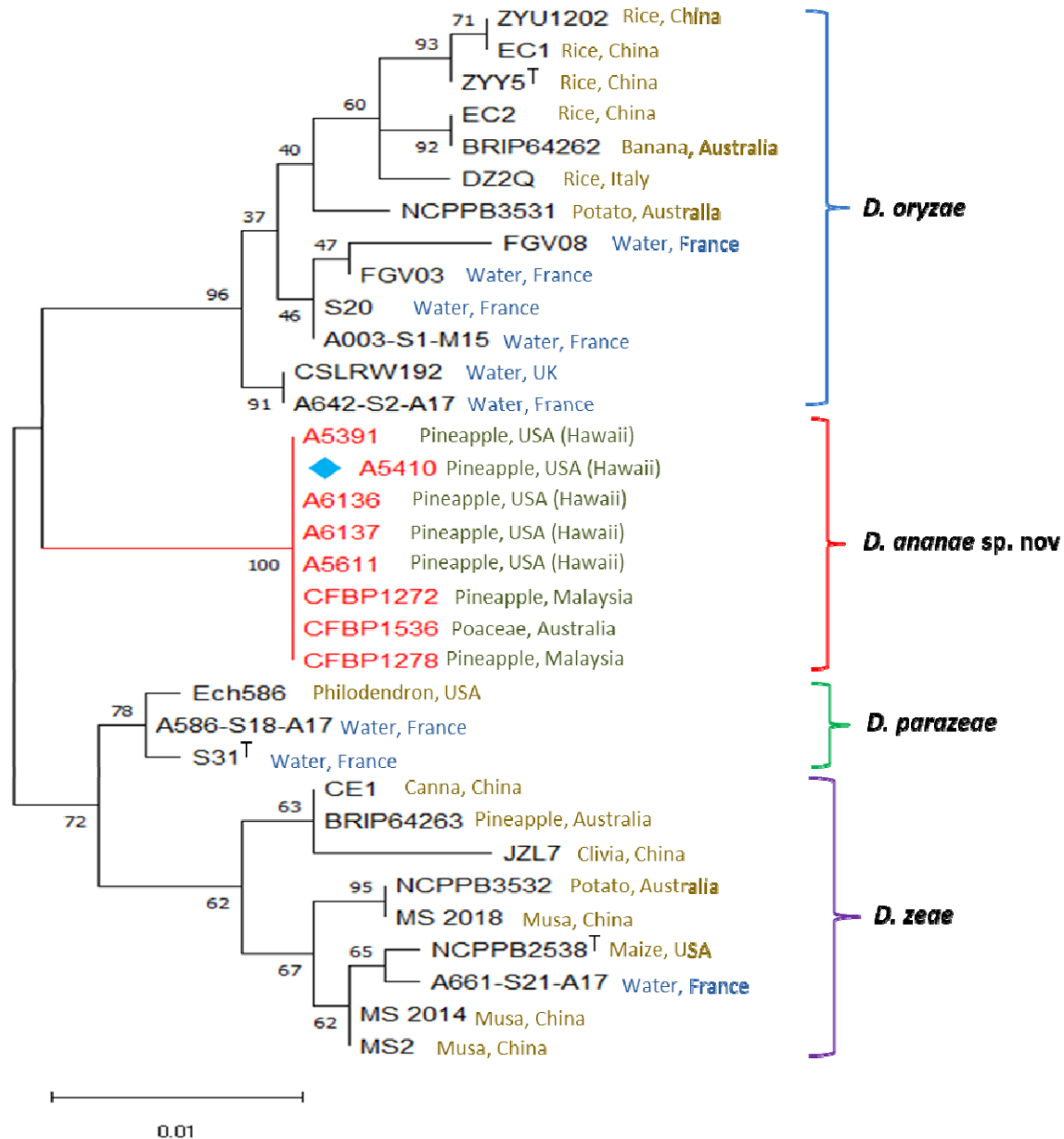
Figure 4. The *gapA* based phylogenetic tree of the *Dickeya* species. This tree includes different strains of each species included in the *D. zeae* complex. Evolutionary history was inferred by using the maximum likelihood method and Tamura-Nei model. The tree with the highest log likelihood (-1409.19) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in number of substitutions per site. This analysis involved 33 nucleotide sequences. There was a total of 702 positions in the final dataset. Evolutionary analyses were conducted in MEGA11.
