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.

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

Author affiliations:

1 Univ Lyon, CNRS, INSA Lyon, UCBL, UMR 5240 Microbiologie Adaptation et Pathogénie, F-69622 Villeurbanne, France
2 Université de Lyon, Université Claude Bernard Lyon 1, CNRS FR 3728 BioEnviS, plateau DTAMB, F-69621 Villeurbanne, France
3 Strathclyde Institute of Pharmacy & Biomedical Sciences, 161 Cathedral Street, Glasgow G4 ORE, UK

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

Email address of authors: nicole.cotte-pattat@insa-lyon.fr, cecile.jacot-des-combes@univ-lyon1.fr, jerome.briolay@univ-lyon1.fr, leighton.pritchard@strath.ac.uk

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

Subject category: Taxonomic description, new taxa - Proteobacteria

Word count: 4250

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

Abbreviations: ANI, average nucleotide identity; dDDH, digital DNA-DNA hybridization; CFBP, Collection Française de Bactéries Phytopathogènes; CE, carbohydrate esterase; GH, glycoside hydrolase; PL,
polysaccharide lyase; RAST, rapid annotations using subsystems technology; T1SS, T2SS, T3SS, T4SS and T6SS, type I, II, III, IV and VI secretion systems.

**ABSTRACT** (238 words)

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

**INTRODUCTION**

For nearly 40 years, our laboratory has been interested in bacteria formerly belonging to *Erwinia* and *Pectobacterium* but now included in the genus *Dickeya*, an important group of plant pathogens that affect a wide range of hosts, including vegetable crops and ornamental plants [1, 2, 3]. Most characterized *Dickeya* strains originate from infected crops or ornamental plants, although a few *Dickeya* strains have been isolated from water. This genus belongs to the *Enterobacteriales* order and more precisely to the *Pectobacteriaceae* family that includes five genera: *Brenneria, Dickeya, Lonsdalea, Pectobacterium*, and *Sodalis* [4]. *Dickeya* and *Pectobacterium* species cause soft-rot diseases on plants due to the action of extracellular pectinases that attack the plant cell wall [1, 2]. Members of these two genera are often designated as soft-rot *Pectobacteriaceae* (SRP).
The history of SRP classification began with the establishment of the genus *Erwinia*, which was founded to gather several Gram negative plant pathogenic bacteria [5]. This genus included, among others, the species *Erwinia chrysanthemi* and *Erwinia carotovora*. A non-official classification in six pathovars related to the host plant was used to describe *E. chrysanthemi* members, namely pv. chrysanthemi, pv. dianthicola, pv. dieffenbachiae, pv. parthenii, pv. zeae and pv. parasiadiaca [6, 7]. Strains of the pathovar paradisiaca were also proposed for elevation to the species *Erwinia paradisiaca* [8]. As a consequence of reclassification in the genus *Erwinia*, SRP were gathered into the genus *Pectobacterium* that included two species *Pectobacterium carotovorum* and *Pectobacterium chrysanthemi*, replacing *E. carotovora* and *E. chrysanthemi*, respectively [9]. However, the species *E. paradisiaca* was included in the genus *Brenneria* and named *Brenneria paradisiaca* [10]. The nomenclature again changed in 2005 with the proposal that strains formerly designated *P. chrysanthemi* or *B. paradisiaca* be placed into the new genus *Dickeya* [11]. At this time, the genus *Dickeya* comprised six recognized species: *D. chrysanthemi*, *D. dadantii*, *D. dieffenbachiae*, *D. dianthicola*, *D. zeae* and *D. paradisiaca* [11]. Thereafter, new changes were proposed in the genus *Dickeya*. Members of the species *D. dieffenbachiae* were reclassified as a subspecies of *D. dadantii* (i.e. *D. dadantii* subsp. *dieffenbachiae*) [12]. More recently, novel species have been characterized: *D. solani* for isolates responsible for potato diseases in Europe [13, 14]; *D. fangzhongdai* isolated from pear trees in China and orchids in different countries [15, 16]; *D. poaceiphila* for strains responsible for a sugarcane disease in Australia [17]; and *D. oryzae* causing rice diseases [18]. Three new species were also identified in water samples: *D. aquatica* from rivers in England and Finland [19]; *D. lacustris* from lakes in France [20]; and *D. undicola* from water samples collected in Malaysia and France [21]. Thus, the genus *Dickeya* currently comprises twelve species with validly accepted names: *D. aquatica*, *D. chrysanthemi*, *D. dadantii*, *D. dianthicola*, *D. fangzhongdai*, *D. lacustris*, *D. oryzae*, *D. paradisiaca*, *D. poaceiphila*, *D. solani*, *D. undicola*, and *D. zeae*.

