

1 **Proposal for the creation of a new genus *Musicola* gen. nov., reclassification of**
2 ***Dickeya paradisiaca* (Samson et al. 2005) as *Musicola paradisiaca* comb. nov. and**
3 **description of a new species *Musicola keenii* sp. nov.**

4

5 Nicole Hugouvieux-Cotte-Pattat ^{1*}, Cécile Jacot-des-Combes², Jérôme Briolay², Leighton Pritchard³

6

7 **Author affiliations :**

8 ¹ Univ Lyon, CNRS, INSA Lyon, UCBL, UMR 5240 Microbiologie Adaptation et Pathogénie, F-69622 Villeurbanne,
9 France

10 ² Université de Lyon, Université Claude Bernard Lyon 1, CNRS FR 3728 BioEnviS, plateforme DTAMB, F-69621
11 Villeurbanne, France

12 ³ Strathclyde Institute of Pharmacy & Biomedical Sciences, 161 Cathedral Street, Glasgow G4 ORE, UK

13

14 *Corresponding author: Nicole Hugouvieux-Cotte-Pattat, Nicole.Cotte-Pattat@insa-lyon.fr

15

16 **Email address of authors:**

17 nicole.cotte-pattat@insa-lyon.fr, cecile.jacot-des-combes@univ-lyon1.fr, jerome.briolay@univ-lyon1.fr,

18 leighton.pritchard@strath.ac.uk

19

20 **Keywords:** phytopathogen, soft-rot *Enterobacterales*, *Pectobacteriaceae*, *Dickeya*, *Musicola paradisiaca*

21 **Subject category:** Taxonomic description, new taxa - Proteobacteria

22

23 **Word count:** 4250

24

25 **Depositories:**

26 The type strain *Musicola keenii* A3967^T (CFBP 8732^T, LMG 31880^T) Whole Genome Shotgun project has been
27 deposited at DDBJ/ENA/GenBank under the accession JAAWVW000000000. The version described in this paper
28 is version JAAWVW010000000. The 16S rRNA sequence accession is MT275741.

29

30 **Abbreviations:** ANI, average nucleotide identity; dDDH, digital DNA-DNA hybridization; CFBP, Collection
31 Française de Bactéries Phytopathogènes; CE, carbohydrate esterase; GH, glycoside hydrolase; PL,

32 polysaccharide lyase; RAST, rapid annotations using subsystems technology; T1SS, T2SS, T3SS, T4SS and T6SS,
33 type I, II, III, IV and VI secretion systems.

34

35

36 **ABSTRACT** (238 words)

37 The *Pectobacteriaceae* family of important plant pathogens includes the genus *Dickeya*. There are currently
38 twelve described species of *Dickeya*, although some are poorly characterized at the genomic level. Only two
39 genomes of *Dickeya paradisiaca*, the type strain CFBP 4178^T and strain Ech703, have previously been
40 sequenced. Members of this species are mostly of tropical or subtropical origin. During investigation of strains
41 present in our laboratory collection we sequenced the atypical strain A3967, registered as CFBP 722, isolated
42 from *Solanum lycopersicum* (tomato) in the South of France in 1965. The genome of strain A3967 shares dDDH
43 and ANI values of 68% and 96%, respectively, with the *D. paradisiaca* type strain CFBP 4178^T. However, ANI
44 analysis showed that *D. paradisiaca* strains are significantly dissimilar to the other *Dickeya* species, such that
45 less than 1/3 of their genomes align to any other *Dickeya* genome. On phenotypic, phylogenetic and genomic
46 grounds, we propose a reassignment of *D. paradisiaca* to the genus level, for which we propose the name
47 *Musicola* gen. nov., with *Musicola paradisiaca* as the type species and CFBP 4178^T (NCPBP 2511^T) as the type
48 strain. Phenotypic analysis showed differences between strain A3967^T and CFBP 4178^T, such as for the
49 assimilation of melibiose, raffinose and *myo*-inositol. These results support the description of two novel
50 species, namely *Musicola paradisiaca* comb. nov. and *Musicola keenii* sp. nov., with CFBP 4178^T (NCPBP 2511^T,
51 LMG 2542^T) and A3967^T (CFBP 8732^T, LMG 31880^T) as the type strain, respectively.

52

53

54 **INTRODUCTION**

55 For nearly 40 years, our laboratory has been interested in bacteria formerly belonging to *Erwinia* and
56 *Pectobacterium* but now included in the genus *Dickeya*, an important group of plant pathogens that affect a
57 wide range of hosts, including vegetable crops and ornamental plants [1, 2, 3]. Most characterized *Dickeya*
58 strains originate from infected crops or ornamental plants, although a few *Dickeya* strains have been isolated
59 from water. This genus belongs to the *Enterobacteriales* order and more precisely to the *Pectobacteriaceae*
60 family that includes five genera: *Brenneria*, *Dickeya*, *Lonsdalea*, *Pectobacterium*, and *Sodalis* [4]. *Dickeya* and
61 *Pectobacterium* species cause soft-rot diseases on plants due to the action of extracellular pectinases that
62 attack the plant cell wall [1, 2]. Members of these two genera are often designated as soft-rot
63 *Pectobacteriaceae* (SRP).

64 The history of SRP classification began with the establishment of the genus *Erwinia*, which was founded to
 65 gather several Gram negative plant pathogenic bacteria [5]. This genus included, among others, the species
 66 *Erwinia chrysanthemi* and *Erwinia carotovora*. A non-official classification in six pathovars related to the host
 67 plant was used to describe *E. chrysanthemi* members, namely pv. *chrysanthemi*, pv. *dianthicola*, pv.
 68 *dieffenbachiae*, pv. *parthenii*, pv. *zuae* and pv. *paradisiaca* [6, 7]. Strains of the pathovar *paradisiaca* were also
 69 proposed for elevation to the species *Erwinia paradisiaca* [8]. As a consequence of reclassification in the genus
 70 *Erwinia*, SRP were gathered into the genus *Pectobacterium* that included two species *Pectobacterium*
 71 *carotovorum* and *Pectobacterium chrysanthemi*, replacing *E. carotovora* and *E. chrysanthemi*, respectively [9].
 72 However, the species *E. paradisiaca* was included in the genus *Brenneria* and named *Brenneria paradisiaca*
 73 [10]. The nomenclature again changed in 2005 with the proposal that strains formerly designated *P.*
 74 *chrysanthemi* or *B. paradisiaca* be placed into the new genus *Dickeya* [11]. At this time, the genus *Dickeya*
 75 comprised six recognized species: *D. chrysanthemi*, *D. dadantii*, *D. dieffenbachiae*, *D. dianthicola*, *D. zuae* and
 76 *D. paradisiaca* [11]. Thereafter, new changes were proposed in the genus *Dickeya*. Members of the species *D.*
 77 *dieffenbachiae* were reclassified as a subspecies of *D. dadantii* (i.e. *D. dadantii* subsp. *dieffenbachiae*) [12].
 78 More recently, novel species have been characterized: *D. solani* for isolates responsible for potato diseases in
 79 Europe [13, 14]; *D. fangzhongdai* isolated from pear trees in China and orchids in different countries [15, 16];
 80 *D. poaceiphila* for strains responsible for a sugarcane disease in Australia [17]; and *D. oryzae* causing rice
 81 diseases [18]. Three new species were also identified in water samples: *D. aquatica* from rivers in England and
 82 Finland [19]; *D. lacustris* from lakes in France [20]; and *D. undicola* from water samples collected in Malaysia
 83 and France [21]. Thus, the genus *Dickeya* currently comprises twelve species with validly accepted names: *D.*
 84 *aquatica*, *D. chrysanthemi*, *D. dadantii*, *D. dianthicola*, *D. fangzhongdai*, *D. lacustris*, *D. oryzae*, *D. paradisiaca*,
 85 *D. poaceiphila*, *D. solani*, *D. undicola*, and *D. zuae*.

86 Previous studies have shown *D. paradisiaca* to be the most basal member of the genus *Dickeya* in terms of
 87 nucleotide identity, pan-genome content, genome synteny and whole-genome phylogeny [22, 23, 24].
 88 Genomic data suggested that differences between the *D. paradisiaca* strains and other members of the genus
 89 *Dickeya* would justify separation into a new genus [22, 24]. However, data on the genetic diversity of *D.*
 90 *paradisiaca* are scarce, and it is one of the least well characterized *Dickeya* species at the genomic level. Most
 91 *D. paradisiaca* strains were isolated from *Musa paradisiaca* (banana trees) in tropical or subtropical countries
 92 (Colombia, Cuba, Jamaica, Panama, etc.) [25]. Prior to this publication, only two *D. paradisiaca* genomes were
 93 available, those of the type strain NCPPB 2511^T [26] and of strain Ech703, isolated from *Musa paradisiaca* in
 94 Colombia in 1970 and from *Solanum tuberosum* in Australia, respectively [27, 28].

95 To better understand diversity in the genus *Dickeya*, we analysed poorly characterized SRP strains stored in our
 96 laboratory collection. Phenotypic and genetic analyses suggested that, in addition to the type strain, five

97 isolates belong to the species *D. paradisiaca*. Four isolates appeared to be similar to the type strain, but strain
 98 A3967 showed genetic divergence and atypical phenotype that appeared sufficient to justify the proposal of a
 99 novel species. Phenotypic, phylogenetic and genomic arguments also justify a reassignment of *D. paradisiaca*
 100 to the genus level, for which we propose the name *Musicola* gen. nov. This novel genus includes two species:
 101 *Musicola paradisiaca* comb. nov. as the type strain, and *Musicola keenii* sp. nov., with NCPPB 2511^T (CFBP
 102 4178^T, LMG 2542^T) and A3967^T (CFBP 8732^T, LMG 31880^T) as the type strain, respectively.

103

104

105 GENETIC AND PHENOTYPIC CHARACTERIZATION OF STRAINS

106 To improve our understanding of SRP diversity, we analyzed eight poorly-characterized wild-type strains
 107 available in the collection of the laboratory MAP (<https://map.insa-lyon.fr/en/>) (Table S1). These strains were
 108 formerly obtained from different laboratories or from the French Collection of Phytopathogenic Bacteria, CFBP
 109 (<https://www6.inrae.fr/cirm/CFBP-Bacteries-associees-aux-Plantes>). To perform preliminary strain
 110 identification, the *gapA* gene was amplified by PCR [29] using the IllustraTM PuReTaqTM Ready-To-GoTM kit (GE
 111 Healthcare) on bacterial cell lysates, with the primers gapAF and gapAR (AAGTGAAAGACGGTCACCTGGT and
 112 CGATCAGGTCCAGAACCTTGTT, respectively). As the primer gapAF is inadequate for amplification of *D.*
 113 *paradisiaca* members, it was replaced by gapAFP (AAGTGAAAAATGGCAATCTGGTCGT) to improve the *gapA*
 114 amplification. Sequences of the *gapA* PCR products were determined by Sanger sequencing (Biofidal, Vaux en
 115 Velin, France).

116 A phylogenetic tree was inferred by the Neighbor-Joining method [30] from distances obtained using the
 117 Maximum Composite Likelihood method with alignment by MUSCLE, conducted in MEGA X (version 10.2.4.). In
 118 the resulting *gapA* tree, sequences from five strains (Table 1) grouped with that of the *D. paradisiaca* type
 119 strain CFBP 4178^T (Fig. 1), and they are clustered as an outgroup to the other *Dickeya* species (Fig. 1). Notably,
 120 the *gapA* sequence of strain A3967 diverges from that of the other strains assigned to the species *D.*
 121 *paradisiaca* (Fig. 1).

122 The 16S rRNA genes of each of the five strains were amplified by PCR and sequenced. A BLASTN search against
 123 nt NCBI database confirmed that the top hits for the 16S rRNA amplicons are annotated as strains of the
 124 species *D. paradisiaca*. A 16S phylogenetic tree was inferred as described above in MEGA X (version 10.2.4.),
 125 there were a total of 1548 positions in the final dataset. In the resulting tree, the branch corresponding to the
 126 *D. paradisiaca* strains is clearly included inside the genus *Dickeya* (Fig. S1). However, the 16S rRNA gene
 127 sequences are not considered as a discriminant marker for *Enterobacteriales* classification [4]. Some divergence
 128 in the 16S rRNA sequences was observed between strain A3967 and the other *D. paradisiaca* strains, including
 129 the two strains whose genome has been sequenced, CFBP 4178^T and Ech703 (Fig. S1).

130 A phenotypic characterization of the six available strains (Table 1) was performed using minimal media (M63)
131 supplemented with one carbon compound (2 g l⁻¹) to determine if they could use each compound as a sole
132 carbon and energy source for growth. Growth was recorded after incubating plates at 30°C for 24 to 72 h
133 (Table 1). Similarly to CFBP 4178^T, the five strains were able to grow in the presence of several
134 monosaccharides or oligosaccharides, such as D-arabinose, L-arabinose, D-fructose, D-galactose, D-
135 galacturonate, D-glucose, D-glucuronate, D-mannose, D-ribose, sucrose, and D-xylose (Table 1). In contrast,
136 they were not able to assimilate D-mannitol, L-rhamnose and D-cellobiose. Strain A3967 differs from the five
137 other strains by its capacity to catabolize *myo*-inositol and its inability to utilize melibiose and raffinose (Table
138 1). Thus, strain A3967 showed an atypical phenotype for sugar assimilation in comparison to the five other
139 strains assigned to the species *D. paradisiaca*, which showed homogeneous profiles.

140 In our collection, A3967 was registered as CFBP 722, a strain isolated from tomato in the South of France in
141 1965 (Table S1) [31]. In the analysis leading to the description of the genus *Dickeya* [11], CFBP 722 was found
142 to belong to phenon 5, whose members are now classified as *D. dianthicola*, while *D. paradisiaca* strains were
143 members of phenon 6. Phenons 5 and 6 differed by three phenotypic characters, i.e., assimilation of D-
144 arabinose, D-mannitol and *myo*-inositol [11]. As A3967 is able to assimilate *myo*-inositol, its phenotype is
145 intermediate between those of phenons 5 and 6. It was not possible to verify the original strain since CFBP 722
146 is no longer available in the CFBP collection. Strain A3967 was recently reintroduced in this collection under the
147 number CFBP 8732 (Table S1).

148 As A3967^T (CFBP 8732^T) appears to be an atypical strain, we decided to compare this strain to the *D.*
149 *paradisiaca* type strain CFBP 4178^T using a large biochemical characterization performed with Biolog plates
150 PM1 and PM2A which contain 190 potential carbon sources [32] (Table S2). Inoculations were performed
151 according to the manufacturer instructions and lecture was made after 48h at 30°C. The data obtained for the
152 *D. paradisiaca* type strain agreed with a previous report also using the Biolog plates PM1 and PM2A [15].
153 However, negative results were obtained for several compounds that allowed the growth of CFBP 4178^T in our
154 analysis of sugar assimilation in minimal media (Table 1). Rather than bacterial growth, the Biolog system
155 detects the metabolic activity of the cells due to substrate assimilation; this activity is visualized by the color
156 change of an oxidoreduction indicator. The Biolog system may have been inappropriate for strain CFBP 4178^T,
157 perhaps due to its low metabolic activity. Such a discrepancy between Biolog data and growth on minimal
158 medium supplemented with each compound was not observed for strain A3967^T (CFBP 8732^T) or *D. dadantii*
159 3937 (Table S2, Table 1). In the Biolog system, dissimilar results between the two strains A3967^T (CFBP 8732^T)
160 and CFBP 4178^T were observed for assimilation of 17 carbon sources (Table S2). Three differences were
161 confirmed by testing the bacterial growth in minimal medium, namely *myo*-inositol, D-melibiose, and D-
162 raffinose. Ten negative results of CFBP 4178^T should be taken with caution as the substrates have not been

163 tested in minimal medium, namely D-fructose-6-phosphate, galactaric acid, D-glucaric acid, D-glucose-6-
164 phosphate, L-lyxose, D-pscicose, L-serine, L-alanine, D-malic acid, and pyruvic acid.

165 The effect of temperature on bacterial growth was analysed in LB medium by incubations ranging from 25 to
166 43°C. Cell density was estimated by measuring optical density at 600 nm (OD₆₀₀) at 24 and 48h. Optimal growth
167 temperature was taken to be that giving the highest OD₆₀₀ value; the maximal growth temperature was the
168 highest temperature allowing for significant growth. Growth of the six selected strains was observed across a
169 wide range of temperatures, with an optimal growth rate from 27 to 33°C and the maximal temperature
170 allowing growth was 40-41°C (Fig. S2).