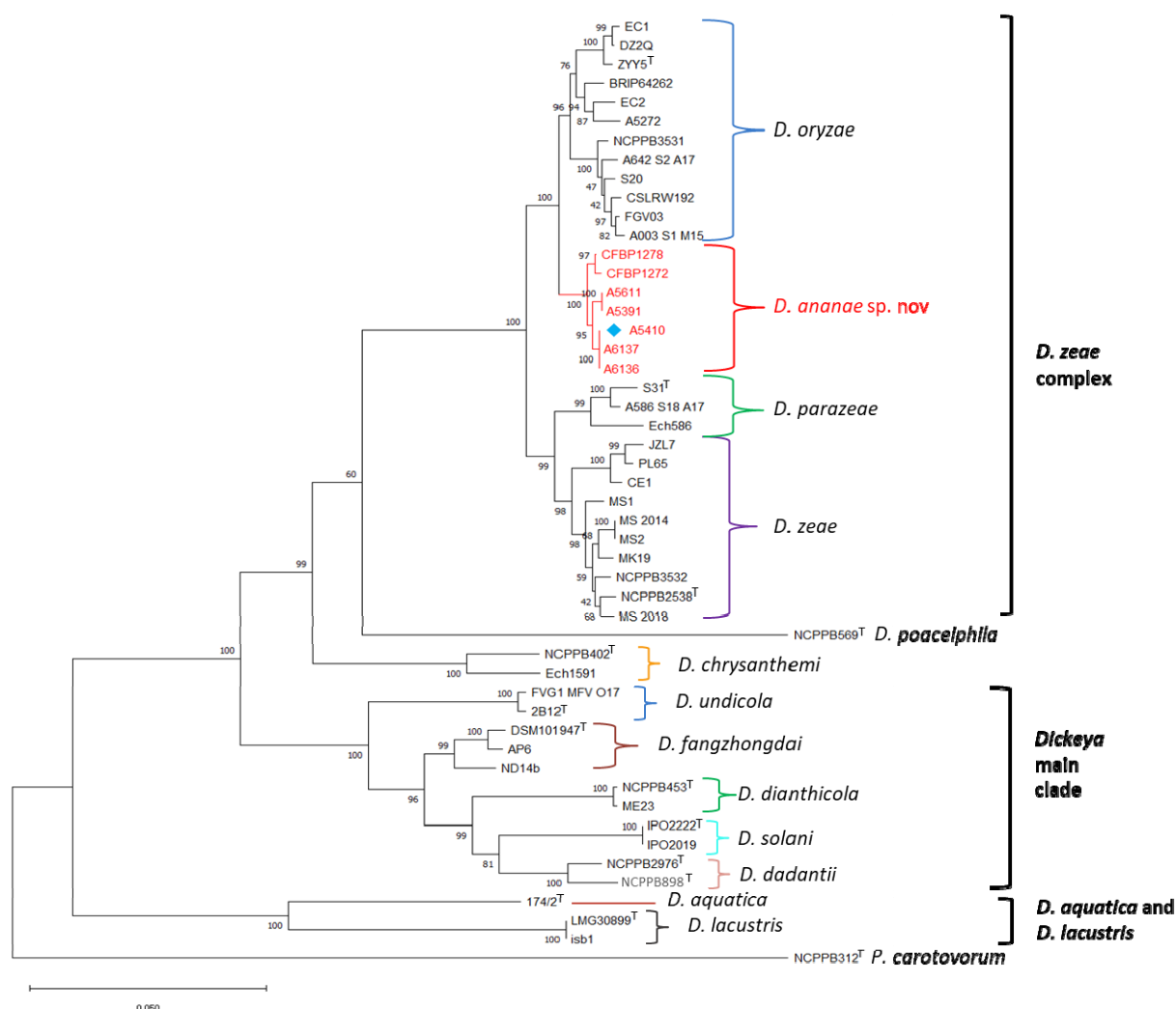
Figure 5. A maximum likelihood phylogenetic tree based on the concatenated sequences of 10 housekeeping genes, *dnaA*, *dnaX*, *fusA*, *gyrB*, *infB*, *recA*, *recN*, *rplB*, *atpD* and *gapA* retrieved from the whole genome sequences of 50 strains including the type and other strains of the other species within genus *Dickeya*. The multiple alignment of each gene sequence and concatenation was performed using Geneious Prime 2021.2.2. The phylogenetic tree was reconstructed using concatenated alignments in MEGA 11 with 1,000 bootstrap replicates. Evolutionary distances were calculated using the Tamura-Nei model. Bootstrap percentages are indicated at the branch points (only values above 60% are shown). The genome of *Pectobacterium carotovorum* was used as an outgroup. The tree scale bar indicates the number of nucleotide substitutions per sequence position. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 50 nucleotide sequences. There was a total of 16,692 positions in the final dataset.