Previous studies have shown *D. paradisiaca* to be the most basal member of the genus *Dickeya* in terms of nucleotide identity, pan-genome content, genome synteny and whole-genome phylogeny [22, 23, 24]. Genomic data suggested that differences between the *D. paradisiaca* strains and other members of the genus *Dickeya* would justify separation into a new genus [22, 24]. However, data on the genetic diversity of *D. paradisiaca* are scarce, and it is one of the least well characterized *Dickeya* species at the genomic level. Most *D. paradisiaca* strains were isolated from *Musa paradisiaca* (banana trees) in tropical or subtropical countries (Colombia, Cuba, Jamaica, Panama, etc.) [25]. Prior to this publication, only two *D. paradisiaca* genomes were available, those of the type strain NCPPB 2511 [26] and of strain Ech703, isolated from *Musa paradisiaca* in Colombia in 1970 and from *Solanum tuberosum* in Australia, respectively [27, 28]. To better understand diversity in the genus *Dickeya*, we analysed poorly characterized SRP strains stored in our laboratory collection. Phenotypic and genetic analyses suggested that, in addition to the type strain, five...
isolates belong to the species *D. paradisiaca*. Four isolates appeared to be similar to the type strain, but strain A3967 showed genetic divergence and atypical phenotype that appeared sufficient to justify the proposal of a novel species. Phenotypic, phylogenetic and genomic arguments also justify a reassignment of *D. paradisiaca* to the genus level, for which we propose the name *Musicola* gen. nov. This novel genus includes two species: *Musicola paradisiaca* comb. nov. as the type strain, and *Musicola keenii* sp. nov., with NCPPB 2511\textsuperscript{T} (CFBP 4178\textsuperscript{T}, LMG 2542\textsuperscript{T}) and A3967\textsuperscript{T} (CFBP 8732\textsuperscript{T}, LMG 31880\textsuperscript{T}) as the type strain, respectively.

**GENETIC AND PHENOTYPIC CHARACTERIZATION OF STRAINS**

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

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

The 16S rRNA genes of each of the five strains were amplified by PCR and sequenced. A BLASTN search against nt NCBI database confirmed that the top hits for the 16S rRNA amplicons are annotated as strains of the species *D. paradisiaca*. A 16S phylogenetic tree was inferred as described above in MEGA X (version 10.2.4.), there were a total of 1548 positions in the final dataset. In the resulting tree, the branch corresponding to the *D. paradisiaca* strains is clearly included inside the genus *Dickeya* (Fig. S1). However, the 16S rRNA gene sequences are not considered as a discriminant marker for *Enterobacterales* classification [4]. Some divergence in the 16S rRNA sequences was observed between strain A3967 and the other *D. paradisiaca* strains, including the two strains whose genome has been sequenced, CFBP 4178\textsuperscript{T} and Ech703 (Fig. S1).
A phenotypic characterization of the six available strains (Table 1) was performed using minimal media (M63) supplemented with one carbon compound (2 g l\(^{-1}\)) to determine if they could use each compound as a sole carbon and energy source for growth. Growth was recorded after incubating plates at 30°C for 24 to 72 h (Table 1). Similarly to CFBP 4178\(^T\), the five strains were able to grow in the presence of several monosaccharides or oligosaccharides, such as D-arabinose, L-arabinose, D-fructose, D-galactose, D-galacturonate, D-glucose, D-glucuronate, D-mannose, D-ribose, sucrose, and D-xylose (Table 1). In contrast, they were not able to assimilate D-mannitol, L-rhamnose and D-cellobiose. Strain A3967 differs from the five other strains by its capacity to catabolize myo-inositol and its inability to utilize melibiose and raffinose (Table 1). Thus, strain A3967 showed an atypical phenotype for sugar assimilation in comparison to the five other strains assigned to the species *D. paradisiaca*, which showed homogeneous profiles.

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

As A3967\(^T\) (CFBP 8732\(^T\)) appears to be an atypical strain, we decided to compare this strain to the *D. paradisiaca* type strain CFBP 4178\(^T\) using a large biochemical characterization performed with Biolog plates PM1 and PM2A which contain 190 potential carbon sources [32] (Table S2). Inoculations were performed according to the manufacturer instructions and lecture was made after 48h at 30°C. The data obtained for the *D. paradisiaca* type strain agreed with a previous report also using the Biolog plates PM1 and PM2A [15]. However, negative results were obtained for several compounds that allowed the growth of CFBP 4178\(^T\) in our analysis of sugar assimilation in minimal media (Table 1). Rather than bacterial growth, the Biolog system detects the metabolic activity of the cells due to substrate assimilation; this activity is visualized by the color change of an oxidoreduction indicator. The Biolog system may have been inappropriate for strain CFBP 4178\(^T\), perhaps due to its low metabolic activity. Such a discrepancy between Biolog data and growth on minimal medium supplemented with each compound was not observed for strain A3967\(^T\) (CFBP 8732\(^T\)) or *D. dadantii* 3937 (Table S2, Table 1). In the Biolog system, dissimilar results between the two strains A3967\(^T\) (CFBP 8732\(^T\)) and CFBP 4178\(^T\) were observed for assimilation of 17 carbon sources (Table S2). Three differences were confirmed by testing the bacterial growth in minimal medium, namely myo-inositol, D-melibiose, and D-raffinose. Ten negative results of CFBP 4178\(^T\) should be taken with caution as the substrates have not been
tested in minimal medium, namely D-fructose-6-phosphate, galactaric acid, D-glucaric acid, D-glucose-6-phosphate, L-lyxose, D-psicose, L-serine, L-alanine, D-malic acid, and pyruvic acid.