171 Since *Dickeya* isolates are characterized by their ability to secrete several plant cell wall degrading enzymes, we
172 used a range of media to detect such activities [20] (Table 2). The bacterial motility was measured by the
173 growth diameter 24 h after inoculation in 0.3% L agar plate for swimming and on 0.6% L agar plate for
174 swarming [20]. The maceration ability was evaluated by the length of macerated tissue observed 24 h after
175 inoculation for chicory leaves and the weight of macerated tissue obtained after 48 h for potato tubers [36]. In
176 each case, *D. dadantii* 3937 was used as a reference strain. The six strains assigned to the species *D.*
177 *paradisiaca* showed a good pectinase activity on medium containing polygalacturonate, and they were able to
178 grow in the presence of this polysaccharide as a sole carbon source (Table 1). They secreted cellulase but no
179 protease or lipase activities were observed (Table 2). All strains were motile but showed various levels of
180 swimming and swarming motilities (Table 2). In comparison to *D. dadantii* 3937, the six strains showed weak
181 ability to macerate plant tissues either on chicory leaves or potato tubers, with A3967^T (CFBP 8732^T) being a
182 little more efficient than the five strains assigned to the species *D. paradisiaca*, especially for maceration of
183 potato tubers (Table 2).

184

185

186 GENOMIC COMPARISONS

187 To gain further information on strain A3967^T (CFBP 8732^T), its genome sequence was determined. The total
188 bacterial genomic DNA was extracted using a NucleoSpin^R bacterial DNA purification kit (Macherey-Nagel).
189 Quantification and quality control of the DNA was performed using a Nanodrop spectrophotometer, a Qubit4
190 fluorimeter and agarose gel electrophoresis. Genomic DNA was sequenced using a MiSeq Illumina platform
191 (Biofidal, Vaux en Velin, France). We assembled the genome using SPAdes version 3.11.1 (2020/10) [33]. The
192 resulting draft genome of strain A3967^T (CFBP 8732^T) comprises 72 contigs (N50=266,679, L50=7, 125X
193 coverage depth) with a total length of 4,402,645 bp and a G+C content (mol%) of 54.4%. The draft genome was
194 automatically annotated using RAST version 2.0 (2020/10) [34], which predicted 4,354 protein coding genes

195 and 76 RNA-coding sequences, including 68 tRNAs and 8 rRNAs (23S, 16S, and 6 x 5S). The draft genome has
196 NCBI accession GCF_014855505.1.

197 To clarify the taxonomic position of strain A3967^T (CFBP 8732^T), we calculated digital DNA-DNA hybridization
198 (dDDH) and average nucleotide identity (ANI) values. The dDDH method was proposed as a means of
199 approximating the wet-lab DDH method [35]. Using the A3967^T (CFBP 8732^T) genome as a reference and
200 *Dickeya* genomes as queries (Table 3), dDDH gave values of 68.3 and 68.4% with *D. paradisiaca* CFBP 4178^T and
201 Ech703, respectively, and values lower than 24% with other *Dickeya* species. The dDDH value of 68.3%
202 obtained by comparing strain A3967^T (CFBP 8732^T) with the *D. paradisiaca* type strain is below the threshold
203 dDDH of 70% commonly used to delineate species [35]. However in practice there is significant variance in the
204 results of *in vitro* DDH; genome sequences sharing ≈70% overall sequence identity may give DDH estimates
205 between 60% and 90% [35, 36]. In addition, there is uncertainty in the mapping of dDDH values to DDH values
206 obtained *in vitro*. Straightforward transfer of a 70% dDDH threshold to species boundary is therefore not
207 reliable.

208 Average Nucleotide Identity (ANI) approaches directly calculate the sequence identity of two genomes [37].
209 ANI values must always be interpreted alongside “aligned fraction” or “coverage” values, which report the
210 proportion of pairwise-homologous regions, as these are essential context regarding the proportion of each
211 genome that is similar [36]. Pairwise ANIm values were calculated using pyani v0.3 for 135 genomes
212 downloaded from the NCBI assembly database corresponding to all publicly-available genomes of *Dickeya*
213 (taxid:204037) and *Brenneria* (taxid:71655) [selection of *Dickeya* strains in Fig. 2; all strains in Fig. S3]. This
214 indicated that no more than 33% of the *D. paradisiaca* genomes (A3967, Ech703, NCPPB 2511) could be aligned
215 with any other *Dickeya* genome. In particular, no more than 18% of any *D. paradisiaca* genome could be
216 aligned with any *D. lacustris* or *D. aquatica* genome (*D. lacustris* and *D. aquatica* can be aligned over at least
217 73% of their genomes). Genomes of the nine species *D. chrysanthemii*, *D. dadantii*, *D. dianthicola*, *D.*
218 *fangzhongdai*, *D. oryzae*, *D. poaceiphila*, *D. solani*, *D. undicola* and *D. zaeae*, could be aligned to each other over
219 at least 59% of their total lengths.

220 Although there is no formal definition of genus delineation on the basis of genome similarity, it has been
221 argued that assigning two organisms to the same genus is not sound when only a small proportion of their
222 genome sequences is recognisably homologous. Empirical evidence from recent approaches using aligned
223 fraction to discriminate at genus level indicates that genus boundary values vary by taxon, but members of the
224 same genus tend to share at least 60% coverage [38, 39]. Consistent with these surveys, the *D. paradisiaca*
225 genomes would constitute a distinct genus from the *D. lacustris*/*D. aquatica* group, and from the remaining
226 group of *Dickeya* genomes including *D. chrysanthemii*, *D. dadantii*, *D. dianthicola*, *D. fangzhongdai*, *D. oryzae*, *D.*
227 *poaceiphila*, *D. solani*, *D. undicola* and *D. zaeae*. In particular, we observe that the alignment coverage between

228 *D. paradisiaca* and any other *Dickeya* genome is no greater than that between isolates of *Escherichia* and
 229 *Salmonella*. Hence we propose reclassification of A3967 and strains previously assigned to the species *D.*
 230 *paradisiaca* into a novel genus.

231 There is substantial evidence for a discontinuity in ANI sequence identity consistent with pre-existing bacterial
 232 species distinctions [40]. The threshold, associated with barriers to homologous recombination, corresponds to
 233 identities around 95±1 % ANI, the exact value differing by genus [41, 42]. There is consequently no precise
 234 universal ANI percentage threshold applicable to all bacterial genera or species, and additional evidence must
 235 be taken into account. In our analysis, isolate A3967^T (CFBP 8732^T) has about 96% ANI identity with the
 236 genomes of the two strains assigned to the species *D. paradisiaca*, consistent with a species-level delineation
 237 that is supported by the phenotypic differences reported above and the genomic differences described below.

238

239

240 **PHYLOGENOMIC ANALYSIS SUPPORTING ESTABLISHMENT OF *MUSICOLA* GEN. NOV.**

241 Whole-genome classification of 49 genomes spanning five genera in the *Pectobacteriaceae* (including *Musicola*)
 242 was performed using pyani v0.3.0b [22] with the ANIm algorithm. Assuming 94-96% identity as an approximate
 243 threshold corresponding to species division, and 40-50% coverage as an approximate threshold corresponding
 244 to genus division, the results support the following eight genus divisions (Fig. S4): (1) *Dickeya* (*D. solani*, *D.*
 245 *dadantii*, *D. fangzhongai*, *D. undicola*, *D. dianthicola*, *D. poaceiphila*, *D. zaeae*, *D. chrysanthemii*); (2) *Musicola* (*M.*
 246 *paradisiaca*, *M. keenii*); (3) a novel genus for *D. aquatica* and *D. lacustris*; (4) *Lonsdalea* (*L. iberica*, *L. quercina*,
 247 *L. britannica*); (5) *Pectobacterium* (*P. atrosepticum*, *P. wasabiae*, *P. parvum*); and three novel genera for *B.*
 248 *roseae*, *B. alni*, and *B. goodwinii*, respectively.

249 A multigene maximum-likelihood phylogenetic reconstruction was performed on same set of genomes (Fig. 3).
 250 To ensure consistency of annotation between genomes, all sequences were reannotated using prodigal v2.6.3
 251 [43] to obtain a predicted proteome. In total 1201 single-copy orthologues were identified across the predicted
 252 proteomes of all 49 genomes, using orthofinder v2.5.2 [44]. The protein sequences for these genes were
 253 aligned using MAFFT v7.480 [45] and the corresponding nucleotide coding sequences threaded using t-coffee
 254 v12.00.7fb08c2 [46]. The threaded DNA sequences were concatenated to generate on sequence per genome
 255 using the Python script concatenate_cds.py, which also generated a partition file (one partition per gene).

256 The partition file and concatenated alignment were used as input to raxml-ng v1.0.2 [47] to generate a best-fit
 257 phylogenetic tree using maximum likelihood. The GTR+F0+G4m+B model was used, parameterized separately
 258 for each of the 1201 genes. A single tree topology was fit for each of 20 starting trees, suggesting that this was
 259 the globally-optimal topology. One hundred bootstrap replicate trees were determined to estimate support
 260 values for each tree partition; MRE-based bootstrapping indicated that convergence could be reached with

261 only 50 replicates. The fitted topology (Fig. 3) supports the genus and species divisions implied by whole-
 262 genome classification (Fig. S4). Thus, both ANIm and a comprehensive multigene phylogeny support the same
 263 genus and species divisions, including establishment of *Musicola* as a novel genus.

264 In conclusion, this study adds further data supporting a reassignment that was suggested by previous genome-
 265 scale analyses of the *Dickeya* genus [22, 24]. Phenotypic, phylogenetic and genomic arguments justify a
 266 reassignment of *D. paradisiaca* to the genus level for which we propose the name *Musicola* gen. nov., with
 267 *Musicola paradisiaca* as the type species and CFBP 4178^T (NCPBP 2511^T) as the type strain for the genus.
 268 Moreover, characterization of strain A3967^T (CFBP 8732^T) allows the description of two species in this new
 269 genus for which we propose the names *Musicola paradisiaca* comb. nov. and *Musicola keenii* sp. nov., with
 270 CFBP 4178^T (NCPBP 2511^T, LMG 2542^T) and A3967^T (CFBP 8732^T, LMG 31880^T) as the type strain, respectively.
 271 The proposal of these two species is supported by genomic and phenotypic analysis revealing clear differences
 272 between strains A3967^T (CFBP 8732^T) and CFBP 4178^T. These analyses indicated that a simple distinction
 273 between isolates of the two *Musicola* species can be obtained by testing the assimilation of *myo*-inositol,
 274 melibiose and raffinose (Table 1, Table S3, Table S4).

275

276

277 GENOME CONTENT OF *MUSICOLA* SPECIES

278 The genome content of *M. keenii* A3967^T (CFBP 8732^T) was compared with that of the *M. paradisiaca* strains
 279 CFBP 4178^T and Ech703 whose genome has been previously sequenced (GCA_000400505.1 and
 280 GCA_000023545.1, respectively). The genomes of CFBP 4178^T or Ech703 are closely related, with only a few
 281 tens of genes predicted to be present in only one of these two strains (data not shown). In contrast, using the
 282 function-based comparison tool in RAST, about 600 genes were predicted to be present in A3967^T (CFBP 8732^T)
 283 and absent in CFBP 4178^T or Ech703 or, *vice versa*, present in CFBP 4178^T or Ech703 and absent in A3967^T (CFBP
 284 8732^T). Among functionally-annotated genes, 110 were specific to *M. keenii* A3967^T (CFBP 8732^T) (Table S3),
 285 including a cluster encoding a pilus of the Inl1 type, the *iol* cluster involved in *myo*-inositol assimilation, a
 286 cluster encoding two VirB4 components of type 4 secretion system (T4SS) and two predicted β -glucosidases.
 287 Conversely, 125 functionally-annotated genes were found in *M. paradisiaca* strains CFBP 4178^T and Ech703 but
 288 absent in A3967^T (CFBP 8732^T), including the genes *rafAT* involved in both melibiose and raffinose assimilation,
 289 a cluster encoding the components of a CRISPR system, a cluster encoding four polyketide synthases (PKS),
 290 three toxin/antitoxin couples and three loci containing predicted prophage genes (Table S4). These genomic
 291 differences confirm a genomic basis for the phenotypic differences observed between the two strains A3967^T
 292 (CFBP 8732^T) and CFBP 4178^T for *myo*-inositol, melibiose and raffinose assimilation (Table 1). By comparison

293 with *Dickeya* species, the lack of the cluster *mtIAD* in the genomes of A3967^T (CFBP 8732^T), CFBP 4178^T and
 294 Ech703 explains the inability of *Musicola* strains to assimilate D-mannitol (Table 1).

295 The A3967^T (CFBP 8732^T) annotation was used to search for genes potentially involved in the degradation of
 296 plant cell walls by focusing on pectate lyases, the main determinant of the soft rot symptoms caused by
 297 *Dickeya* [48]. Similar genes involved in the degradation of pectic polysaccharides were found in A3967^T (CFBP
 298 8732^T), CFBP 4178^T and Ech703 genomes. They contain five genes encoding pectate lyases of the family PL1
 299 (*pelA*, *pelD*, *pelB*, *pelC*, *pelZ*), one of family PL2 (*pelW*), three of family PL9 (*pelL*, *pelN*, *pelX*), a potential pectin
 300 lyase of the family PL1 (*pnIG*), an oligogalacturonate lyase of the family PL22 (*ogl*), two putative
 301 polygalacturonases of the family GH28 (*pehK*, *pehX*) and two esterases of families CE8 and CE12 (*pemA* and
 302 *paeY*, respectively). The three genomes encode a cellulase of the family GH5 (CelZ) and contain a cluster
 303 encoding a type II secretion system (T2SS) which is responsible for specific secretion of pectate lyases and of
 304 the cellulase CelZ in the genus *Dickeya*. In contrast to the *Dickeya* species, the three *Musicola* strains possess
 305 neither genes encoding metalloproteases nor the type I secretion system (T1SS) responsible for their secretion.
 306 This is consistent with the absence of protease secretion in *Musicola* (Table 1). The three *Musicola* genomes
 307 encode the major regulators known to be involved in *Dickeya* virulence: KdgR, PecS, PecT, Crp, RsmA,
 308 RsmC/HexY, MfbR, SlyA and the quorum sensing system, Vfm [49]. However, the three *Musicola* strains have a
 309 partial N-acyl homoserine lactone (AHL) dependent quorum sensing system as they encode the regulator ExpR
 310 but not the AHL synthase Expl. In comparison to *Dickeya*, the three *Musicola* strains are poorly equipped to
 311 fight against oxidative stresses as their genomes lack the genes *katE*, *sodC*, *indABC*, *hmpX* and *sufABCDSE*. As
 312 previously observed for CFBP 4178^T and Ech703, the cluster of genes involved in flagellum biosynthesis of
 313 A3967^T (CFBP 8732^T) is different of that found in *Dickeya* species and more related to the *Lonsdalea* and
 314 *Brenneria* flagellum gene cluster [24].

315 Several differences presented by CFBP 4178^T and Ech703 in comparison with other *Dickeya* species were
 316 previously observed by Pedron and Van Gijsegem who use genome data to investigate the diversity in the
 317 genus *Dickeya* [24]. By examination of the *Dickeya* core genome, they noticed that only 1 800 genes,
 318 representing about 40% of the gene content, are common between *D. paradisiaca* strains, i. e. CFBP 4178^T and
 319 Ech703, and other *Dickeya* species [24], a value consistent with our results on genome coverage. Moreover, a
 320 pangenome analysis based on the presence/absence of the genes, showed that the two strains CFBP 4178^T and
 321 Ech703 cluster outside the *Dickeya* genus [24]. The authors also noticed that CFBP 4178^T and Ech703 lack
 322 several genes known to be involved in *Dickeya* virulence [24]. These dissimilarities, also observed in this study
 323 for strain A3967^T (CFBP 8732^T), suggest notable differences in the virulence strategies of *Dickeya* and *Musicola*
 324 members. Indeed, virulence tests have shown that the *Musicola* strains have a weak maceration activity on

325 potato tubers and chicory leaves, two plant models classically used to evaluate the virulence of *Dickeya* strains
326 (Table 2).

327

328

329 DESCRIPTION OF THE NEW GENUS AND SPECIES

330 Description of *Musicola* gen. nov.

331 *Musicola* [Mu.si'co.la, N.L. fem. n. *Musa* the genus of the banana; L. suff. *-cola* from L. masc. or fem. *incola*
332 inhabitant, dweller; N.L. fem n. *Musicola* an inhabitant of *Musa*].

333 This taxon was previously described as *Dickeya paradisiaca* (Samson et al 2005).

334 Other synonyms: *Erwinia paradisiaca*; *Erwinia chrysanthemi* pathovar *paradisiaca*; *Erwinia chrysanthemi*
335 phenom 6; *Pectobacterium chrysanthemi* biovar 4; *Brenneria paradisiaca*.