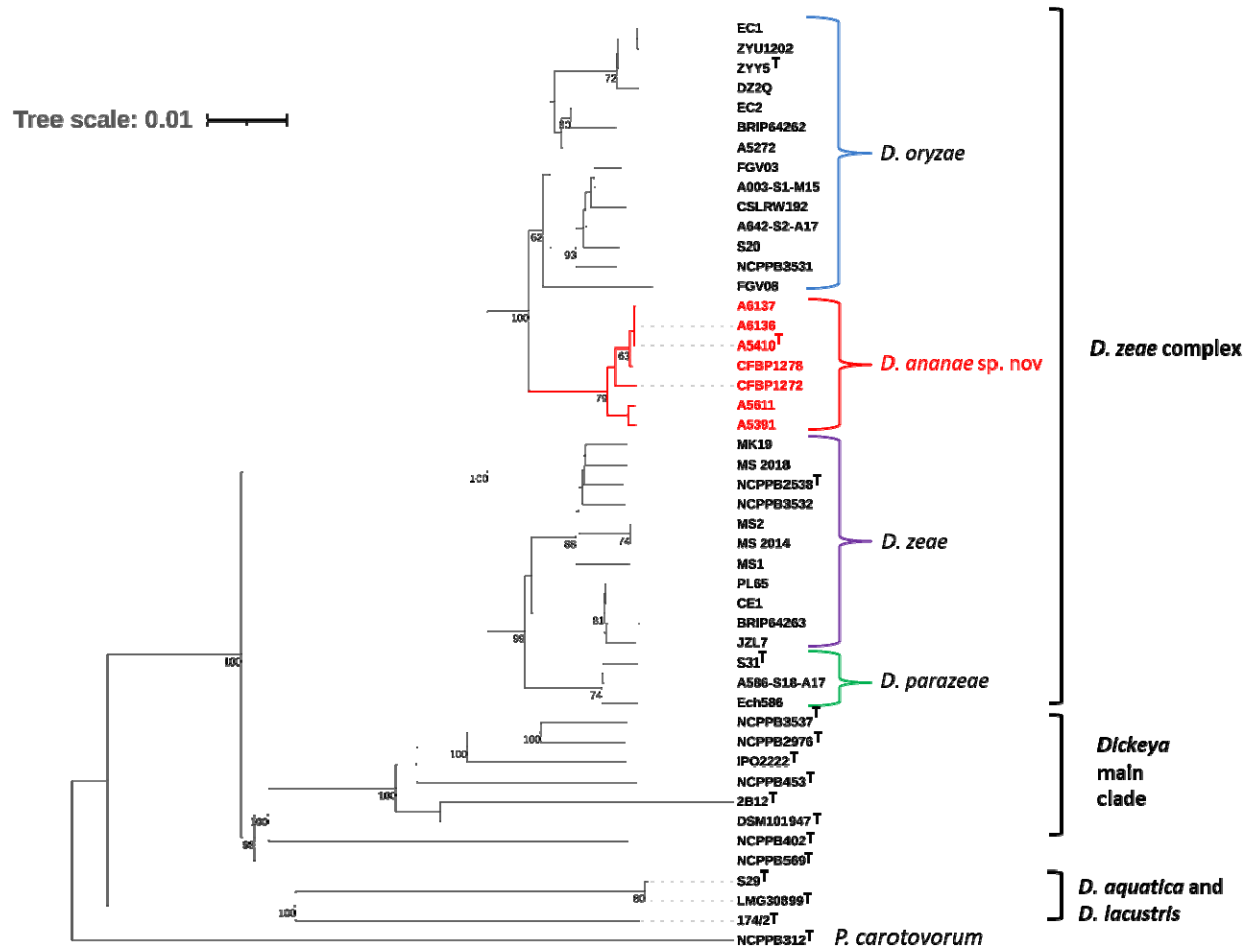
Figure 6. Whole genome sequence-based phylogenetic tree of the genus *Dickeya*. This tree was constructed using the Type Strain Genome Server (TYGS), it shows the relationships between *D. ananae* sp. nov. A5410[T] and strains of different *Dickeya* species. The tree was inferred with FastME 2.1.6.1 using the Genome BLAST Distance Phylogeny (GBDP) approach, with distances calculated from genome sequences via the TYGS server. Branch lengths are scaled in terms of GBDP distance formula d5. The numbers above the branches are GBDP pseudo-bootstrap support values >60% from 100 replications, with an average branch support of 92.7%. The tree was rooted at the midpoint (Farris, 1972). Leaf labels are annotated by affiliation to species