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

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

**GENOMIC COMPARISONS**

To gain further information on strain A3967$^T$ (CFBP 8732$^T$), its genome sequence was determined. The total bacterial genomic DNA was extracted using a NucleoSpin® bacterial DNA purification kit (Macherey-Nagel). Quantification and quality control of the DNA was performed using a Nanodrop spectrophotometer, a Qubit4 fluorimeter and agarose gel electrophoresis. Genomic DNA was sequenced using a MiSeq Illumina platform (Biofidal, Vaux en Velin, France). We assembled the genome using SPAdes version 3.11.1 (2020/10) [33]. The resulting draft genome of strain A3967$^T$ (CFBP 8732$^T$) comprises 72 contigs (N50=266,679, L50=7, 125X coverage depth) with a total length of 4,402,645 bp and a G+C content (mol%) of 54.4%. The draft genome was automatically annotated using RAST version 2.0 (2020/10) [34], which predicted 4,354 protein coding genes.
and 76 RNA-coding sequences, including 68 tRNAs and 8 rRNAs (23S, 16S, and 6 x 5S). The draft genome has NCBI accession GCF_014855505.1.

To clarify the taxonomic position of strain A3967ᵀ (CFBP 8732ᵀ), we calculated digital DNA-DNA hybridization (dDDH) and average nucleotide identity (ANI) values. The dDDH method was proposed as a means of approximating the wet-lab DDH method [35]. Using the A3967ᵀ (CFBP 8732ᵀ) genome as a reference and Dickeya genomes as queries (Table 3), dDDH gave values of 68.3 and 68.4% with D. paradisiaca CFBP 4178ᵀ and Ech703, respectively, and values lower than 24% with other Dickeya species. The dDDH value of 68.3% obtained by comparing strain A3967ᵀ (CFBP 8732ᵀ) with the D. paradisiaca type strain is below the threshold dDDH of 70% commonly used to delineate species [35]. However in practice there is significant variance in the results of in vitro DDH; genome sequences sharing ≈70% overall sequence identity may give DDH estimates between 60% and 90% [35, 36]. In addition, there is uncertainty in the mapping of dDDH values to DDH values obtained in vitro. Straightforward transfer of a 70% dDDH threshold to species boundary is therefore not reliable.

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

Although there is no formal definition of genus delineation on the basis of genome similarity, it has been argued that assigning two organisms to the same genus is not sound when only a small proportion of their genome sequences is recognisably homologous. Empirical evidence from recent approaches using aligned fraction to discriminate at genus level indicates that genus boundary values vary by taxon, but members of the same genus tend to share at least 60% coverage [38, 39]. Consistent with these surveys, the D. paradisiaca genomes would constitute a distinct genus from the D. lacustris/D. aquatica group, and from the remaining group of Dickeya genomes including D. chrysanthemi, D. dadantii, D. dianthicola, D. fangzhongdai, D. oryzae, D. poaceiphila, D. solani, D. undicola and D. zeae. In particular, we observe that the alignment coverage between
D. paradisiaca and any other Dickeya genome is no greater than that between isolates of Escherichia and Salmonella. Hence we propose reclassification of A3967 and strains previously assigned to the species D. paradisiaca into a novel genus. There is substantial evidence for a discontinuity in ANI sequence identity consistent with pre-existing bacterial species distinctions [40]. The threshold, associated with barriers to homologous recombination, corresponds to identities around 95±1 % ANI, the exact value differing by genus [41, 42]. There is consequently no precise universal ANI percentage threshold applicable to all bacterial genera or species, and additional evidence must be taken into account. In our analysis, isolate A3967T (CFBP 8732T) has about 96% ANI identity with the genomes of the two strains assigned to the species D. paradisiaca, consistent with a species-level delineation that is supported by the phenotypic differences reported above and the genomic differences described below.

PHYLOGENOMIC ANALYSIS SUPPORTING ESTABLISHMENT OF MUSICOLA GEN. NOV.

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

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

The partition file and concatenated alignment were used as input to raxml-ng v1.0.2 [47] to generate a best-fit phylogenetic tree using maximum likelihood. The GTR+F0+G4m+B model was used, parameterized separately for each of the 1201 genes. A single tree topology was fit for each of 20 starting trees, suggesting that this was the globally-optimal topology. One hundred bootstrap replicate trees were determined to estimate support values for each tree partition; MRE-based bootstrapping indicated that convergence could be reached with
only 50 replicates. The fitted topology (Fig. 3) supports the genus and species divisions implied by whole-genome classification (Fig. S4). Thus, both ANIm and a comprehensive multigene phylogeny support the same genus and species divisions, including establishment of Musicola as a novel genus.