336 *Musicola* members are gram-negative, non-sporeforming, facultatively anaerobic pectinolytic bacteria. Cells
337 have average dimensions of 0.6 by 1.7 µm. They are motile with peritrichous flagella. After 48h at 30°C on LB
338 medium (5 g.l⁻¹ tryptone, 3 g.l⁻¹ yeast extract, 5 g.l⁻¹ NaCl and 15 g.l⁻¹ agar), they form pale cream-colored
339 colonies of 0.5-2.5 mm in diameter with translucent appearance. The optimum temperature for bacterial
340 growth is 25-36°C and they can grow up to 40-41°C. They are able to grow in the presence of D-arabinose, L-
341 arabinose, N-acetyl-D-glucosamine, D-fructose, D-galactose, D-galacturonic acid, D-gluconic acid, D-glucose, D-
342 glucuronic acid, glycerol, D-mannose, D-ribose, sucrose, D-xylose, arbutin, salicin, citric acid, lactic acid,
343 succinic acid, polygalacturonate or pectin as the sole carbon source but they unable to utilize D-arabitol, D-
344 cellobiose, gentiobiose, D-lactose, D-maltose, D-mannitol, L-rhamnose, D-sorbitol, or D-trehalose. They
345 produce extracellular pectinases, a cellulase but no proteases. They do not produce the blue pigment
346 indigoidine.

347 The type species of the genus is *Musicola paradisiaca*, with NCPPB 2511^T (CFBP 4178^T; LMG 2542^T) as the type
348 strain.

349 Description of *Musicola paradisiaca* comb. nov.

350 *Musicola paradisiaca* [pa.ra.di.si.a'ca L. fem. adj. *paradisiaca*, referring to the isolation of most strains from
351 *Musa paradisiaca*]. General description as for the genus. The optimum temperature for bacterial growth is
352 about 33°C and they can grow up to 40°C. The *M. paradisiaca* type strain is able to grow in the presence of D-
353 arabinose, L-arabinose, N-acetyl-D-glucosamine, D-fructose, D-galactose, D-galacturonic acid, D-gluconic acid,
354 D-glucose, D-glucuronic acid, glycerol, D-mannose, D-melibiose, D-raffinose, D-ribose, sucrose, D-xylose,
355 arbutin, salicin, citric acid, lactic acid, succinic acid, polygalacturonate or pectin as the sole carbon source but it
356 is unable to utilize *myo*-inositol.

357 The DNA G+C content of the type strain NCPPB 2511^T (CFBP 4178^T; LMG 2542^T) is 55.0% based on the genome
 358 sequence. Other characterized members of this species include strain Ech703, identified on the basis of its
 359 genome analysis, and strains CFBP 1445, CFBP 1446, CFBP 1451, CFBP 3477, CFBP 3696, and CFBP 3699
 360 identified on the basis of phenotypic data. Among these *M. paradisiaca* members, six strains were isolated in
 361 Colombia from *Musa paradisiaca*, from 1968 to 1972, and two strains were isolated in Cuba in 1987, from
 362 *Musa* sp. and *Zea mays*, respectively.

363 **Description of *Musicola keenii* sp. nov.**

364 *Musicola keenii* [keen'i.i N.L. gen n. *keenii* in honour of the American molecular biologist Noel T. Keen] [50].
 365 General description as for the genus. The optimum temperature for bacterial growth is about 33°C and the
 366 type strain can grow up to 41°C. *M. keenii* A3967^T is able to grow in the presence of D-arabinose, L-arabinose,
 367 N-acetyl-D-glucosamine, D-fructose, D-fructose-6-phosphate, D-galactose, D-galacturonic acid, glucaric acid, D-
 368 gluconic acid, D-glucose, D-glucose-6-phosphate, D-glucuronic acid, glycerol, *myo*-inositol, D-mannose, D-
 369 ribose, D-psicose, sucrose, D-xylose, arbutin, salicin, citric acid, lactic acid, pyruvic acid, succinic acid,
 370 polygalacturonate or pectin as the sole carbon source but it is unable to utilize D-melibiose or D-raffinose.
 371 The DNA G+C content of the type strain A3967^T (CFBP 8732^T; LMG 31880^T) is 54.4% based on the genome
 372 sequence (sequence accession JAAWVW000000000).

373

374

375 **AUTHORS' STATEMENTS**

376 **Funding information**

377 This work was supported by funding of CNRS, University Lyon 1, and INSA Lyon to UMR 5240.

378

379 **Acknowledgements**

380 We thank Véronique Utzinger for medium preparation and contribution to phenotypic analyses and Perrine
 381 Portier for exchanges on the strains of CFBP (Collection Française de Bactéries Phytopathogènes, Beaucauzé,
 382 France). This work was supported by CNRS, University Lyon 1, and INSA Lyon. It has benefited from the
 383 expertise and facilities of the platform DTAMB of FR 3728 (Villeurbanne, France), with sequencing performed
 384 by BIOFIDAL (Vaux-en-Velin, France).

385

386 **Supporting Information**

387 Scripts and data enabling reproduction of the phylogenomic analysis presented in this manuscript can be
 388 obtained at https://widdowquinn.github.io/SI_Hugouvieux-Cotte-Pattat_2021/

389

390 **Conflicts of interest**

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

392

393 **ABBREVIATIONS**

394 ANI, average nucleotide identity; dDDH, digital DNA-DNA hybridization; CE, carbohydrate esterase; GH,
 395 glycoside hydrolase; PL, polysaccharide lyase; RAST, rapid annotations using subsystems technology; T1SS,
 396 T2SS, T3SS, T4SS and T6SS, type I, II, III, IV and VI secretion systems.

397

398 **REFERENCES**

- 399 1. **Charkowski A, Blanco C, Condemine G, Expert D, Franza T et al.** The role of secretion systems and
 400 small molecules in soft rot *Enterobacteriaceae* pathogenicity. *Annu Rev Phytopathol* 2012; 50: 425-449.
- 401 2. **Hugouvieux-Cotte-Pattat N, Condemine G, Gueguen E, Shevchik VE.** *Dickeya* plant pathogens. *eLS*
 402 2020; March. DOI: [10.1002/9780470015902.a0028932](https://doi.org/10.1002/9780470015902.a0028932)
- 403 3. **Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M et al.** Top 10 plant pathogenic bacteria in
 404 molecular plant pathology. *Mol Plant Pathol* 2012; 13:614-629.
- 405 4. **Adeolu M, Alnajjar S, Naushad S, Gupta RS.** Genome-based phylogeny and taxonomy of the
 406 '*Enterobacteriales*': proposal for *Enterobacterales* ord. nov. divided into the families
 407 *Enterobacteriaceae*, *Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov.,
 408 *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. *Int J Syst Evol Microbiol*
 409 2016; 66: 5575-5599.
- 410 5. **Winslow CE, Broadhurst J, Buchanan RE, Krumwiede C, Rogers LA, Smith GH.** The families and genera
 411 of the bacteria. Final report of the Committee of the Society of American Bacteriologists on
 412 characterization and classification of bacterial types. *J Bacteriol* 1920; 5:191-229.
- 413 6. **Young JM, Dye DW, Bradbury JF, Panagopoulos CG, Robbs CF.** A proposed nomenclature and
 414 classification for plant-pathogenic bacteria. *N Z J Agric Res* 1978; 21:153-177.
- 415 7. **Lelliott RA, Dickey RS.** Genus VII. *Erwinia*. In *Bergey's Manual of Systematic Bacteriology*. Krieg NR and
 416 Holt JG (eds). Williams & Wilkins, Baltimore. 1984; vol. 1, pp. 469-476.
- 417 8. **Victoria J I, Barros O.** Etiologia de una nueva enfermedad bacterial del platano (*Musa paradisiaca* L.) en
 418 Colombia. *Inst Colomb Agropecu Revista ICA* 1969 4:173-190.

- 419 9. **Brenner DJ, Steigerwalt AG, Miklos GV, Fanning GR.** Deoxyribonucleic acid relatedness among
 420 erwiniae and other *Enterobacteriaceae*: the soft-rot organisms (genus *Pectobacterium* Waldee). *Int J*
 421 *Syst Bacteriol* 1973; 23:205–216.
- 422 10. **Hauben L, Moore ERB, Vauterin L, Steenackers M, Mergaert J et al.** Phylogenetic position of
 423 phytopathogens within the *Enterobacteriaceae*. *Syst Appl Microbiol* 1998; 21:384–397.
- 424 11. **Samson R, Legendre JB, Christen R, Fischer-Le Saux M, Achouak W et al.** Transfer of *Pectobacterium*
 425 *chrysanthemi* (Burkholder et al. 1953) Brenner et al. 1973 and *Brenneria paradisiaca* to the genus
 426 *Dickeya* gen. nov. as *Dickeya chrysanthemi* comb. nov. and *Dickeya paradisiaca* comb. nov. and
 427 delineation of four novel species, *Dickeya dadantii* sp. nov., *Dickeya dianthicola* sp. nov., *Dickeya*
 428 *dieffenbachiae* sp. nov. and *Dickeya zae* sp. nov. *Int J Syst Evol Microbiol* 2005; 55:1415–1427.
- 429 12. **Brady CL, Cleenwerck I, Denman S, Venter SN, Rodriguez-Palenzuela P et al.** Proposal to reclassify
 430 *Brenneria quercina* (Hildebrand & Schroth 1967) Hauben et al. 1999 into a novel genus, *Lonsdalea* gen.
 431 nov., as *Lonsdalea quercina* comb. nov., descriptions of *Lonsdalea quercina* subsp. *quercina* comb. nov.,
 432 *Lonsdalea quercina* subsp. *Iberica* subsp. nov., and *Lonsdalea quercina* subsp. *britannica* subsp. nov.,
 433 emendation of the description of the genus *Brenneria*, reclassification of *Dickeya dieffenbachiae* as
 434 *Dickeya dadantii* subsp. *dieffenbachiae* comb. nov., and emendation of the description of *Dickeya*
 435 *dadantii*. *Int J Syst Evol Microbiol* 2012; 62:1592-1602.
- 436 13. **Slawiak M, van Beckhoven JRCM, Speksnijder AGCL, Czajkowski R, Grabe G et al.** Biochemical and
 437 genetical analysis reveal a new clade of biovar 3 *Dickeya* spp. strains isolated from potato in Europe.
 438 *Eur J Plant Pathol* 2009; 125:245–261.
- 439 14. **van der Wolf JM, Nijhuis EH, Kowalewska MJ, Saddler GS, Parkinson N et al.** *Dickeya solani* sp. nov., a
 440 pectinolytic plant-pathogenic bacterium isolated from potato (*Solanum tuberosum*). *Int J Syst Evol*
 441 *Microbiol* 2014; 64:768–774.
- 442 15. **Tian Y, Zhao Y, Yuan X, Yi J, Fan J et al.** *Dickeya fangzhongdai* sp. nov., a plant-pathogenic bacterium
 443 isolated from pear trees (*Pyrus pyrifolia*). *Int J Syst Evol Microbiol* 2016; 66:2831–2835.
- 444 16. **Alič Š, Van Gijsegem F, Pédrón J, Ravnikar M, Dreó T.** Diversity within the novel *Dickeya fangzhongdai*
 445 sp., isolated from infected orchids, water and pears. *Plant Pathol* 2018; 67:1612–1620.
- 446 17. **Hugouvieux-Cotte-Pattat N, Brochier-Armanet C, Flandrois JP, Sylvie Reverchon S.** *Dickeya*
 447 *poaceaphila* sp. nov., a plant-pathogenic bacterium isolated from sugar cane (*Saccharum officinarum*).
 448 *Int J Syst Evol Microbiol* 2020; 70: 4508-4514. doi:10.1099/ijsem.0.004306
- 449 18. **Wang X, He SW, Guo HB, Han JG, Thin KK, et al.** *Dickeya oryzae* sp. nov., isolated from the roots of
 450 rice. *Int J Syst Evol Microbiol* 2020; 70: 4171-4178. doi:10.1099/ijsem.0.004265

- 451 19. **Parkinson N, DeVos P, Pirhonen M, Elphinstone J.** *Dickeya aquatica* sp. nov., isolated from waterways.
452 *Int J Syst Evol Microbiol* 2014; 64: 2264–2266.
- 453 20. **Hugouvieux-Cotte-Pattat N, Jacot-des-Combes C, Briolay J.** *Dickeya lacustris* sp. nov., a water-living
454 pectinolytic bacterium isolated from lakes in France. *Int J Syst Evol Microbiol* 2019; 69:721-726.
- 455 21. **Oulghazi S, Pédrón J, Cigna J, Lau YY, Mounni M et al.** *Dickeya undicola* sp. nov., a novel species for
456 pectinolytic isolates from surface waters in Europe and Asia. *Int J Syst Evol Microbiol* 2019; 69:2440-
457 2444.
- 458 22. **Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK.** Genomics and taxonomy in diagnostics
459 for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods* 2016; 8:12-24.
- 460 23. **Duprey A, Taib N, Leonard S, Garin T, Flandrois JP et al.** The phytopathogenic nature of *Dickeya*
461 *aquatica* 174/2 and the dynamic early evolution of *Dickeya* pathogenicity. *Environ Microbiol* 2019;
462 21:2809-2835.
- 463 24. **Pédrón J, Van Gijsegem F.** Diversity in the bacterial *Dickeya* genus grouping plant pathogens and
464 waterways isolates. *OBM Genetics* 2019; 3(4):22. doi:10.21926
- 465 25. **Dickey RS, Victoria JI.** Taxonomy and emended description of strains of *Erwinia* isolated from *Musa*
466 *paradisiaca* Lineaus. *Int J Syst Bacteriol* 1980; 30:129-134.
- 467 26. **Pritchard L, Humphris S, Saddler GS, Elphinstone JG, Pirhonen M, Toth IK.** Draft genome sequences of
468 17 isolates of the plant pathogenic bacterium *Dickeya*. *Genome Announc* 2013; 1(6):e00978-13.
469 doi:10.1128/genomeA.00978-13
- 470 27. **Fernandez-Borrero O, Lopez-Duques S.** Pudricion acuosa del susedo tallo del platana (*Musa*
471 *paradisiaca*) causade por *Erwinia paradisiaca* n. sp. *Cenicafe* 1970; 21:3-44.
- 472 28. **Marrero G, Schneider KL, Jenkins DM, Alvarez AM.** Phylogeny and classification of *Dickeya* based on
473 multilocus sequence analysis. *Int J Syst Evol Microbiol* 2013; 63:3524-3539. doi: 10.1099/ijls.0.046490-0.
- 474 29. **Cigna J, Dewaegeneire P, Beury A, Gobert V, Faure D.** A *gapA* PCR sequencing assay for identifying the
475 *Dickeya* and *Pectobacterium* potato pathogens. *Plant Disease* 2017; 101:1278-1282.
- 476 30. **Saitou N, Nei M.** The neighbour-joining method: A new method for reconstructing phylogenetic trees.
477 *Mol Biol Evol* 1987; 4:406-425.
- 478 31. **Barzic MR, Samson R, Trigalet A.** Pourriture bactérienne de la tomate cultivée en serre. *Ann*
479 *Phytopathol* 1976; 8:237-240.
- 480 32. **Bochner BR.** Global phenotypic characterization of bacteria. *FEMS Microbiol Rev* 2009; 33:191-205.
- 481 33. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al.** SPAdes: a new genome assembly
482 algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 2012, **19**: 455-477.

- 483 34. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T et al.** The RAST server: rapid annotations using
484 subsystems technology. *BMC Genomics* 2008; 9:75.
- 485 35. **Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM.** DNA–DNA
486 hybridization values and their relationship to whole genome sequence similarities. *Int J Syst Evol*
487 *Microbiol* 2007; 57: 81-91.
- 488 36. **Richter M, Rosselló-Móra R.** Shifting the genomic gold standard for the prokaryotic species definition.
489 *Proc Natl Acad Sci U S A.* 2009; 106:19126-19131.
- 490 37. **Meier-Kolthoff JPM, Auch AF, Klenk HP, Göker M.** Genome sequence-based species delimitation with
491 confidence intervals and improved distance functions. *BMC Bioinformatics* 2013; 14:60.
- 492 38. **Barco RA, Garrity GM, Scott JJ, Amend JP, Neilson KH, Emerson D.** A genus definition for bacteria and
493 archaea based on a standard genome relatedness index. *mBio.* 2020; 11(1):e02475-19. doi:
494 10.1128/mBio.02475-19.
- 495 39. **Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A.**
496 Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* 2015; 43:6761-6771.
497 doi: 10.1093/nar/gkv657.
- 498 40. **Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S.** High throughput ANI analysis of 90K
499 prokaryotic genomes reveals clear species boundaries. *Nat Commun.* 2018; 9(1):5114. doi:
500 10.1038/s41467-018-07641-9.
- 501 41. **Murray CS., Gao Y, Wu M.** There is no evidence of a universal genetic boundary among microbial
502 species. *bioRxiv* 2020. 07.27.223511; doi: <https://doi.org/10.1101/2020.07.27.223511>
- 503 42. **Olm MR, Crits-Christoph A, Diamond S, Lavy A, Matheus Carnevali PB, Banfield JF.** Consistent
504 metagenome-derived metrics verify and delineate bacterial species boundaries. *mSystems* 2020;
505 5(1):e00731-19. doi: 10.1128/mSystems.00731-19
- 506 43. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** Prodigal: prokaryotic gene recognition
507 and translation initiation site identification. *BMC Bioinformatics* 2010; 11:119. doi: 10.1186/1471-2105-
508 11-119.
- 509 44. **Emms DM; Kelly S.** OrthoFinder: phylogenetic orthology inference for comparative genomics. [Genome](#)
510 [Biol](#) 2019; [20:238](#)
- 511 45. **Nakamura T, Yamada KD, Tomii K, Katoh K.** Parallelization of MAFFT for large-scale multiple sequence
512 alignments. *Bioinformatics* 2018; 34:2490–2492
- 513 46. **Notredame C, Higgins DG, Heringa J.** T-Coffee: A novel method for multiple sequence alignments. *J*
514 *Mol Biol* 2000; 30:205-217.