and subspecies clusters, genomic G+C content, δ values and overall genome sequence length, number of proteins, and the kind of strain. The phylogenetic trees with bootstrapping value and colored annotations were created using web-based tool Interactive Tree of Life (iTOL v6).
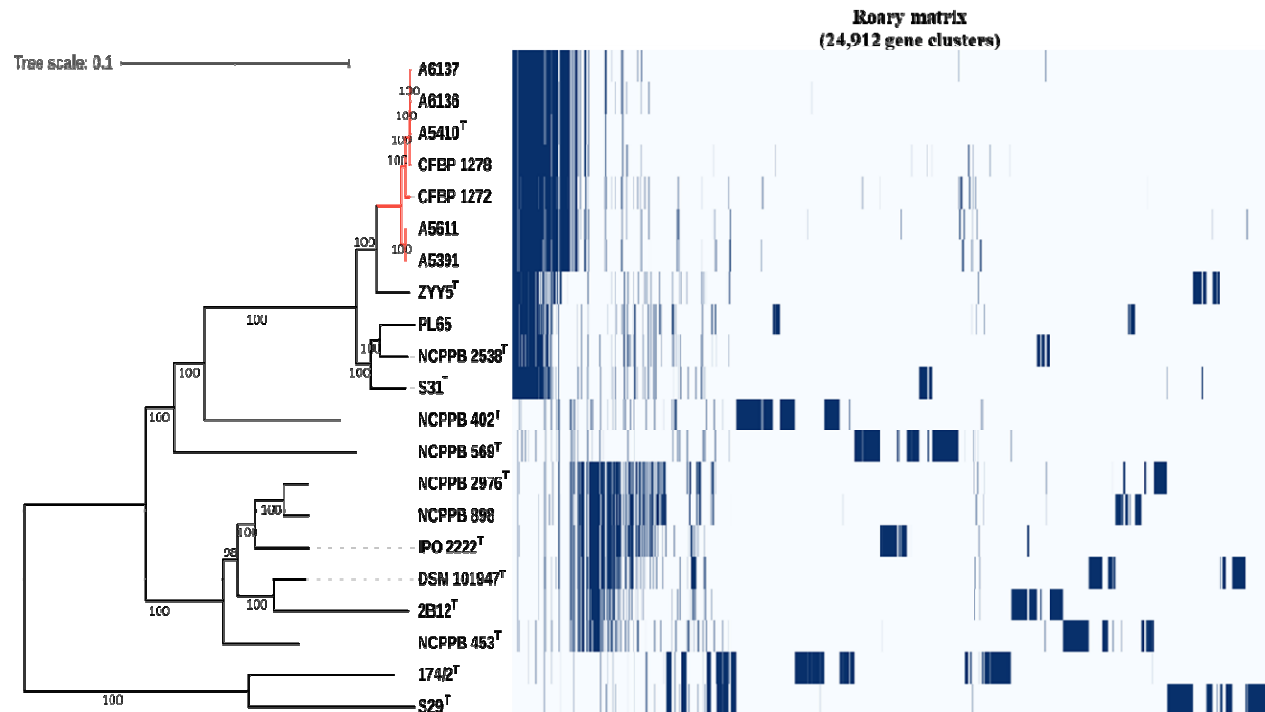
Figure 7. Core and pan genome analyses of *Dickeya* species including *D. ananae* sp. nov. The tree was constructed using 21 strains belonging to various *Dickeya* species. Left: Core-genome-based ML phylogenetic tree. Tree scale bar represents the nucleotide substitutions per site. Right: Pan genome-based Roary matrix of present and absent genes among the *Dickeya* species. The presence (blue) and absence (white) of accessory genome elements is shown.