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

GENOME CONTENT OF MUSICOLA SPECIES

The genome content of M. keenii A3967\textsuperscript{T} (CFBP 8732\textsuperscript{T}) was compared with that of the M. paradisiaca strains CFBP 4178\textsuperscript{T} and Ech703 whose genome has been previously sequenced (GCA_000400505.1 and GCA_000023545.1, respectively). The genomes of CFBP 4178\textsuperscript{T} or Ech703 are closely related, with only a few tens of genes predicted to be present in only one of these two strains (data not shown). In contrast, using the function-based comparison tool in RAST, about 600 genes were predicted to be present in A3967\textsuperscript{T} (CFBP 8732\textsuperscript{T}) and absent in CFBP 4178\textsuperscript{T} or Ech703 or, vice versa, present in CFBP 4178\textsuperscript{T} or Ech703 and absent in A3967\textsuperscript{T} (CFBP 8732\textsuperscript{T}). Among functionally-annotated genes, 110 were specific to M. keenii A3967\textsuperscript{T} (CFBP 8732\textsuperscript{T}) (Table S3), including a cluster encoding a pilus of the IncI1 type, the iol cluster involved in myo-inositol assimilation, a cluster encoding two VirB4 components of type 4 secretion system (T4SS) and two predicted \(\beta\)-glucosidases. Conversely, 125 functionally-annotated genes were found in M. paradisiaca strains CFBP 4178\textsuperscript{T} and Ech703 but absent in A3967\textsuperscript{T} (CFBP 8732\textsuperscript{T}), including the genes rafAT involved in both melibiose and raffinose assimilation, a cluster encoding the components of a CRISPR system, a cluster encoding four polyketide synthases (PKS), three toxin/antitoxin couples and three loci containing predicted prophage genes (Table S4). These genomic differences confirm a genomic basis for the phenotypic differences observed between the two strains A3967\textsuperscript{T} (CFBP 8732\textsuperscript{T}) and CFBP 4178\textsuperscript{T} for myo-inositol, melibiose and raffinose assimilation (Table 1). By comparison
with *Dickeya* species, the lack of the cluster *mtlAD* in the genomes of A3967\(^T\) (CFBP 8732\(^T\)), CFBP 4178\(^T\) and Ech703 explains the inability of *Musicola* strains to assimilate D-mannitol (Table 1).

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

Several differences presented by CFBP 4178\(^T\) and Ech703 in comparison with other *Dickeya* species were previously observed by Pedron and Van Gijsegem who use genome data to investigate the diversity in the genus *Dickeya* [24]. By examination of the *Dickeya* core genome, they noticed that only 1 800 genes, representing about 40% of the gene content, are common between *D. paradisiaca* strains, i.e. CFBP 4178\(^T\) and Ech703, and other *Dickeya* species [24], a value consistent with our results on genome coverage. Moreover, a pangenome analysis based on the presence/absence of the genes, showed that the two strains CFBP 4178\(^T\) and Ech703 cluster outside the *Dickeya* genus [24]. The authors also noticed that CFBP 4178\(^T\) and Ech703 lack several genes known to be involved in *Dickeya* virulence [24]. These dissimilarities, also observed in this study for strain A3967\(^T\) (CFBP 8732\(^T\)), suggest notable differences in the virulence strategies of *Dickeya* and *Musicola* members. Indeed, virulence tests have shown that the *Musicola* strains have a weak maceration activity on...
potato tubers and chicory leaves, two plant models classically used to evaluate the virulence of *Dickeya* strains (Table 2).

**DESCRIPTION OF THE NEW GENUS AND SPECIES**

**Description of *Musicola* gen. nov.**

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

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

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

*Musicola* members are gram-negative, non-sporeforming, facultatively anaerobic pectinolytic bacteria. Cells have average dimensions of 0.6 by 1.7 pm. They are motile with peritrichous flagella. After 48h at 30°C on LB 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 colonies of 0.5-2.5 mm in diameter with translucent appearance. The optimum temperature for bacterial 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-arabinose, N-acetyl-D-glucosamine, D-fructose, D-galactose, D-galacturonic acid, D-gluconic acid, D-glucose, D-glucuronic acid, glycerol, D-mannose, D-ribose, sucrose, D-xylose, arbutin, salicin, citric acid, lactic acid, succinic acid, polygalacturonate or pectin as the sole carbon source but they unable to utilize D-arabitol, D-cellobiose, gentiobiose, D-lactose, D-maltose, D-mannitol, L-rhamnose, D-sorbitol, or D-trehalose. They produce extracellular pectinases, a cellulase but no proteases. They do not produce the blue pigment indigoidine.

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

**Description of *Musicola paradisiaca* comb. nov.**

*Musicola paradisiaca* [pa.ra.di.si.a’ca L. fem. adj. *paradisiaca*, referring to the isolation of most strains from *Musa paradisiaca*]. General description as for the genus. The optimum temperature for bacterial growth is 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-arabinose, L-arabinose, N-acetyl-D-glucosamine, D-fructose, D-galactose, D-galacturonic acid, D-gluconic acid, D-glucose, D-glucuronic acid, glycerol, D-mannose, D-melibiose, D-raffinose, D-ribose, sucrose, D-xylose, arbutin, salicin, citric acid, lactic acid, succinic acid, polygalacturonate or pectin as the sole carbon source but it is unable to utilize *myo*-inositol.
The DNA G+C content of the type strain NCPPB 2511\textsuperscript{T} (CFBP 4178\textsuperscript{T}; LMG 2542\textsuperscript{T}) is 55.0% based on the genome sequence. Other characterized members of this species include strain Ech703, identified on the basis of its genome analysis, and strains CFBP 1445, CFBP 1446, CFBP 1451, CFBP 3477, CFBP 3696, and CFBP 3699 identified on the basis of phenotypic data. Among these \textit{M. paradisiaca} members, six strains were isolated in Colombia from \textit{Musa paradisiaca}, from 1968 to 1972, and two strains were isolated in Cuba in 1987, from \textit{Musa} sp. and \textit{Zea mays}, respectively.