- 515 47. **Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A.** RAxML-NG: A fast, scalable, and user-friendly
516 tool for maximum likelihood phylogenetic inference. *Bioinformatics* 2019; btz305
517 [doi:10.1093/bioinformatics/btz305](https://doi.org/10.1093/bioinformatics/btz305)
- 518 48. **Hugouvieux-Cotte-Pattat N, Condemine G, Shevchik VE.** Bacterial pectate lyases, structural and
519 functional diversity. *Environ Microbiol Rep* 2014; 6:427-440.
- 520 49. **Reverchon S, Nasser W.** *Dickeya* ecology, environment sensing and regulation of virulence programme.
521 *Environ Microbiol Rep* 2013; 5:622–636.
- 522 50. **Collmer A, Gold S.** Noel T. Keen-pioneer leader in molecular plant pathology. *Annu Rev Phytopathol*
523 2007; 45:25-42.
-

524

525

CONFIDENTIAL

526 FIGURES AND TABLES

527

528 **Fig. 1.** Phylogenetic position of A3967 and different strains based on *gapA* gene sequences.

529 This analysis was performed using available *gapA* gene sequences from type strains of *Dickeya* species, and
 530 sequences of the *gapA* PCR product (represented by a circle) for the laboratory strains A1816, A3967, A4507,
 531 A6358, A6375 and A6065^T. Type strains of *Pectobacterium carotovorum*, *P. atrosepticum* and *P. wasabiae* were
 532 used as outgroups. The evolutionary history was inferred using the neighbour-joining method [30], with
 533 bootstrap support values indicated (1000 bootstrap replicates). The evolutionary distances were computed
 534 using the maximum composite likelihood method and are in the units of the number of base substitutions per
 535 site (718 positions). Evolutionary analyses were conducted in MEGA X version 10.2.4.

536

537

538

539 **Fig. 2.** ANIm percentage identity and coverage for *Dickeya* type strains.

540 Heatmaps of (a) ANIm identity and (b) ANIm coverage for strains A3967, CFBP 4178^T, Ech703, the type strains
 541 of eleven *Dickeya* species, and for *Pectobacterium carotovorum* CFBP 2046^T.

542 In (a) pairwise comparisons with >95% identity are filled red; comparisons with <95% identity are filled blue;
 543 comparisons with ≈95% identity are filled white (approximating a species boundary). Red blocks along the
 544 diagonal indicate each type strain represents a distinct species, and strain A3967 shares 96% identity with the
 545 *D. paradisiaca* type strain CFBP 4178^T. A complete heatmap showing ANIm values for all publicly available
 546 *Dickeya* genomes is given in Fig. S3b.

547 In (b) pairwise comparisons with >50% coverage (also known as "alignment fraction") are filled red;
 548 comparisons with <50% coverage are filled blue; comparisons with ≈50% coverage are filled white
 549 (approximating a genus boundary). The coherent red blocks imply that *P. carotovorum* CFBP 2046^T belongs to a
 550 discrete genus, distinct from the *Dickeya* genomes. Likewise, the three *D. paradisiaca* genomes belong to a
 551 discrete genus distinct from the other *Dickeya* genomes. The *D. lacustris* and *D. aquatica* type strains might
 552 also belong to a distinct genus, separate from the other *Dickeya*. The remaining eight *Dickeya* species appear to
 553 constitute a single coherent genus group sharing at least 61% of their complete genomes in homologous
 554 pairwise alignment. A complete heatmap showing ANIm coverage for all publicly available *Dickeya* genomes is
 555 given in Fig. S3b.

556

557 **Fig. 3. Multigene phylogenetic reconstruction on a set of 49 genomes spanning five genera in the**
558 ***Pectobacteriaceae***

559 Maximum-likelihood phylogenetic reconstruction obtained using raxml-ng for 49 Pectobacteriaceae genomes,
560 constructed from 1201 single-copy orthologues shared by all genomes. The GTR+F0+G4m+B model was used
561 and parametrized separately for each orthologue. The best-fit topology shown was obtained for each of 20
562 distinct starting trees and is likely to be the global optimum; support for internal bipartitions was obtained
563 using 100 bootstraps. The topology shown was midpoint-rooted, manually annotated and coloured using
564 figtree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>). The tree supports reassignment of *D. paradisiaca* to
565 genus level, with proposed name *Musicola*.

566

567

CONFIDENTIAL

568

569 **Table 1.** Carbon assimilation

570 Strains were inoculated onto M63 plates supplemented with a sole carbon source (2 g l⁻¹). The sign - indicates
 571 no growth after 72h at 30°C; +, indicates growth at 24h; w, indicates weak growth (visible after 48 or 72h).

572

	A3967 ^T	A6065 ^T	A1816	A4507	A6358	A6375	<i>D. dadantii</i>
	CFBP 8732 ^T	CFBP 4178 ^T	CFBP 1445	CFBP 3477	CFBP 3696	CFBP 3699	3937
577 D-arabinose	w*	w*	w	w	w	w	w
578 L-arabinose	+	+	+	+	+	+	+
579 D-galactose	+	+	+	+	+	+	+
580 D-glucose	+	+	+	+	+	+	+
581 D-fructose	+	+	+	+	+	+	+
582 D-mannose	+	+	+	+	+	+	+
583 L-rhamnose	-	-	-	-	-	-	-
584 D-ribose	+	+	+	+	+	+	+
585 D-xylose	+	+	+	+	+	+	+
586 D-galacturonate	+	+	+	+	+	+	+
587 D-glucuronate	+	+	+	+	+	+	-
588 D-cellobiose	-	-	-	-	-	-	+
589 D-melibiose	-	+	+	+	+	+	+
590 D-raffinose	-	+	+	+	+	+	+
591 Sucrose	+	+	+	+	+	+	+
592 Glycerol	+	+	+	+	+	+	+
593 D-mannitol	-	-	-	-	-	-	+
594 myo-Inositol	+	-	-	-	-	-	+
595 Citrate	w	w*	w	w	w	w	w
596 L-lactate	w	w*	w	w	w	w	w
597 Polygalacturonate	+	+	+	+	+	+	+

598

599 *For these compounds, a discrepancy was observed with data from the Biolog plates PM1 or PM2A that gave a
 600 negative result (see Table S2).

601

602 **Table 2.** Enzyme secretion, motility and maceration ability

603 *D. dadantii* 3937 was used as a reference strain. The enzyme secretion was assessed on plates containing an
 604 enzyme substrate [20]: +, positive; -, negative. Motility was estimated in 0.3% L agar plate for swimming and on
 605 0.6% L agar plate for swarming [20]. The length of macerated tissue was measured 24 h after inoculation for
 606 chicory leaves and the weight of macerated tissue was measured after 48 h for potato tubers. For each
 607 measurement, the mean value is given with the standard deviation.

	A3967 ^T CFBP 8732 ^T	A6065 ^T CFBP 4178 ^T	A1816 CFBP 1445	A4507 CFBP 3477	A6358 CFBP 3696	A6375 CFBP 3699	<i>D. dadantii</i> 3937
Pectinase secretion	+	+	+	+	+	+	+
Cellulase secretion	+	+	+	+	+	+	+
Protease secretion	-	-	-	-	-	-	+
Lipase secretion	-	-	-	-	-	-	+
Swimming motility (mm)	6±3	14±2	5±1	11±4	16±2	1±0	19±4
Swarming motility (mm)	38±6	7±2	6±2	41±8	37±6	8±3	51±9
Chicory leaf maceration (mm)	10.7±1.9	7.2±3.2	2.2±2.2	9.3±2.9	9±1.9	6.1±3.8	37.7±8.6
Potato tuber maceration (g)	0.43±0.09	0.16±0.07	0.06±0.01	0.11±0.06	0.21±0.08	0.26±0.12	1.68±0.52

633 **Table 3.** ANI and dDDH values between *Dickeya* type strains and *Musicola* strains A3967^T (CFBP 8732^T), CFBP
634 4178^T and Ech703

635 The left lower triangle displays the ANI values (%) and the right-upper triangle displays the dDDH values. The
636 ANI values were calculated based on pairwise comparisons between A3967^T (CFBP 8732^T) and other genomes
637 (<http://enve-omics.ce.gatech.edu/ani/>) [36]. The dDDH values were calculated using the A3967 genome as a
638 reference and other *Dickeya* genomes as queries (<http://ggdc.dsmz.de/>) [37]. This analysis was performed
639 using the genomes of strains CFBP 4178^T (GCA_000400505.1), Ech703 (GCA_000023545.1), *D. aquatica* CFBP
640 8348^T (GCA_900095885.1), *D. chrysanthemi* CFBP 2048^T (GCA_000406105.1), *D. dadantii* CFBP 1269^T
641 (GCA_003049785.1), *D. dianthicola* CFBP 1200^T (GCA_000365305.1), *D. fangzhongdai* CFBP 8607^T
642 (GCA_002812485.1), *D. lacustris* S29^T (GCA_003934295.1), *D. poaceiphila* CFBP 8731^T (GCA_007858975.2), *D.*
643 *solani* CFBP 7345^T (GCA_001644705.1), *D. undicola* CFBP 8650^T (GCA_000784735.1), *D. zeae* CFBP 2052^T
644 (GCA_000406165.1) and *P. carotovorum* CFBP 2046^T (GCA_000749855.1).

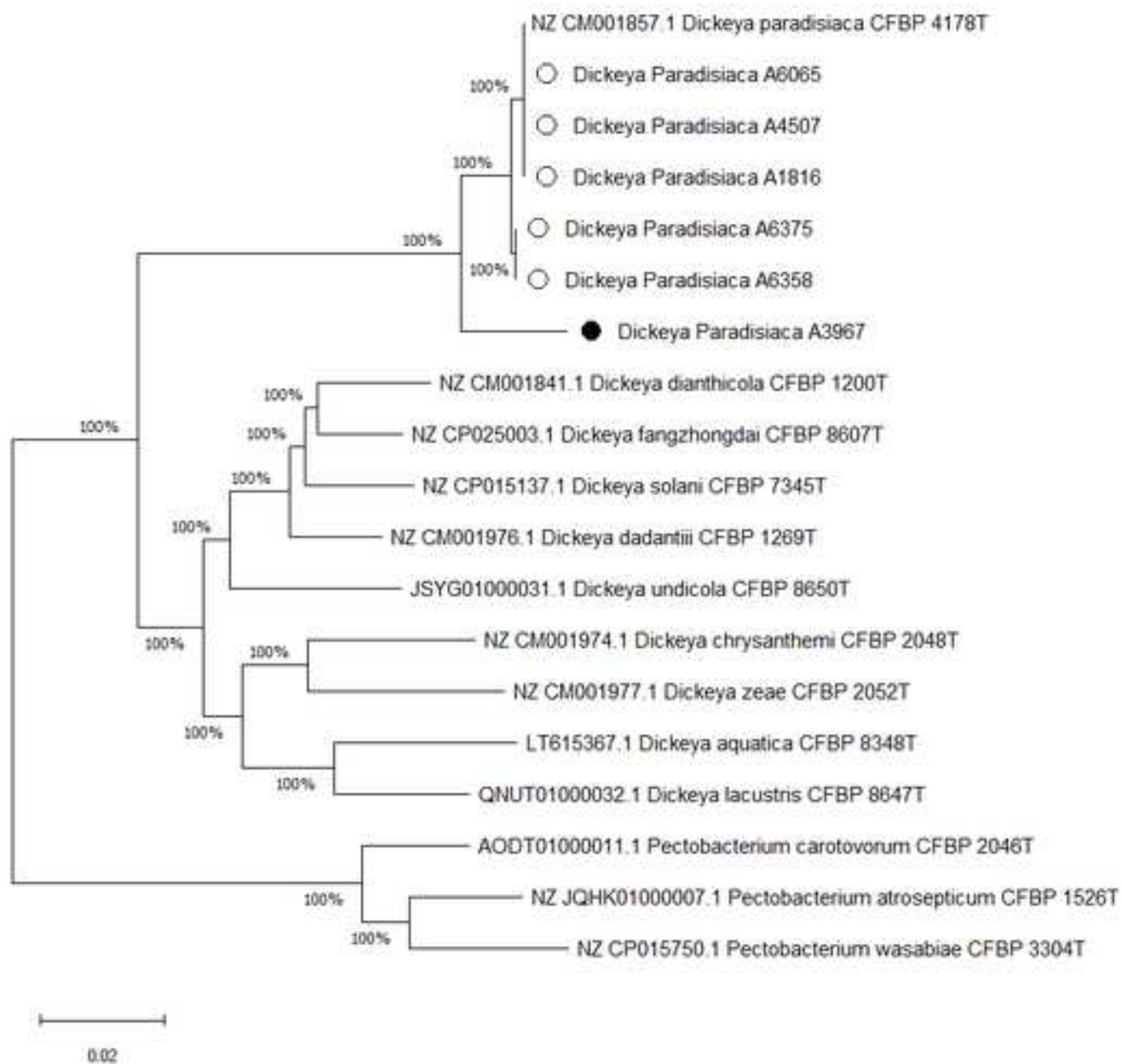
645 A table showing ANI identity values for all publicly available *Dickeya* genomes is given in Fig. S3a.