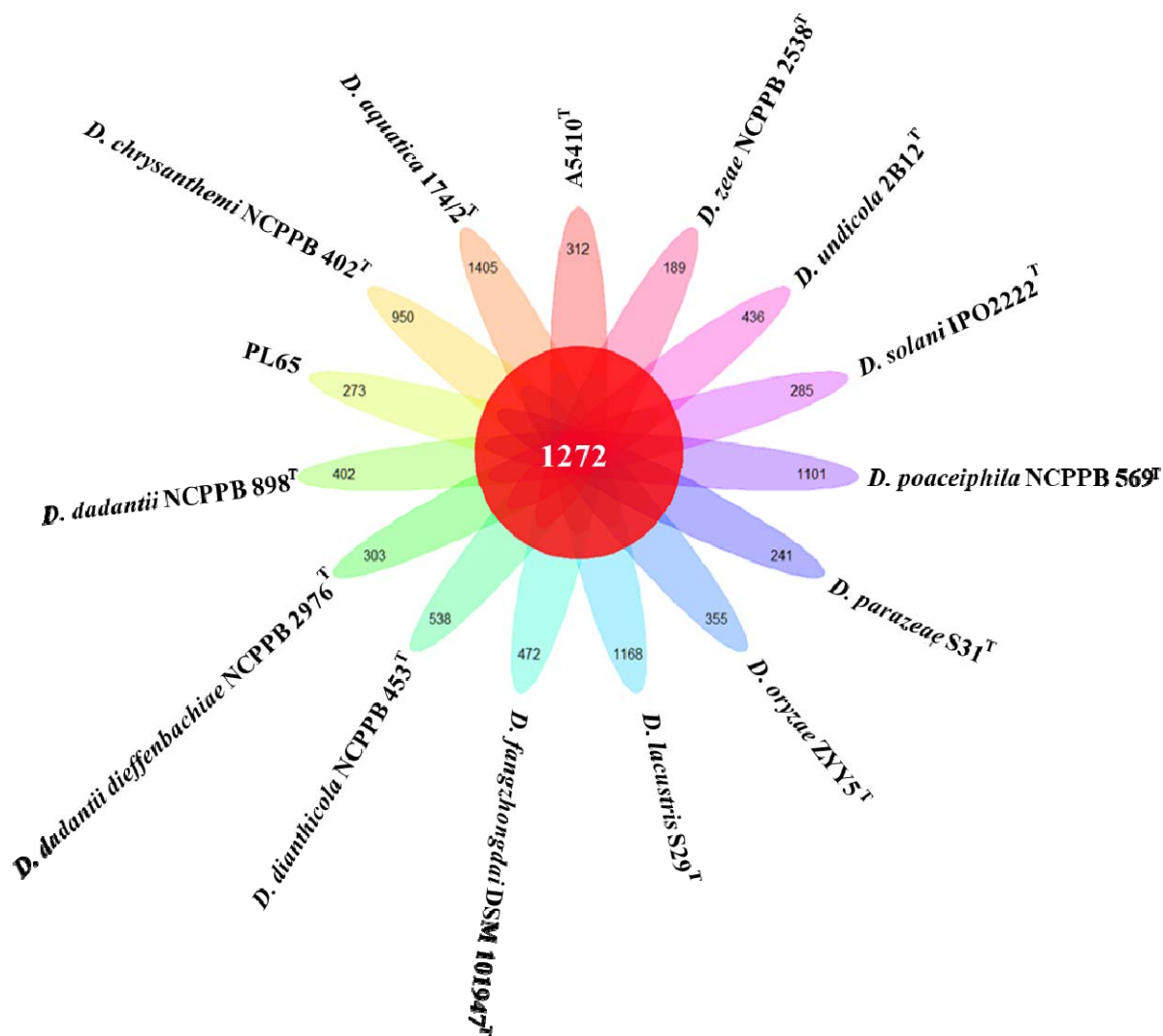
Figure 8. Pan-genome analyses of *Dickeya* species including *D. ananae* sp. nov. Pan-genome analyses of *D. ananae* sp. nov. (A5410[T]) with type strains of the other *Dickeya* species. Floral plot depicting the number of core orthologous genes (center) and the number of unique genes (petals). T in superscript denotes type strain of respective *Dickeya* species.