**Description of \textit{Musicola keenii} sp. nov.**

\textit{Musicola keenii} [keen\textsuperscript{1}.i N.L. gen. n. keenii in honour of the American molecular biologist Noel T. Keen] [50].

General description as for the genus. The optimum temperature for bacterial growth is about 33\textdegree C and the type strain can grow up to 41\textdegree C. \textit{M. keenii} A3967\textsuperscript{T} is able to grow in the presence of D-arabinose, L-arabinose, N-acetyl-D-glucosamine, D-fructose, D-fructose-6-phosphate, D-galactose, D-galacturonic acid, glucaric acid, D-gluconic acid, D-glucose, D-glucose-6-phosphate, D-glucuronic acid, glycerol, \textit{myo-inositol}, D-mannose, D-ribose, D-psicose, sucrose, D-xylose, arbutin, salicin, citric acid, lactic acid, pyruvic acid, succinic acid, polygalacturonate or pectin as the sole carbon source but it is unable to utilize D-melibiose or D-raffinose.

The DNA G+C content of the type strain A3967\textsuperscript{T} (CFBP 8732\textsuperscript{T}; LMG 31880\textsuperscript{T}) is 54.4% based on the genome sequence (sequence accession JAAWVW000000000).

---

**AUTHORS’ STATEMENTS**

**Funding information**

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

**Acknowledgements**

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

**Supporting Information**

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

The authors declare that there are no conflicts of interest.

ABBREVIATIONS

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

REFERENCES


47. 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; btz305
doi:10.1093/bioinformatics/btz305

48. Hugouvieux-Cotte-Pattat N, Condemine G, Shevchik VE. Bacterial pectate lyases, structural and

49. Reverchon S, Nasser W. *Dickeya* ecology, environment sensing and regulation of virulence programme.

2007; 45:25-42.
**FIGURES AND TABLES**

**Fig. 1.** Phylogenetic position of A3967 and different strains based on gapA gene sequences.
This analysis was performed using available gapA gene sequences from type strains of Dickeya species, and sequences of the gapA PCR product (represented by a circle) for the laboratory strains A1816, A3967, A4507, A6358, A6375 and A6065T. Type strains of Pectobacterium carotovorum, P. atrosepticum and P. wasabiae were used as outgroups. The evolutionary history was inferred using the neighbour-joining method [30], with bootstrap support values indicated (1000 bootstrap replicates). The evolutionary distances were computed using the maximum composite likelihood method and are in the units of the number of base substitutions per site (718 positions). Evolutionary analyses were conducted in MEGA X version 10.2.4.

**Fig. 2.** ANIm percentage identity and coverage for Dickeya type strains.
Heatmaps of (a) ANIm identity and (b) ANIm coverage for strains A3967, CFBP 4178T, Ech703, the type strains of eleven Dickeya species, and for Pectobacterium carotovorum CFBP 2046T.

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

In (b) 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 (approximating a genus boundary). The coherent red blocks imply that P. carotovorum CFBP 2046T belongs to a discrete genus, distinct from the Dickeya genomes. Likewise, the three D. paradisiaca genomes belong to a discrete genus distinct from the other Dickeya genomes. The D. lacustris and D. aquatica type strains might also belong to a distinct genus, separate from the other Dickeya. The remaining eight Dickeya species appear to constitute a single coherent genus group sharing at least 61% of their complete genomes in homologous pairwise alignment. A complete heatmap showing ANIm coverage for all publicly available Dickeya genomes is given in Fig. S3b.
Fig. 3. Multigene phylogenetic reconstruction on a set of 49 genomes spanning five genera in the Pectobacteriaceae

Maximum-likelihood phylogenetic reconstruction obtained using raxml-ng for 49 Pectobacteriaceae genomes, constructed from 1201 single-copy orthologues shared by all genomes. The GTR+F0+G4m+B model was used and parametrized separately for each orthologue. The best-fit topology shown was obtained for each of 20 distinct starting trees and is likely to be the global optimum; support for internal bipartitions was obtained using 100 bootstraps. The topology shown was midpoint-rooted, manually annotated and coloured using figtree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). The tree supports reassignment of *D. paradisiaca* to genus level, with proposed name *Musicola.*
Table 1. Carbon assimilation

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

<table>
<thead>
<tr>
<th></th>
<th>A3967(^T)</th>
<th>A6065(^T)</th>
<th>A1816</th>
<th>A4507</th>
<th>A6358</th>
<th>A6375</th>
<th>D. dadantii</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFBP 8732 (^T)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6358</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6375</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3937</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- D-arabinose  w*  w*  w  w  w  w  w
- L-arabinose  +  +*  +  +  +  +  +
- D-galactose  +  +*  +  +  +  +  +
- D-glucose  +  +  +  +  +  +  +
- D-fructose  +  +  +  +  +  +  +
- D-mannose  +  +  +  +  +  +  +
- L-rhamnose  -  -  -  -  -  -  -
- D-ribose  +  +*  +  +  +  +  +
- D-xylose  +  +*  +  +  +  +  +
- D-galacturonate  +  +  +  +  +  +  +
- D-glucuronate  +  +  +  +  +  +  -
- D-cellobiose  -  -  -  -  -  -  -
- D-melibiose  -  +  +  +  +  +  +
- D-raffinose  -  +  +  +  +  +  +
- Sucrose  +  +  +  +  +  +  +
- Glycerol  +  +  +  +  +  +  +
- D-mannitol  -  -  -  -  -  -  -
- myo-Inositol  +  -  -  -  -  -  +
- Citrate  w  w*  w  w  w  w  w
- L-lactate  w  w*  w  w  w  w  w
- Polygalacturonate  +  +  +  +  +  +  +