dDDH

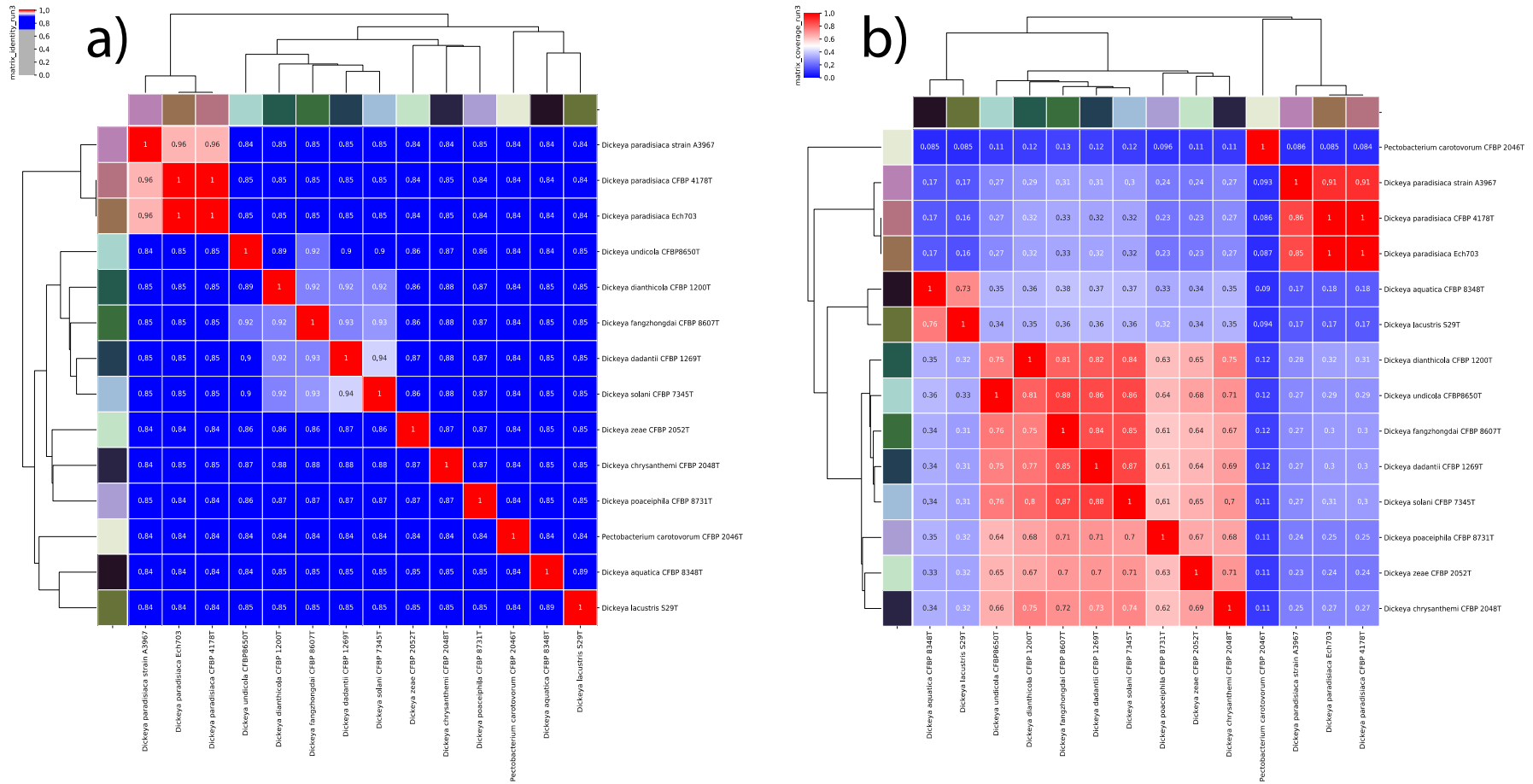
	<i>Dickeya zeae</i> CFBP 2052 ^T	<i>Dickeya chrysanthemi</i> CFBP 2048 ^T	<i>Dickeya poaceiphila</i> NCPPB 569 ^T	<i>Dickeya fangzhongdai</i> CFBP 8607 ^T	<i>Dickeya dianthicola</i> CFBP 1200 ^T	<i>Dickeya solani</i> CFBP 7345 ^T	<i>Dickeya dadantii</i> CFBP 1269 ^T	<i>Dickeya undicola</i> CFBP 8650 ^T	<i>Dickeya lacustris</i> CFBP 8647 ^T	<i>Dickeya aquatica</i> CFBP 8348 ^T	Ech 703	CFBP 4178 ^T	A3967 ^T (CFBP 8732 ^T)
ANI <i>Dickeya zeae</i> CFBP 2052 ^T	100	31.2	29.5	28.8	29.0	28.7	28.6	27.6	23.3	23.6	22.0	22.0	22.1
<i>Dickeya chrysanthemi</i> CFBP 2048 ^T	86.57	100	30.5	31.8	32.2	31.5	32.3	29.1	23.3	23.4	22.4	22.4	22.2
<i>Dickeya poaceiphila</i> NCPPB 569 ^T	85.55	86.10	100	30.9	30.6	30.5	30.6	28.5	22.9	23.0	22.4	22.4	22.3
<i>Dickeya fangzhongdai</i> CFBP 8607 ^T	85.09	86.68	86.00	100	45.9	48.0	46.7	44.5	23.4	23.7	23.9	23.5	23.1
<i>Dickeya dianthicola</i> CFBP 1200 ^T	85.23	87.12	86.09	92.07	100	46.3	45.7	36.1	23.4	23.4	23.5	23.5	22.9
<i>Dickeya solani</i> CFBP 7345 ^T	85.02	86.68	86.08	92.62	92.15	100	54.7	37.4	23.2	23.5	23.5	23.2	23.1
<i>Dickeya dadantii</i> CFBP 1269 ^T	84.93	87.03	86.15	92.43	91.94	94.07	100	37.1	23.2	23.2	23.3	23.3	22.8
<i>Dickeya undicola</i> CFBP 8650 ^T	84.31	85.53	84.90	91.66	89.01	89.36	89.26	100	22.8	23.0	22.6	22.5	22.3
<i>Dickeya lacustris</i> CFBP 8647 ^T	79.63	79.89	79.59	80.40	80.03	79.82	79.87	79.73	100	32.9	21.8	21.5	21.3
<i>Dickeya aquatica</i> CFBP 8348 ^T	80.21	80.34	80.09	80.50	80.57	80.39	80.42	80.11	87.45	100	22.0	21.5	21.5
Ech 703	78.30	79.04	78.42	80.00	79.82	79.77	79.61	79.05	77.04	77.43	100	100	68.4
CFBP 4178^T	78.25	79.17	78.46	80.02	79.82	79.82	79.69	78.69	77.12	77.20	99.98	100	68.3
A3967^T (CFBP 8732^T)	78.17	78.71	78.54	79.35	79.24	79.42	79.44	78.50	76.78	77.20	96.17	96.21	100
<i>Pectobacterium carotovorum</i> CFBP 2046 ^T	75.16	75.43	75.34	75.85	75.78	75.84	75.73	75.51	74.77	74.78	75.07	74.99	74.89

CONFIDENTIAL

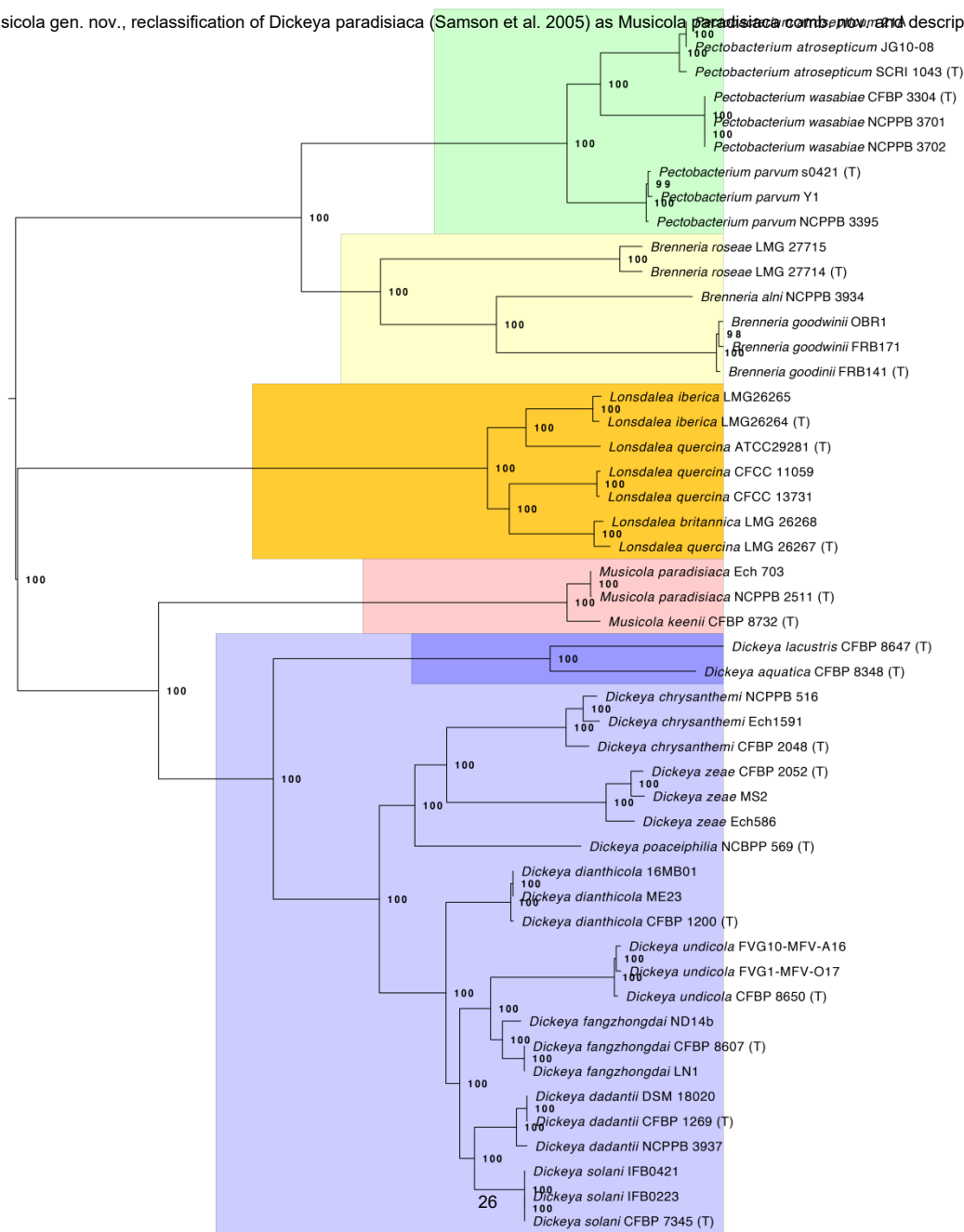
Proposal for the creation of a new genus *Musicola* gen. nov., reclassification of *Dickeya paradisiaca* (Samson et al. 2005) as *Musicola paradisiaca* comb. nov. and description of a new species *Musicola keenii* sp. nov.



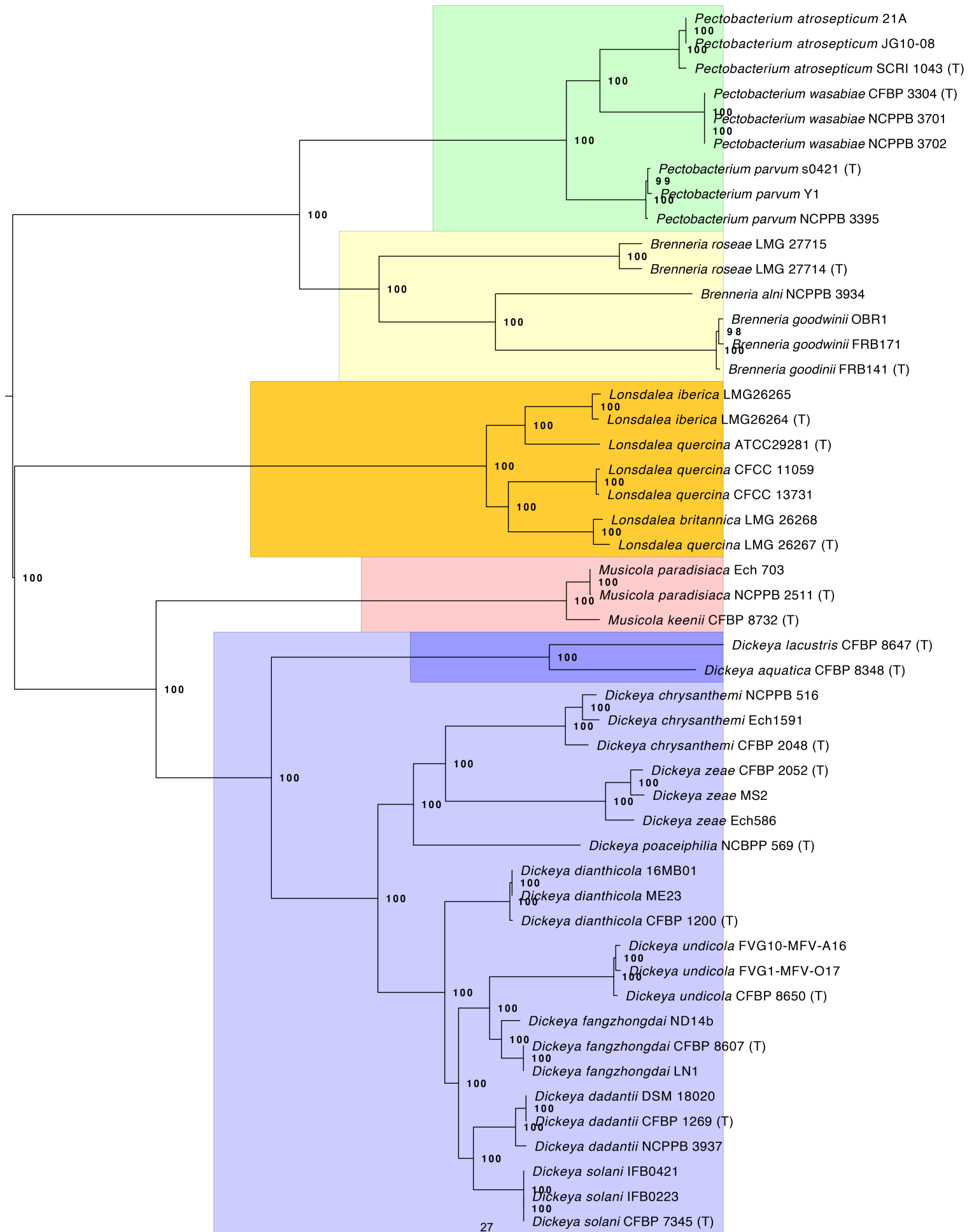
Proposal for the creation of a new genus *Musicola* gen. nov., reclassification of *Dickeya paradisiaca* (Samson et al. 2005) as *Musicola paradisiaca* comb. nov. and description of a new species *Musicola keenii* sp. nov.



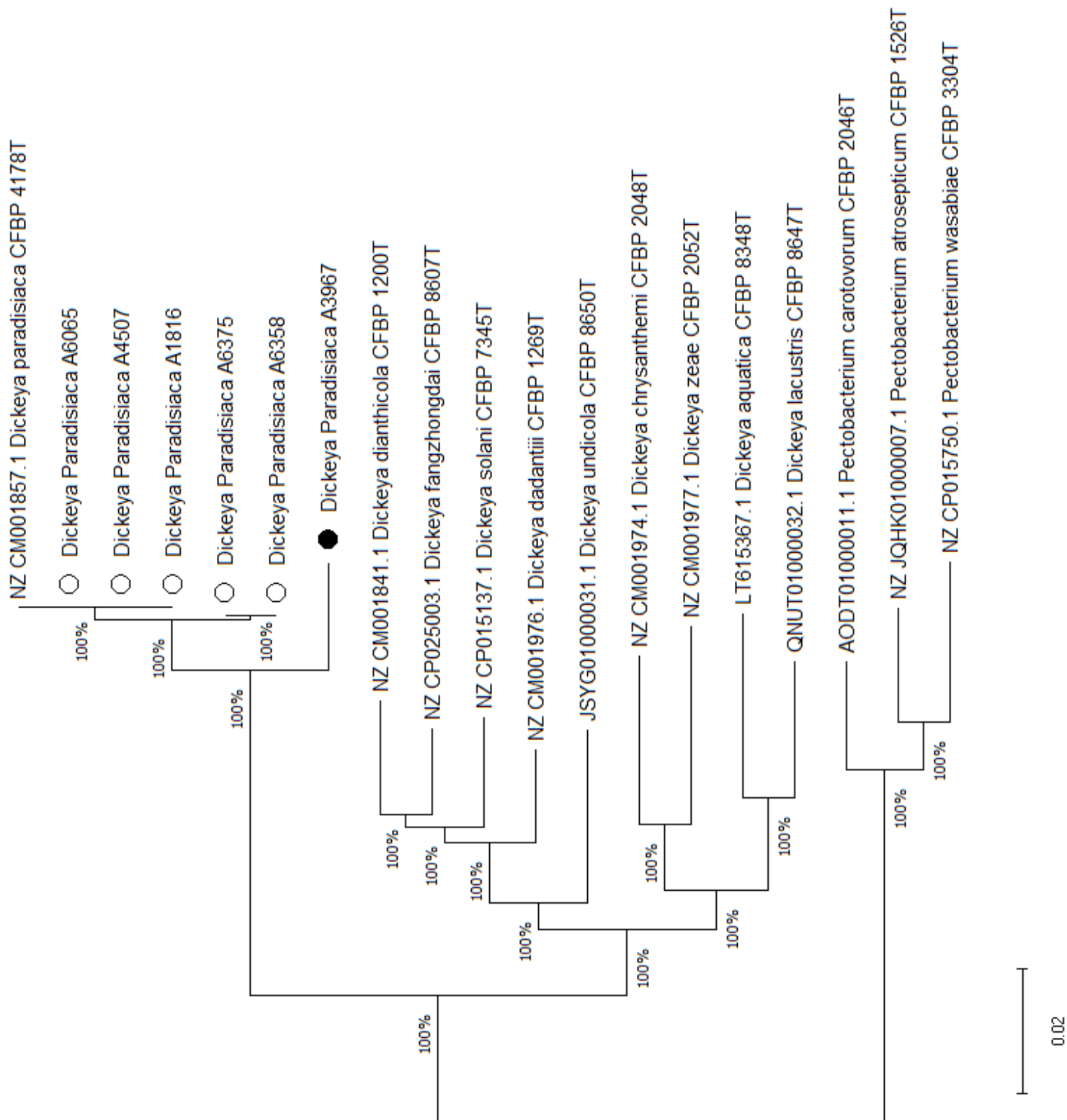
Proposal for the creation of a new genus *Musicola* gen. nov., reclassification of *Dickeya paradisiaca* (Samson et al. 2005) as *Musicola paradisiaca* comb. nov. and description of a new species *Musicola keenii* sp. nov.



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*



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 keeni*



Supplementary data

Fig. S1. Phylogenetic position of A3967 and different strains based on 16S rRNA gene sequences.