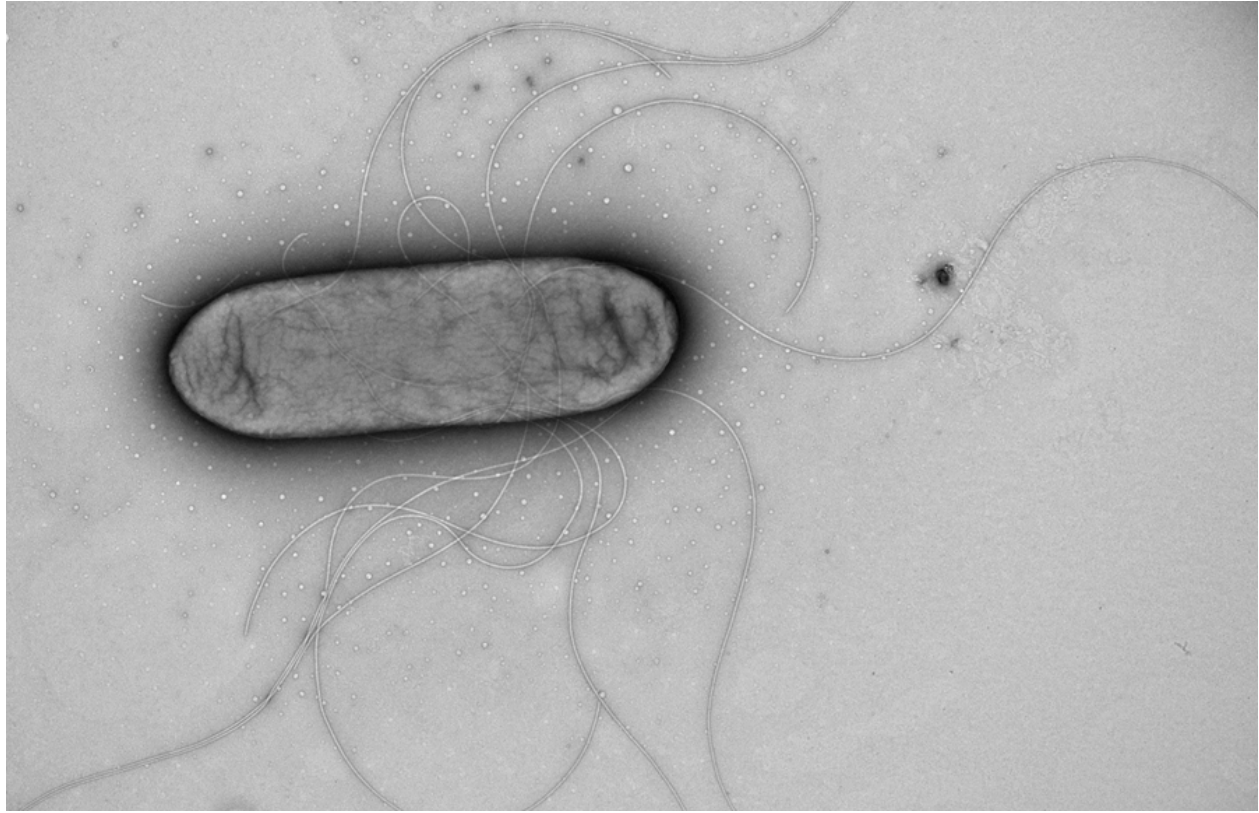
Figure 9. Transmission electron micrograph of *D. ananae* sp. nov. A5410[T]. A minimum of 10 images were captured for the analyzed strain across three biological replicates. A representative image is shown. Scale bar indicates 800 nm.