*For these compounds, a discrepancy was observed with data from the Biolog plates PM1 or PM2A that gave a negative result (see Table S2).
Table 2. Enzyme secretion, motility and maceration ability

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

<table>
<thead>
<tr>
<th></th>
<th>A3967&lt;sup&gt;T&lt;/sup&gt;</th>
<th>A6065&lt;sup&gt;T&lt;/sup&gt;</th>
<th>A1816</th>
<th>A4507</th>
<th>A6358</th>
<th>A6375</th>
<th>D. dadantii 3937</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pectinase secretion</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Cellulase secretion</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Protease secretion</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Lipase secretion</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Swimming motility (mm)</td>
<td>6±3</td>
<td>14±2</td>
<td>5±1</td>
<td>11±4</td>
<td>16±2</td>
<td>1±0</td>
<td>19±4</td>
</tr>
<tr>
<td>Swarming motility (mm)</td>
<td>38±6</td>
<td>7±2</td>
<td>6±2</td>
<td>41±8</td>
<td>37±6</td>
<td>8±3</td>
<td>51±9</td>
</tr>
<tr>
<td>Chicory leaf maceration (mm)</td>
<td>10.7±1.9</td>
<td>7.2±3.2</td>
<td>2.2±2.2</td>
<td>9.3±2.9</td>
<td>9±1.9</td>
<td>6.1±3.8</td>
<td>37.7±8.6</td>
</tr>
<tr>
<td>Potato tuber maceration (g)</td>
<td>0.43±0.09</td>
<td>0.16±0.07</td>
<td>0.06±0.01</td>
<td>0.11±0.06</td>
<td>0.21±0.08</td>
<td>0.26±0.12</td>
<td>1.68±0.52</td>
</tr>
</tbody>
</table>
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.

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

The lower left triangle displays the ANI values (%) and the right-upper triangle displays the dDDH values. The ANI values were calculated based on pairwise comparisons between A3967\(^T\) (CFBP 8732\(^T\)) and other genomes ([http://enve-omics.ce.gatech.edu/ani/](http://enve-omics.ce.gatech.edu/ani/)) [36]. The dDDH values were calculated using the A3967 genome as a reference and other *Dickeya* genomes as queries ([http://ggdc.dsmz.de/](http://ggdc.dsmz.de/)) [37]. This analysis was performed using the genomes of strains CFBP 4178\(^T\) (GCA_000400505.1), Ech703 (GCA_000023545.1), *D. aquatica* CFBP 8348\(^T\) (GCA_900095885.1), *D. chrysanthemi* CFBP 2048\(^T\) (GCA_000406105.1), *D. dianthicola* CFBP 1269\(^T\) (GCA_003049785.1), *D. dianthicola* CFBP 1200\(^T\) (GCA_000365305.1), *D. fangzhongdai* CFBP 8607\(^T\) (GCA_002812485.1), *D. lacustris* 529\(^T\) (GCA_003934295.1), *D. poaceiphila* CFBP 8731\(^T\) (GCA_007858975.2), *D. solani* CFBP 7345\(^T\) (GCA_001644705.1), *D. undicola* CFBP 8650\(^T\) (GCA_000784735.1), *D. zeae* CFBP 2052\(^T\) (GCA_000406165.1) and *P. carotovorum* CFBP 2046\(^T\) (GCA_000749855.1).