This analysis was performed using 16S rRNA gene sequences from type strains of *Dickeya* species, of the strains CFBP 4178^T, CFBP3477 and Ech703, and sequences of the PCR product (represented by a circle) for A1816, A3967, A6358 (CFBP 3696) and A6375 (CFBP 3699). The sequences of two related strains E353 and 572, registered as *Erwinia chrysanthemi*, were also included as they clustered with the 16S rRNA gene sequences of the studied strains. Strain E353 (EU684953.1) was isolated in China and strain 572 (= Dickey 141) (AF373200.1) was isolated from *Musa paradisiaca*. Type strains of *Pectobacterium carotovorum*, *P. atrosepticum* and *P. wasabiae* were also included. Phylogenetic trees were constructed using the neighbour-joining method, with bootstrap support values indicated (1000 bootstrap replicates). For DNA, the evolutionary distances (number of base substitutions per site) were computed using the maximum composite likelihood method (1548 positions). Evolutionary analyses were conducted in MEGA X version 10.2.4.

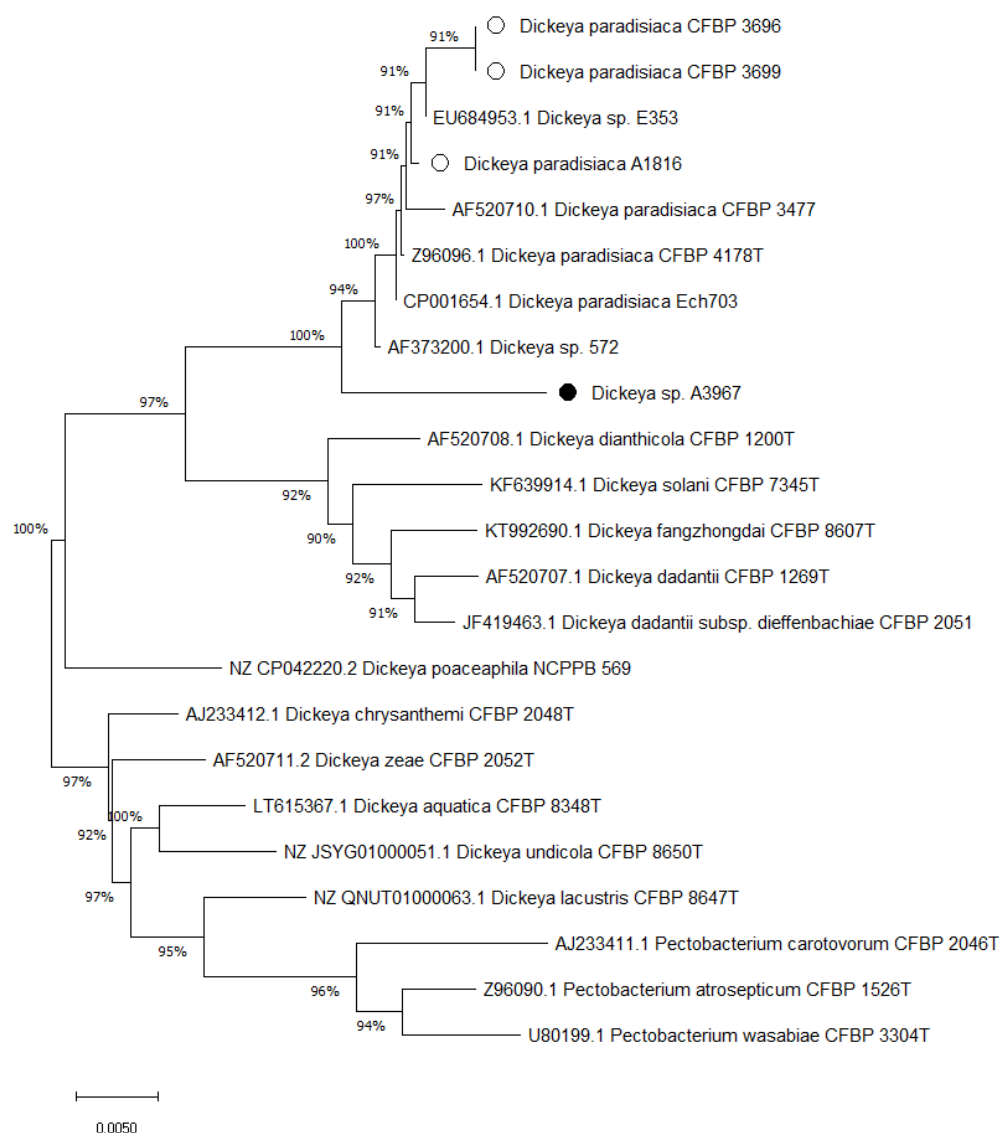
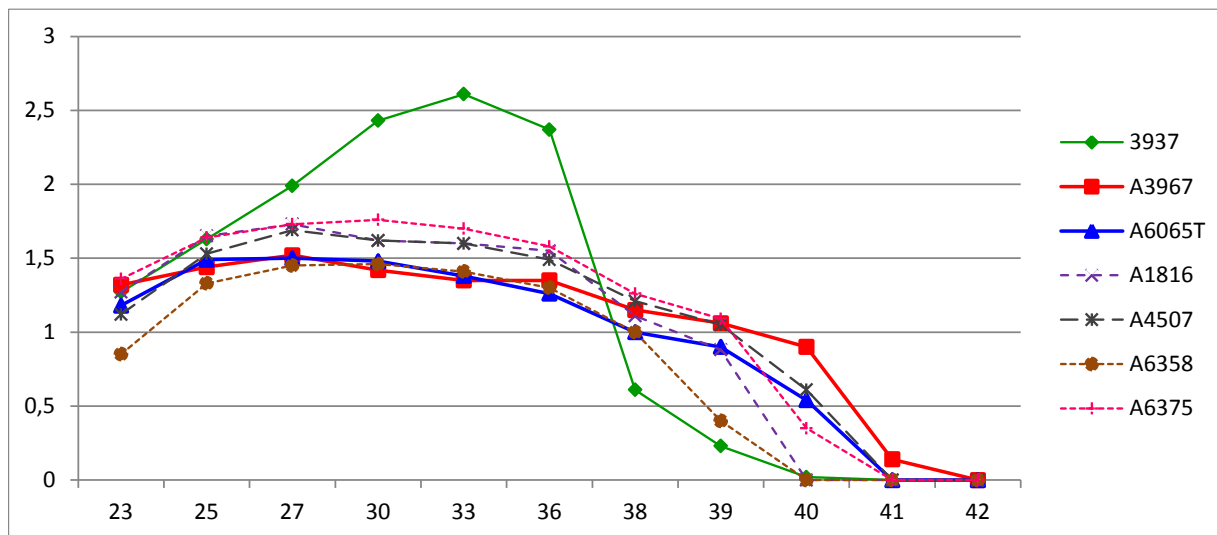


Fig S2. Determination of optimal and maximal growth temperatures of *Musicola* strains

To analyse the growth temperature, bacterial cultures were performed in LB medium in the range of 23 to 42°C. The cell density was estimated by measuring the optical density at 600 nm (OD_{600}) after 24 h. The optimal growth temperature was that giving the highest OD_{600} value, the maximal growth temperature was the highest temperature allowing for a significant growth ($OD_{600} > 0.1$). The *D. dadantii* strain 3937 was used for comparison. *M. paradisiaca* type strain A6065^T = CFBP 4178^T; *M. keenii* type strain A3967^T = CFBP 8732^T.



CONFIDENTIAL

Fig. S3. ANI percentage identity and coverage for 135 publicly-available genomes of *Dickeya* and *Brenneria*

(a) Heatmap of ANIm identity. Pairwise comparisons with >95% identity are filled red; comparisons with <95% identity are filled blue; comparisons with ≈95% identity are filled white. The red blocks along the diagonal indicate groups of genomes with at least 95% identity to all other members of the block and are taken to indicate discrete species groups. These support delineation of the following *Dickeya* species at ≈95% ANIm identity: *D. solani*, *D. dianthicola*, *D. dadantii*, *D. fangzhongdai*, *D. paradisiaca*, *D. aquatica*, *D. lacustris*, *D. zaeae*, *D. undicola*, *D. poaceiphilia*, and *D. chrysanthemi*.

(b) Heatmap of ANIm coverage. Pairwise comparisons with >50% coverage (also known as "alignment fraction") are filled red; comparisons with <50% identity are filled blue; comparisons with ≈50% identity are filled white. Red blocks along the diagonal indicate groups of genomes sharing at least 50% of their genome as recognizable homologous sequence alignment with all other members of that block. As described in the text, membership of the same coherent red block approximates membership of the same genus. Two genomes that are not members of the same block share less than 50% of their genome in homologous alignment and are considered to be members of distinct genera.

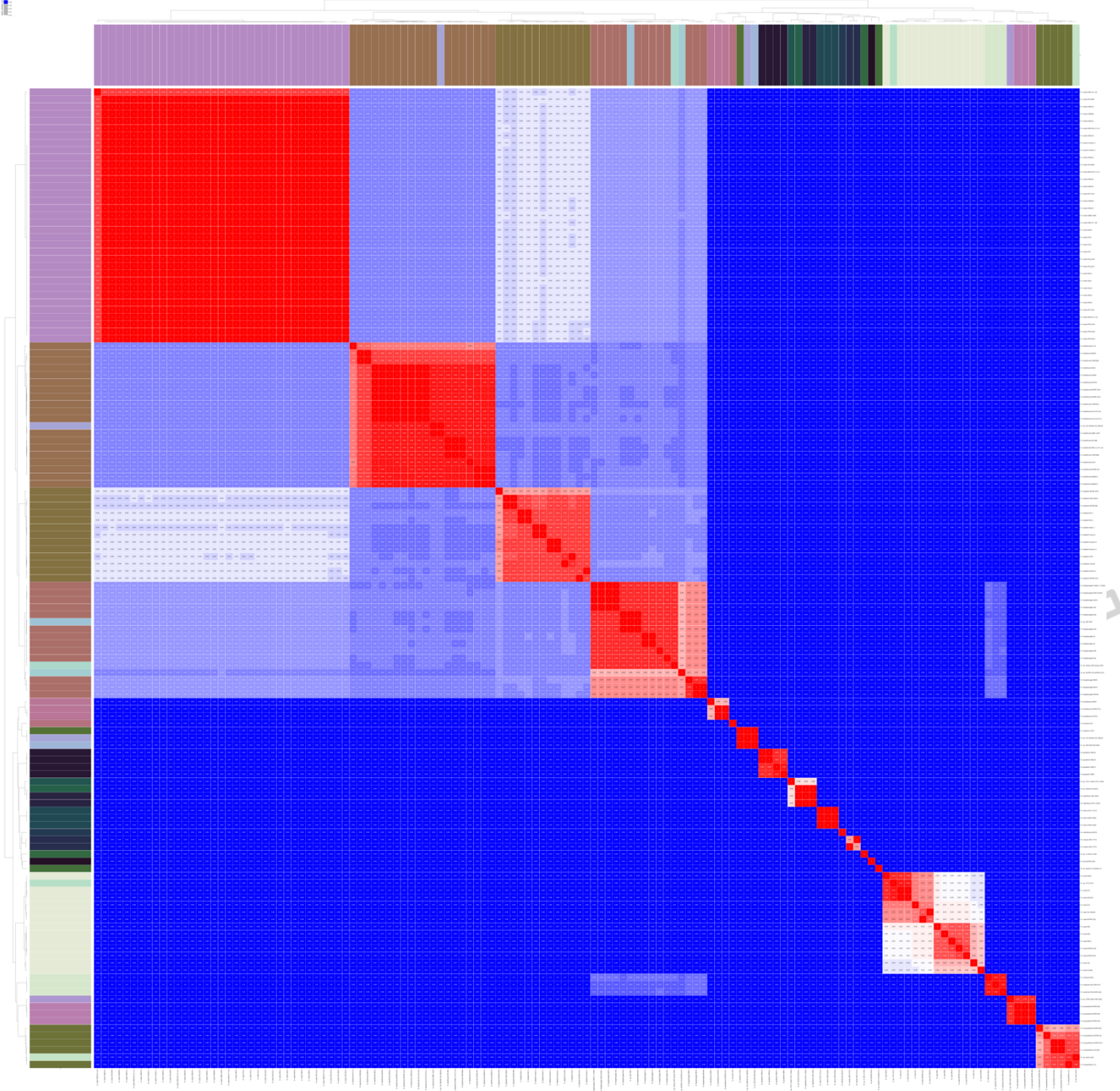
All *Dickeya* species groups except *D. paradisiaca*, *D. lacustris*, and *D. aquatica* form a single coherent red block, indicating that they belong to the same genus. *D. paradisiaca* genomes form a coherent red block distinct from all other *Dickeya* genomes (minimum coverage: 91%; maximum coverage with other *Dickeya* genomes: 33%), indicating that they constitute a distinct, discrete genus. *D. lacustris* and *D. aquatica* are members of the same red block (minimum coverage: 73%; maximum coverage with other *Dickeya* genomes: 37%), indicating that they can be considered members of the same genus, also distinct from *Dickeya*.

For a better visualization, heatmaps will be available online with the figures in full / arbitrary size which can be zoomed in to see the details.

https://github.com/widdowquinn/Sl_Hugovieux-Cotte-Pattat_2021/raw/main/figures/figure_S3_a.pdf

https://github.com/widdowquinn/Sl_Hugovieux-Cotte-Pattat_2021/raw/main/figures/figure_S3_b.pdf

CONFIDENTIAL



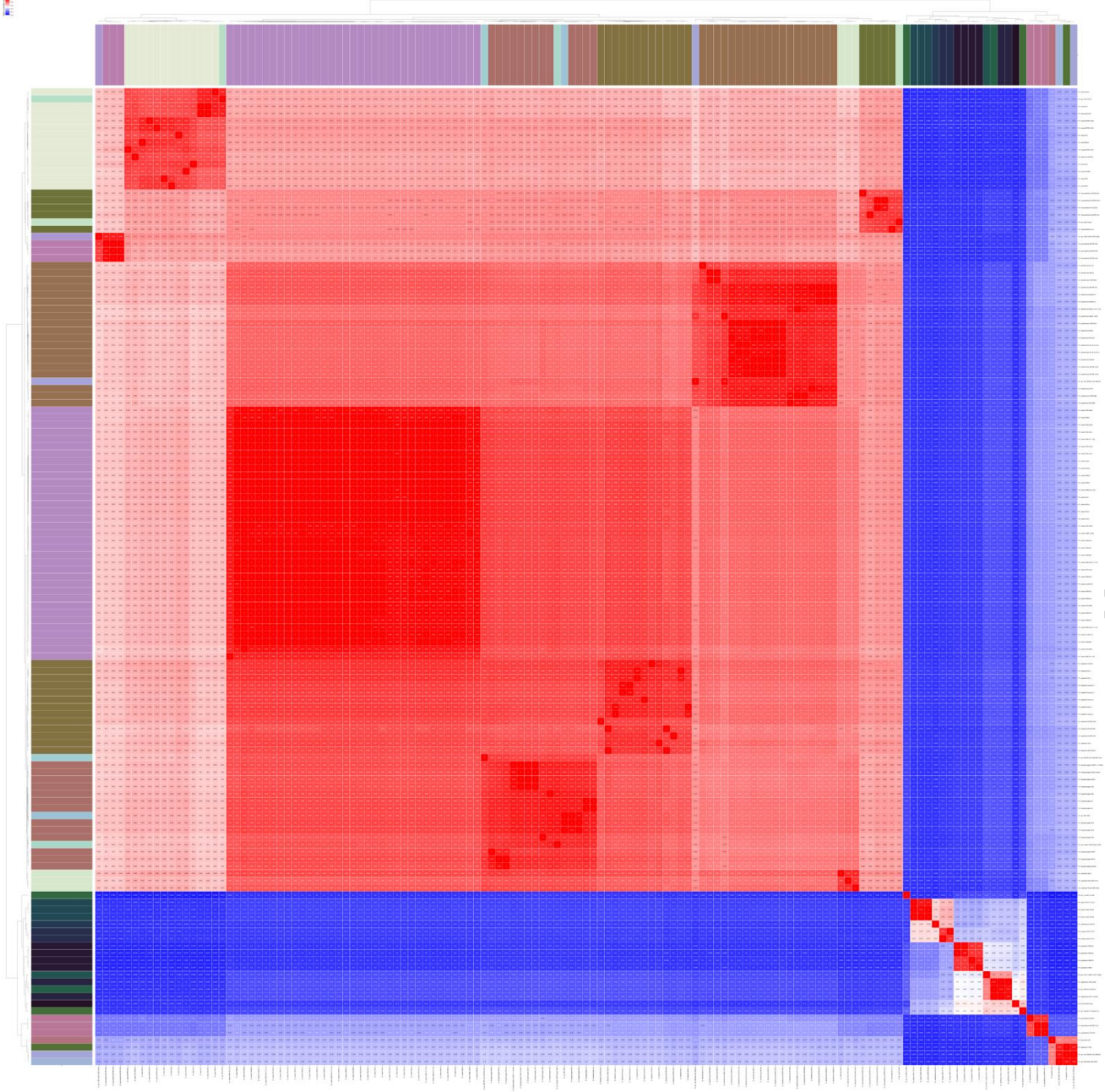


Fig. S4. Heatmaps of ANIm identity and ANIm coverage for whole-genome classification of 49 genomes spanning five genera in the *Pectobacteriaceae*.

Whole-genome classification using pyani v0.3.0b (ANIm).

(a) Heatmap of ANIm identity using 94-96% identity as an approximate threshold for species delineation. Pairwise comparisons with >95% identity are filled red; comparisons with <95% identity are filled blue; comparisons with ≈95% identity are filled white. The red blocks along the diagonal indicate groups of genomes with at least 95% identity to all other members of the block and are taken to indicate discrete species groups.

(b) Heatmap of ANIm coverage using 40-50% coverage for genus delineation. Pairwise comparisons with >50% coverage (also known as "alignment fraction") are filled red; comparisons with <50% coverage are filled blue; comparisons with ≈50% coverage are filled white. Red blocks along the diagonal indicate groups of genomes sharing at least 50% of their genome as recognizable homologous sequence alignment with all other members of that block. As described in the text, membership of the same coherent red block approximates membership of the same genus.

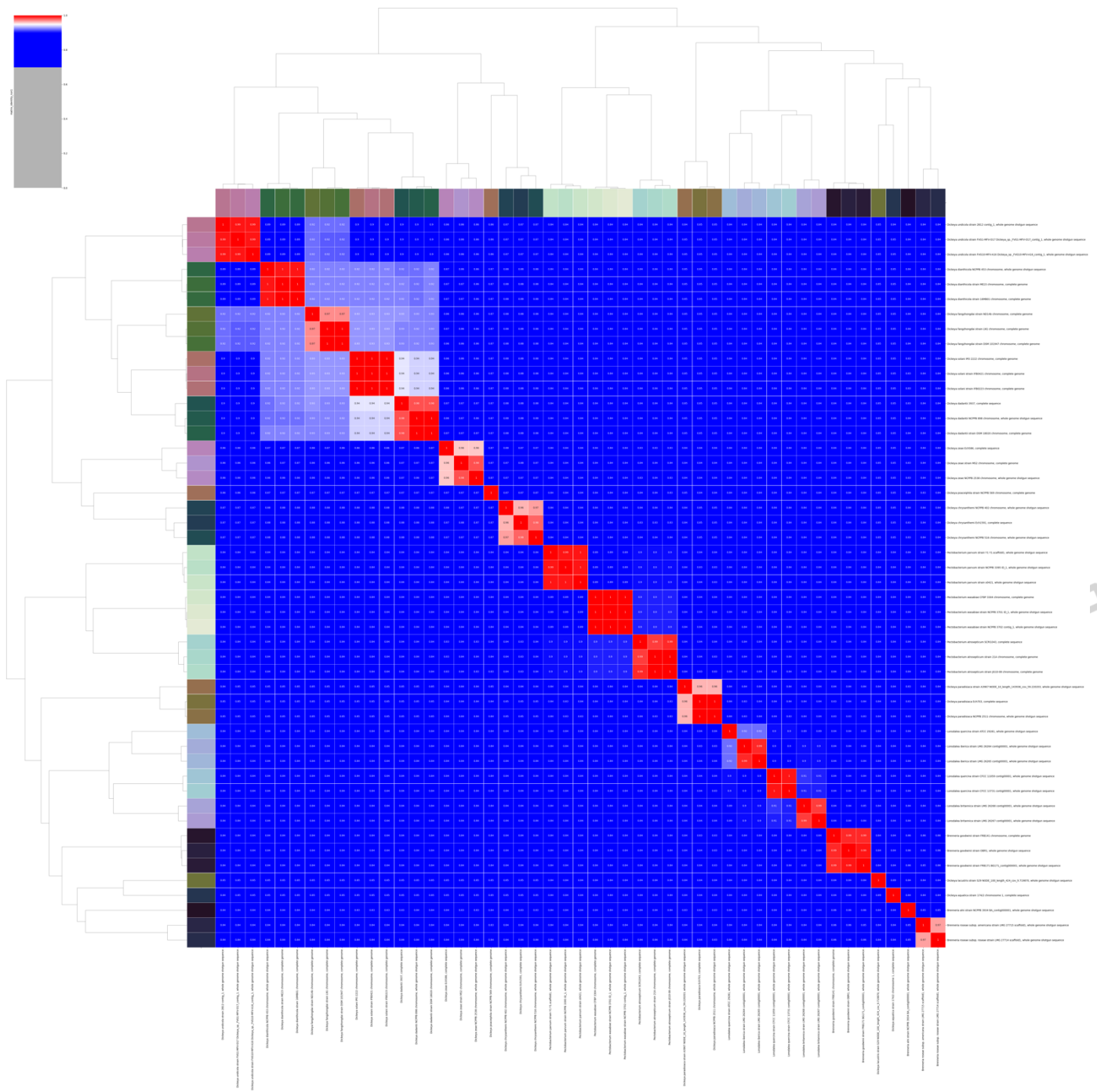
Taken together, the results support the following eight genus divisions: (1) *Dickeya* (*D. solani*, *D. dadantii*, *D. fangzhongai*, *D. undicola*, *D. dianthicola*, *D. poaceiphila*, *D. zea*, *D. chrysanthemi*); (2) *Musicola* (*M. paradisiaca*, *M. keenii*); (3) Gen. nov. I (*D. aquatica*, *D. lacustris*); *Lonsdalea* (*L. iberica*, *L. quercina*, *L. britannica*); *Pectobacterium* (*P. atrosepticum*, *P. wasabiae*, *P. parvum*); Gen. nov. II (*B. roseae*); Gen. nov. III (*B. alni*); Gen. nov. IV (*B. goodwinii*).

For a better visualization, heatmaps will be available online with the figures in full / arbitrary size which can be zoomed in to see the details.

https://github.com/widdowquinn/SI_Hugouvieux-Cotte-Pattat_2021/raw/main/figures/figure_S4_a.pdf

https://github.com/widdowquinn/SI_Hugouvieux-Cotte-Pattat_2021/raw/main/figures/figure_S4_b.pdf

CONFIDENTIAL



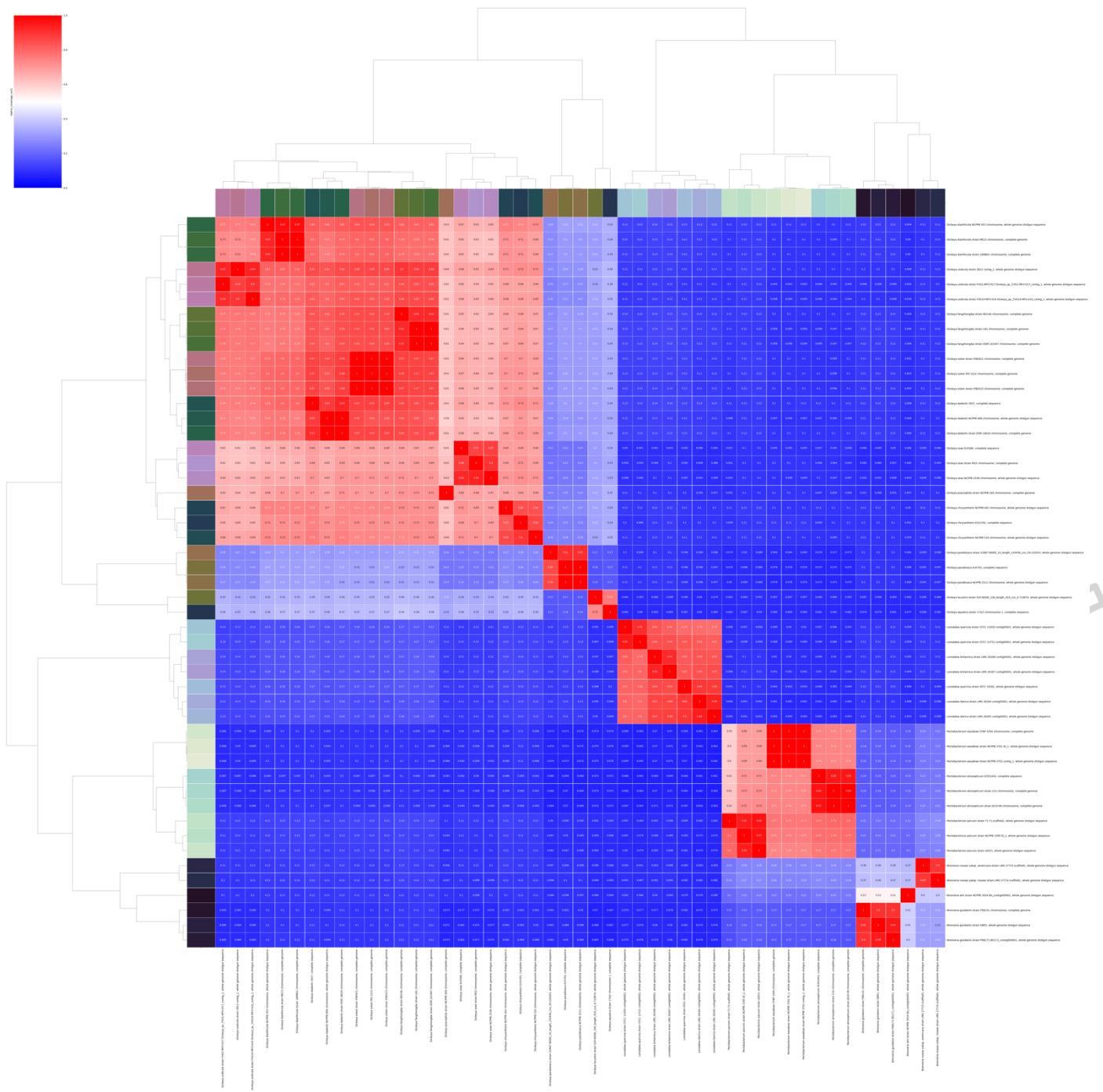


Table S1. The *Musicola* strains used in this study

Strain designations	Origin: country, year, * plant	Isolated by	New classification
A6065 ^T CFBP 4178 ^T NCPBP 2511 ^T LMG 2542 ^T	Columbia, 1970, <i>Musa paradisiaca</i>	Fernandez-Borrero O	<i>Musicola paradisiaca</i> type strain
A1816 CFBP 1445	Columbia, 1972, <i>Musa paradisiaca</i>	Victoria JI	<i>Musicola paradisiaca</i>
A4507 CFBP 3477	Columbia, 1968, <i>Musa paradisiaca</i>	Victoria JI	<i>Musicola paradisiaca</i>
A6358 CFBP 3696	Cuba, 1987, <i>Musa paradisiaca</i>	Rivera N	<i>Musicola paradisiaca</i>
A6375 CFBP 3699 NCPBP 4430	Cuba, 1987, <i>Zea mays</i>	Rivera N	<i>Musicola paradisiaca</i>
A3967 ^T CFBP 722, CFBP 8732 ^T LMG 31880 ^T	France, 1965, <i>Solanum lycopersicon</i>	Prunier JP	<i>Musicola keenii</i> type strain
CFBP1446	Columbia, 1972, <i>Musa paradisiaca</i>	Victoria JI	<i>Musicola paradisiaca</i>
CFBP1451	Columbia, 1972, <i>Musa paradisiaca</i>	Granada G	<i>Musicola paradisiaca</i>

*A : collection of the laboratory Microbiology, Adaptation and Pathogenicity, Lyon, France
 CFBP : Collection Française de Bactéries Phytopathogènes, Beaucauzé, France
 LMG : collection of the Laboratory of Microbiology, Ghent, Belgium
 NCPBP : National Collection of Plant Pathogenic Bacteria, York, UK

Table S2. Metabolic capacities of the type strains of *Musicola keenii* A3967^T (CFBP 8732^T) and *Musicola paradisiaca* CFBP 4178^T

The *D. dadantii* strain 3937 was used for comparison. Metabolic capacities were tested using Biolog plates PM1 and PM2A. Plaques were inoculated with bacteria recovered in the inoculation fluid IF-0 supplemented with dye A, according to the recommendations of the supplier (Biolog, US). Bacterial growth was determined after 48h at 30°C, by measurement of optical density (OD) at $\lambda=590\text{nm}$: -, indicates $\text{OD} < 0.2$; w, indicates $0.2 \leq \text{OD} \leq 0.5$; +, indicates $\text{OD} > 0.5$.

The characters differentiating the two strains are shown in bold letters. The star (*) indicates compounds giving a positive growth when added as the sole carbon source in minimal medium (see Table 1).

Carbon sources	Strains Species	A6065 ^T (CFBP 4178 ^T) <i>M. paradisiaca</i>	A3967 ^T (CFBP 8732 ^T) <i>M. keenii</i>	3937 <i>D. dadantii</i>
L-Arabinose		-	+	+
N-Acetyl-D-glucosamine		+	+	+
D-Saccharic acid (glucaric acid)		-	+	+
Succinic acid		w	+	w
D-Galactose		-	+	+
L-Aspartic acid		w	+	+
L-Proline		-	-	-
D-Alanine		-	-	-
D-Trehalose		-	-	-
D-Mannose		+	+	+
Dulcitol		-	-	-
D-Serine		-	-	-
D-Sorbitol		-	-	-
Glycerol		w	+	+
L-Fucose		-	-	-
D-Glucuronic acid		w	+	-
D-Gluconic acid		+	+	+
D,L- α -Glycerol- phosphate		w	w	w
D-Xylose		-	+	+
L-Lactic acid		-	-	w
Formic acid		-	-	w
D-Mannitol		-	-	+
L-Glutamic acid		-	-	w
D-Glucose-6-phosphate		-	+	+
D-Galactonic acid- γ -lactone		-	-	-
D,L-Malic acid		w	+	+
D-Ribose		-	+	+
Tween 20		-	-	-
L-Rhamnose		-	-	-
D-Fructose		+	+	+
Acetic acid		-	-	w
α -D-Glucose		+	+	+
Maltose		-	-	-
D-Melibiose		+	-	+
Thymidine		-	-	-
L-Asparagine		w	+	+
D-Aspartic acid		-	-	+
D-Glucosaminic acid		-	-	-
1,2-Propanediol		-	-	-
Tween 40		-	-	-
α -Keto-glutaric acid		-	-	-
α -Keto-butyric acid		-	-	-
α -Methyl-D-galactoside		-	-	+
α -D-Lactose		-	-	-
Lactulose		-	-	-
Sucrose		+	+	+
Uridine		-	-	-
L-Glutamine		-	-	w
m-Tartaric Acid		-	-	w
D-Glucose-1-Phosphate		-	-	+

D-Fructose-6-Phosphate	-	+	+
Tween 80	-	-	-
α -Hydroxy glutaric acid- γ -lactone	-	-	-
α -Hydroxy butyric acid	-	-	-
β -Methyl-D-glucoside	w	+	+
Adonitol	-	-	-
Maltotriose	-	-	-
2-Deoxy adenosine	-	-	-
Adenosine	-	-	-
Glycyl-L-aspartic acid	-	-	-
Citric acid	-	-	+
myo-Inositol	-	+	+
D-Threonine	-	-	-
Fumaric acid	+	+	+
Bromo succinic acid	w	w	w
Propionic acid	-	-	-
Mucic acid (galactaric acid)	-	w	+
Glycolic acid	-	-	-
Glyoxylic acid	-	-	-
D-Cellobiose	-	-	+
Inosine	-	-	-
Glycyl-L-glutamic acid	-	-	-
Tricarballic acid	-	-	-
L-Serine	-	w	+
L-Threonine	-	-	-
L-Alanine	-	w	-
L-Alanyl-glycine	-	-	-
Acetoacetic acid	-	-	-
N-Acetyl- β -D-mannosamine	-	-	-
Mono methyl succinate	-	-	-
Methyl pyruvate	+	+	+
D-Malic acid	-	w	w
L-Malic acid	w	+	+
Glycyl-L-proline	-	-	-
p-Hydroxy phenyl acetic acid	-	-	-
m-Hydroxy phenyl acetic acid	-	-	-
Tyramine	-	-	-
D-Psicose	-	+	-
L-Lyxose	-	w	-
Glucuronamide	-	-	-
Pyruvic acid	-	+	+
L-Galactonic acid- γ -lactone	-	-	-
D-Galacturonic acid	+	+	+
Phenylethyl-amine	-	-	-
2-Aminoethanol	-	-	-
Chondroitin sulfate C	-	-	-
α -Cyclodextrin	-	-	-
β -Cyclodextrin	-	-	-
γ -Cyclodextrin	-	-	-
Dextrin	-	-	-
Gelatin	-	-	-
Glycogen	-	-	-
Inulin	-	-	-
Laminarin	-	-	-
Mannan	-	-	-
Pectin	+	+	+
N-Acetyl-D-galactosamine	-	-	-
N-Acetyl-neuraminic acid	-	-	-
β -D-Allose	-	-	-
Amygdalin	-	-	-
D-Arabinose	-	-	w
D-Arabitol	-	-	-
L-Arabitol	-	-	-
Arbutin	+	+	+
2-Deoxy-D-ribose	-	-	-