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

<table>
<thead>
<tr>
<th></th>
<th>Dickeya zeae CFBP 2052(^T)</th>
<th>Dickeya chrysanthemi CFBP 2048(^T)</th>
<th>Dickeya poaceiphila NCPPB 569(^T)</th>
<th>Dickeya fangzhongdai CFBP 8607(^T)</th>
<th>Dickeya dianthicola CFBP 1200(^T)</th>
<th>Dickeya solani CFBP 7345(^T)</th>
<th>Dickeya dadantii CFBP 1269(^T)</th>
<th>Dickeya undicola CFBP 8650(^T)</th>
<th>Dickeya lacustris CFBP 8647(^T)</th>
<th>Dickeya aquatica CFBP 8348(^T)</th>
<th>Ech 703</th>
<th>CFBP 4178(^T)</th>
<th>A3967(^T) (CFBP 8732(^T))</th>
<th>Pectobacterium carotovorum CFBP 2046(^T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANI</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>75.16</td>
</tr>
<tr>
<td>Dickeya zeae CFBP 2052(^T)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>75.16</td>
</tr>
<tr>
<td>Dickeya chrysanthemi CFBP 2048(^T)</td>
<td>86.57</td>
<td>81.65</td>
<td>81.25</td>
<td>79.50</td>
<td>80.05</td>
<td>80.75</td>
<td>80.75</td>
<td>80.75</td>
<td>80.75</td>
<td>80.75</td>
<td>22.0</td>
<td>22.0</td>
<td>22.0</td>
<td>75.16</td>
</tr>
<tr>
<td>Dickeya poaceiphila NCPPB 569(^T)</td>
<td>85.55</td>
<td>85.10</td>
<td>84.65</td>
<td>85.09</td>
<td>85.09</td>
<td>85.09</td>
<td>85.09</td>
<td>85.09</td>
<td>85.09</td>
<td>85.09</td>
<td>22.0</td>
<td>22.0</td>
<td>22.0</td>
<td>75.16</td>
</tr>
<tr>
<td>Dickeya fangzhongdai CFBP 8607(^T)</td>
<td>85.09</td>
<td>85.09</td>
<td>85.09</td>
<td>85.09</td>
<td>85.09</td>
<td>85.09</td>
<td>85.09</td>
<td>85.09</td>
<td>85.09</td>
<td>85.09</td>
<td>22.0</td>
<td>22.0</td>
<td>22.0</td>
<td>75.16</td>
</tr>
<tr>
<td>Dickeya dianthicola CFBP 1200(^T)</td>
<td>85.23</td>
<td>85.23</td>
<td>85.23</td>
<td>85.23</td>
<td>85.23</td>
<td>85.23</td>
<td>85.23</td>
<td>85.23</td>
<td>85.23</td>
<td>85.23</td>
<td>22.0</td>
<td>22.0</td>
<td>22.0</td>
<td>75.16</td>
</tr>
<tr>
<td>Dickeya solani CFBP 7345(^T)</td>
<td>85.02</td>
<td>85.02</td>
<td>85.02</td>
<td>85.02</td>
<td>85.02</td>
<td>85.02</td>
<td>85.02</td>
<td>85.02</td>
<td>85.02</td>
<td>85.02</td>
<td>22.0</td>
<td>22.0</td>
<td>22.0</td>
<td>75.16</td>
</tr>
<tr>
<td>Dickeya dadantii CFBP 1269(^T)</td>
<td>84.93</td>
<td>84.93</td>
<td>84.93</td>
<td>84.93</td>
<td>84.93</td>
<td>84.93</td>
<td>84.93</td>
<td>84.93</td>
<td>84.93</td>
<td>84.93</td>
<td>22.0</td>
<td>22.0</td>
<td>22.0</td>
<td>75.16</td>
</tr>
<tr>
<td>Dickeya undicola CFBP 8650(^T)</td>
<td>84.31</td>
<td>84.31</td>
<td>84.31</td>
<td>84.31</td>
<td>84.31</td>
<td>84.31</td>
<td>84.31</td>
<td>84.31</td>
<td>84.31</td>
<td>84.31</td>
<td>22.0</td>
<td>22.0</td>
<td>22.0</td>
<td>75.16</td>
</tr>
<tr>
<td>Dickeya lacustris CFBP 8647(^T)</td>
<td>79.63</td>
<td>79.63</td>
<td>79.63</td>
<td>79.63</td>
<td>79.63</td>
<td>79.63</td>
<td>79.63</td>
<td>79.63</td>
<td>79.63</td>
<td>79.63</td>
<td>22.0</td>
<td>22.0</td>
<td>22.0</td>
<td>75.16</td>
</tr>
<tr>
<td>Dickeya aquatica CFBP 8348(^T)</td>
<td>80.21</td>
<td>80.21</td>
<td>80.21</td>
<td>80.21</td>
<td>80.21</td>
<td>80.21</td>
<td>80.21</td>
<td>80.21</td>
<td>80.21</td>
<td>80.21</td>
<td>22.0</td>
<td>22.0</td>
<td>22.0</td>
<td>75.16</td>
</tr>
<tr>
<td>Ech 703</td>
<td>78.30</td>
<td>78.30</td>
<td>78.30</td>
<td>78.30</td>
<td>78.30</td>
<td>78.30</td>
<td>78.30</td>
<td>78.30</td>
<td>78.30</td>
<td>78.30</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>68.4</td>
</tr>
<tr>
<td>CFBP 4178(^T)</td>
<td>78.25</td>
<td>78.25</td>
<td>78.25</td>
<td>78.25</td>
<td>78.25</td>
<td>78.25</td>
<td>78.25</td>
<td>78.25</td>
<td>78.25</td>
<td>78.25</td>
<td>99.98</td>
<td>100</td>
<td>100</td>
<td>68.3</td>
</tr>
<tr>
<td>A3967(^T) (CFBP 8732(^T))</td>
<td>78.17</td>
<td>78.17</td>
<td>78.17</td>
<td>78.17</td>
<td>78.17</td>
<td>78.17</td>
<td>78.17</td>
<td>78.17</td>
<td>78.17</td>
<td>78.17</td>
<td>96.17</td>
<td>96.21</td>
<td>100</td>
<td>74.89</td>
</tr>
</tbody>
</table>
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.
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.
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 4178T, 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$^T$. 