I-Erythritol	-	-	-
D-Fucose	-	-	-
3-O-β-D-Galacto-pyranosyl-D-arabinose	-	-	-
Gentiobiose	-	-	-
L-Glucose	-	-	-
Lactitol	-	-	-
D-Melezitose	-	-	-
Maltitol	-	-	-
α-Methyl-D-glucoside	-	-	-
β-Methyl-D-galactoside	-	-	w
3-Methyl Glucose	-	-	-
β-Methyl-D-glucuronic acid	-	-	-
α-Methyl-D-mannoside	-	-	-
β-Methyl-D-xyloside	-	-	-
Palatinose	-	-	-
D-Raffinose	w	-	+
Salicin	+	+	+
Sedoheptulosan	-	-	-
L-Sorbose	-	-	-
Stachyose	-	-	-
D-Tagatose	-	-	-
Turanose	-	-	-
Xylitol	-	-	-
N-Acetyl-D-glucosaminitol	-	-	-
γ-Amino butyric acid	-	-	-
δ-Amino valeric acid	-	-	-
Butyric acid	-	-	-
Capric acid	-	-	-
Caproic acid	-	-	-
Citraconic acid	-	-	-
Citramalic acid	-	-	-
D-Glucosamine	+	+	+
2-Hydroxy benzoic acid	-	-	-
4-Hydroxy benzoic acid	-	-	-
β-Hydroxy butyric acid	-	-	-
γ-Hydroxy butyric acid	-	-	-
α-Keto-valeric acid	-	-	-
Itaconic acid	-	-	-
5-Keto-D-gluconic acid	-	-	-
D-Lactic acid methyl ester	-	-	-
Malonic acid	-	-	w
Melibionnic acid	-	-	-
Oxalic acid	-	-	-
Oxalomalic acid	-	-	-
Quinic acid	-	-	-
D-Ribono-1,4-lactone	-	-	-
Sebacic acid	-	-	-
Sorbic acid	-	-	-
Succinamic acid	w	+	+
D-Tartaric acid	-	-	-
L-Tartaric acid	-	-	-
Acetamide	-	-	-
L-Alaninamide	-	-	-
N-Acetyl-L-glutamic acid	-	-	-
L-Arginine	-	-	-
Glycine	-	-	-
L-Histidine	-	-	-
L-Homoserine	-	-	-
Hydroxy-L-proline	-	-	-
L-Isoleucine	-	-	-
L-Leucine	-	-	-
L-Lysine	-	-	-
L-Methionine	-	-	-
L-Ornithine	-	-	-

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* ke...

L-Phenylalanine	-	-	-
L-Pyroglutamic acid	-	-	-
L-Valine	-	-	-
D,L-Carnitine	-	-	-
Sec-Butylamine	-	-	-
D,L-Octopamine	-	-	-
Putrescine	-	-	-
Dihydroxy Acetone	-	-	-
2,3-Butanediol	-	-	-
2,3-Butanedione	-	-	-
3-Hydroxy 2-Butanone	-	-	-

CONFIDENTIAL

Table S3. Genes present in *M. keenii* A3967^T (CFBP 8732^T) but absent in *M. paradisiaca* CFBP 4178^T

Genes of interest are shown in bold letters. Phage-related gene clusters are in italics. Genes involved in *myo*-inositol catabolism are in red letters.

Gene ID Number	Protein Length (aa)	Protein annotation
1	56	transposase for IS1001 element
2	54	4-oxalocrotonate tautomerase
128	150	UPF0306 protein YhbP
134	271	ABC transporter, permease protein 2 (cluster 5, nickel/peptides/opines)
277	366	Predicted cell-wall-anchored protein SasA (LPXTG motif)
280	509	PilV-like protein
286	473	Conjugal transfer protein traA
287	21	Mobile element protein
694	230	Expansin-YoaJ
797	29	Mobile element protein
884	21	Mobile element protein
885	42	Type I restriction-modification system, restriction subunit R (EC 3.1.21.3)
915	91	Integrase
982	234	ATPase
984	258	putative membrane protein
985	697	Coupling protein VirD4, ATPase required for T-DNA transfer
991	343	Putative ATP-binding protein
993	412	Incl1 plasmid conjugative transfer pilus-tip adhesin protein PilV
994	218	Type-IV secretion leader peptidase/N-methyltransferase
995	163	Incl1 plasmid conjugative transfer putative membrane protein PilT
996	193	Incl1 plasmid conjugative transfer prepilin PilS
997	370	Incl1 plasmid conjugative transfer inner membrane protein PilR
999	167	Incl1 plasmid pilus assembly protein PilP
1001	565	Incl1 plasmid conjugative transfer lipoprotein PilN
1003	403	Incl1 plasmid conjugative transfer protein Pill
1006	160	Putative uncharacterized protein STY4534
1016	586	Protein with ParB-like nuclease domain in PFGI-1-like cluster
1021	296	Chromosome partitioning ATPase in PFGI-1-like cluster, ParA-like
1028	37	Mobile element protein
1037	321	DNA methylase N-4/N-6
1042	78	Mobile element protein
1043	247	Transposase
1046	160	Arsenical pump-driving ATPase (EC 3.6.3.16) TEMP
1047	122	Arsenical resistance operon trans-acting repressor ArsD
1134	40	Transposase, IS3/IS911 family
1259	338	CiaB PROTEIN
1706	377	Putative exported protein precursor
1898	49	Mobile element protein
2085	61	Uncharacterized MFS-type transporter
2151	91	Transcriptional regulator, AsnC family
2328	130	SSU ribosomal protein S11p (S14e)
2361	107	Ykfl toxin protein
2362	112	YafW protein (antitoxin to Ykfl)
2363	158	UPF0758 family protein
2364	274	UPF0380 proteins YafZ and homologs
2403	21	Mobile element protein

2421	229	YjbF outer membrane lipoprotein
2423	713	Uncharacterized lipoprotein YjbH
2435	480	2-methylcitrate dehydratase (EC 4.2.1.79)
2437	402	Acyltransferase 3
2439	733	ATP-dependent helicase
2440	439	ATP/GTP-binding protein
2441	770	probable membrane protein YPO2297
2442	468	Mobile element protein
2444	21	Mobile element protein
2451	45	Assimilatory nitrate reductase large subunit (EC 1.7.99.4)
2588	49	COG5499: Predicted transcription regulator containing HTH domain
2781	357	Undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase (EC 2.7.8.33)
2782	379	Putative polysaccharide export protein YccZ precursor
2783	123	Low molecular weight protein-tyrosine-phosphatase (EC 3.1.3.48) => Etp
2784	723	Tyrosine-protein kinase (EC 2.7.10.2)
2800	402	Radical SAM domain protein
2822	257	Branched-chain amino acid ABC transporter, amino acid-binding protein (TC 3.A.1.4.1)
2825	21	Mobile element protein
2838	53	Homoserine kinase (EC 2.7.1.39)
3124	56	transposase for IS1001 element
3175	129	Transposase IS3/IS911
3388	32	Transcriptional regulator, GntR family domain
3389	40	DNA-binding transcriptional regulator, MocR family
3404	216	Transcriptional regulator KPN_02146, AcrR family
3417	72	transposase for IS1001 element
3418	179	transposase for IS1001 element
3420	141	Oligopeptide transport ATP-binding protein OppD (TC 3.A.1.5.1)
3421	54	transposase for IS1001 element
3422	109	Flagellar hook-length control protein FliK
3434	276	5-deoxy-glucuronate isomerase (EC 5.3.1.-) lolB, myo-inositol catabolism
3435	297	Inosose dehydratase (EC 4.2.1.44) lolE, myo-inositol catabolism
3445	154	DNA gyrase inhibitory protein
3447	504	Malonate-semialdehyde dehydrogenase [inositol] (EC 1.2.1.18) lolA, myo-inositol catabolism
3509	122	Methyl-accepting chemotaxis protein I (serine chemoreceptor protein)
3529	429	L-rhamnonate transporter (predicted by genome context)
3626	123	Putative regulatory protein
3629	233	(adenine-N6-)-methyltransferase homolog
3637	168	<i>Regulatory protein CII bacteriophage 186</i>
3638	100	<i>Phage regulatory protein</i>
3686	65	Integrase
3732	86	hypothetical protein formerly called flagellar hook-length control protein FliK
3733	85	hypothetical protein formerly called flagellar hook-length control protein FliK
3735	71	hypothetical protein formerly called flagellar hook-length control protein FliK
3977	280	ATPase
3981	71	probable membrane protein STY4566
3986	138	putative lipoprotein
3987	940	Type IV secretory pathway, VirB4 components
3988	129	conserved hypothetical protein
3990	67	corresponds to STY4575 from Accession AL513382: Salmonella typhi CT18
3991	48	corresponds to STY4575 from Accession AL513382: Salmonella typhi CT18
3993	47	Type IV secretory pathway, VirB4 components
4002	119	transcriptional regulator, MerR family
4003	516	putative membrane protein

4022	260	VrIR-like protein
4023	1170	DEAD/DEAH box helicase domain protein
4025	581	Bipolar DNA helicase HerA
4235	189	D-alanyl-D-alanine dipeptidase (EC 3.4.13.22)
4275	39	Maltodextrin ABC transporter, substrate-binding protein MdxE
4297	304	Siderophore achromobactin ABC transporter, substrate-binding protein
4316	88	beta-glucosidase (EC 3.2.1.21); 6-phospho-beta-glucosidase (EC 3.2.1.86)
4317	156	beta-glucosidase (EC 3.2.1.21); 6-phospho-beta-glucosidase (EC 3.2.1.86)

CONFIDENTIAL

Table S4. Genes present in *M. paradisiaca* CFBP 4178^T but absent in strain *M. keenii* A3967^T (CFBP 8732^T);

Genes of interest are shown in bold letters. Phage-related gene clusters are in italics. Genes involved in melibiose and raffinose catabolism are in red letters.

Gene ID Number	Protein Length	Protein annotation
16	926	DNA helicase
26	466	DNA-cytosine methyltransferase (EC 2.1.1.37)
39	342	Uncharacterized protein RhuM
489	41	Mobile element protein
577	187	Acetyltransferase, GNAT family
716	50	Cell division protein FtsQ
760	51	Alcohol dehydrogenase (EC 1.1.1.1)
796	627	Retron-type RNA-directed DNA polymerase (EC 2.7.7.49)
807	185	CRISPR-associated protein, Csy4 family
808	337	CRISPR-associated protein, Csy3 family
809	317	CRISPR-associated protein, Csy2 family
810	442	CRISPR-associated protein, Csy1 family
811	1096	CRISPR-associated helicase Cas3
812	334	CRISPR-associated protein Cas1
813	166	Sensory box histidine kinase
982	38	Mobile element protein
1069	401	Macrolide-efflux protein
1070	224	Thymidylate kinase (EC 2.7.4.9)
1071	381	Biotin synthase-related enzyme
1073	188	Cell division trigger factor (EC 5.2.1.8)
1401	227	Oxidoreductase
1404	178	Homoserine kinase (EC 2.7.1.39)
1450	363	ABC transporter, substrate-binding protein (cluster 10, nitrate/sulfonate/bicarbonate)
1488	427	Raffinose and melibiose permease, Raft
1489	710	alpha-galactosidase (EC 3.2.1.22), RafA, raffinose and melibiose catabolism
1504	457	Maltodextrin ABC transporter, substrate-binding protein MdxE
1508	288	Metal-dependent hydrolase
1566	448	N-acetylglucosamine-regulated outer membrane porin
1592	104	Type I restriction-modification system, restriction subunit R (EC 3.1.21.3)
1596	421	Polyketide biosynthesis 3-hydroxy-3-methylglutaryl-ACP synthase PksG
1598	84	Acyl carrier protein
1601	435	Monooxygenase, flavin-binding family
1604	6877	Modular polyketide synthase
1609	5613	Modular polyketide synthase
1611	275	Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39); Enoyl-[acyl-carrier-protein] reductase [FMN] (EC 1.3.1.9)
1612	216	Modular polyketide synthase
1614	161	Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39)
1803	188	Transcriptional regulator, AcrR family
1900	268	Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39)
1990	232	Putative preQ0 transporter
2041	352	Oxidoreductase
2042	281	putative prolyl aminopeptidase
2045	401	cobalt dependent X-Pro dipeptidase
2062	217	Transcriptional regulator, AcrR family
2086	64	Sodium-dependent phosphate transporter
2250	138	FIG00904844: hypothetical protein
2257	104	Death on curing protein, Doc toxin

2348	63	<i>Tail-specific protease precursor</i> (EC 3.4.21.102)
2359	170	<i>Replication gene B protein</i>
2365	49	<i>Putative phage tail protein</i>
2366	94	<i>Putative phage tail protein</i>
2455	313	putative membrane protein
2457	188	Adenine phosphoribosyltransferase (EC 2.4.2.7)
2550	38	Hypothetical MFS-type transporter protein YcaD
2668	94	Integrase
2685	226	<i>prophage protein</i>
2686	67	<i>prophage protein</i>
2687	879	<i>DNA primase</i> (EC 2.7.7.-)
2798	54	ISSod13, transposase
2836	413	Formyl-coenzyme A transferase (EC 2.8.3.16)
2838	79	Formyl-coenzyme A transferase (EC 2.8.3.16)
2839	347	Formyl-coenzyme A transferase (EC 2.8.3.16)
2939	59	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase (EC 2.1.1.61)
2966	210	Transcriptional regulator, TetR family
2972	125	Nitrogenase (iron-iron) delta chain (EC 1.18.6.1)
2974	140	DNA recombination protein RmuC
2975	209	AnfO protein, required for Mo- and V-independent nitrogenase
2979	301	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase (EC 2.3.1.117)
3091	207	Collagenase and related proteases
3103	47	Inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205) / CBS domain
3348	488	<i>Phage integrase</i>
3436	248	TriL protein
3437	73	CcdA protein (antitoxin to CcdB)
3438	102	CcdB toxin protein
3440	670	<i>FIG006126: DNA helicase, restriction/modification system component YeeB</i>
3441	872	<i>FIG045374: Type II restriction enzyme, methylase subunit YeeA</i>
3442	190	<i>Phage DNA invertase</i>
3451	129	Putative integrase protein
3452	266	carbamoyltransferase
3453	434	Nodulation protein nolO (EC 2.1.3.-)
3461	535	Polyketide synthase?
3464	114	Glyoxalase family protein
3465	4484	Polyketide synthase modules and related proteins
3472	389	Split AAA-ATPase protein PA0787
3484	351	Acyl-coenzyme A:6-aminopenicillanic-acid-acyltransferase 40 kDa form (EC 2.3.1.164)
3499	261	Thioesterase
3502	231	putative membrane protein
3505	73	CcdA protein (antitoxin to CcdB)
3506	102	CcdB toxin protein
3507	424	Putative cryptic D-serine deaminase (EC 4.3.1.18)
3512	131	RidA/YER057c/UK114 superfamily protein
3516	143	Arsenate reductase (EC 1.20.4.1) glutaredoxin-coupled, glutaredoxin-like family
3518	239	putative membrane protein
3520	351	Arsenical-resistance protein ACR3
3521	56	Arsenical pump-driving ATPase (EC 3.6.3.16)
3557	241	Predicted permeases
3698	381	group 1 glycosyl transferase
3755	180	RNA polymerase ECF-type sigma factor
3756	328	Fe ²⁺ -dicitrate sensor, membrane component
3852	248	TriL protein

3853	73	CcdA protein (antitoxin to CcdB)
3854	102	CcdB toxin protein
3855	333	Retron-type RNA-directed DNA polymerase (EC 2.7.7.49)
3941	83	Potassium uptake protein TrkH
3946	364	ABC transporter, ATP-binding protein
3985	69	UbiD family decarboxylase, <i>Lactobacillus brevis</i> type
3993	52	Replicative helicase RepA
4000	219	Transposase
4003	682	Replicative helicase RepA
4163	233	Homolog of eukaryotic DNA ligase III
4179	79	Mobile element protein
4181	38	Mobile element protein
4205	70	DNA-binding transcriptional regulator, MocR family / aminotransferase domain
4264	174	Transposase
4265	75	Mobile element protein
4317	182	endolysin
4325	371	Zn peptidase with DNA binding
4335	92	Cox
4337	52	C protein
4338	328	Integrase
4521	224	putative lipoprotein
4522	118	putative lipoprotein
4523	220	putative lipoprotein
4534	363	4-hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27)
4543	223	Toxin HigB / Protein kinase domain of HipA
4544	84	HipB protein, Antitoxin HigA

CONFIDENTIAL



[Click here to access/download](#)

Additional Material for Reviewer

For reviewer_Fig S3ab_Fig S4ab.pptx



CONFIDENTIAL