![Graph showing growth temperatures](image-url)
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. zeae, 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.
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.
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. diantico, D. poaceiphila, D. zeae, 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.

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

<table>
<thead>
<tr>
<th>Strain designations</th>
<th>Origin: country, year, plant</th>
<th>Isolated by</th>
<th>New classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFBP 4178&lt;sup&gt;T&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCPPB 2511&lt;sup&gt;T&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMG 2542&lt;sup&gt;T&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1816</td>
<td>Columbia, 1972, <em>Musa paradisiaca</em></td>
<td>Victoria JI</td>
<td><em>Musicola paradisiaca</em></td>
</tr>
<tr>
<td>CFBP 1445</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4507</td>
<td>Columbia, 1968, <em>Musa paradisiaca</em></td>
<td>Victoria JI</td>
<td><em>Musicola paradisiaca</em></td>
</tr>
<tr>
<td>CFBP 3477</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6358</td>
<td>Cuba, 1987, <em>Musa paradisiaca</em></td>
<td>Rivera N</td>
<td><em>Musicola paradisiaca</em></td>
</tr>
<tr>
<td>CFBP 3696</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6375</td>
<td>Cuba, 1987, <em>Zea mays</em></td>
<td>Rivera N</td>
<td><em>Musicola paradisiaca</em></td>
</tr>
<tr>
<td>CFBP 3699</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFBP 4430</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3967&lt;sup&gt;T&lt;/sup&gt;</td>
<td>France, 1965, <em>Solanum lycopersicon</em></td>
<td>Prunier JP</td>
<td><em>Musicola keenii</em> type strain</td>
</tr>
<tr>
<td>CFBP 722, CFBP 8732&lt;sup&gt;T&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMG 31880&lt;sup&gt;T&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFBP 1446</td>
<td>Columbia, 1972, <em>Musa paradisiaca</em></td>
<td>Victoria JI</td>
<td><em>Múscola paradisiaca</em></td>
</tr>
<tr>
<td>CFBP 1451</td>
<td>Columbia, 1972, <em>Musa paradisiaca</em></td>
<td>Granada G</td>
<td><em>Musicola paradisiaca</em></td>
</tr>
</tbody>
</table>

*A* : collection of the laboratory Microbiology, Adaptation and Pathogenicity, Lyon, France  
CFBP : Collection Française de Bactéries Phytopathogènes, Beaucouzé, France  
LMG : collection of the Laboratory of Microbiology, Ghent, Belgium  
NCPPB : National Collection of Plant Pathogenic Bacteria, York, UK
Table S2. Metabolic capacities of the type strains of \textit{Musicola keenii} A3967\textsuperscript{T} (CFBP 8732\textsuperscript{T}) and \textit{Musicola paradisiaca} CFBP 4178\textsuperscript{T}

The \textit{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 $\lambda$=590nm: -, indicates OD <0.2; w, indicates 0.2≤OD≤0.5; +, indicates 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).

<table>
<thead>
<tr>
<th>Carbon sources</th>
<th>Strains</th>
<th>A6065\textsuperscript{T} (CFBP 4178\textsuperscript{T})</th>
<th>A3967\textsuperscript{T} (CFBP 8732\textsuperscript{T})</th>
<th>3937 D. dadantii</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Arabinose</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>N-Acetyl-D-glucosamine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D-Saccharic acid (glucaric acid)</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Succinic acid</td>
<td>w</td>
<td>+</td>
<td>w</td>
<td>w</td>
</tr>
<tr>
<td>D-Galactose</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>L-Aspartic acid</td>
<td>w</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>L-Proline</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-Alanine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-Trehalose</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-Mannose</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Dulcitol</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-Serine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-Sorbitol</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Glycerol</td>
<td>w</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>L-Fucose</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-Glucuronic acid</td>
<td>w</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-Glucosonic acid</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D,L-α-Glycerol- phosphate</td>
<td>w</td>
<td>w</td>
<td>w</td>
<td>w</td>
</tr>
<tr>
<td>D-Xylose</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>L-Lactic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Formic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-Mannitol</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>L-Glutamic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>w</td>
</tr>
<tr>
<td>D-Glucose-6-phosphate</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D-Galactonic acid-γ-lactone</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D,L-Malic acid</td>
<td>w</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D-Ribose</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Tween 20</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>L-Rhamnose</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-Fructose</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Acetic acid</td>
<td>-</td>
<td>-</td>
<td>w</td>
<td>-</td>
</tr>
<tr>
<td>α-D-Glucose</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Maltose</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-Melibiose</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Thymidine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>L-Asparagine</td>
<td>w</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D-Aspartic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>D-Glucosaminic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1,2-Propanediol</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tween 40</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>α-Keto-glutaric acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>α-Keto-butyric acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>α-Methyl-D-galactoside</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>α-D-Lactose</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lactulose</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sucrose</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Uridine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>L-Glutamine</td>
<td>-</td>
<td>-</td>
<td>w</td>
<td>-</td>
</tr>
<tr>
<td>m-Tartaric Acid</td>
<td>-</td>
<td>-</td>
<td>w</td>
<td>-</td>
</tr>
<tr>
<td>D-Glucose-1-Phosphate</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Compound</td>
<td>Data 1</td>
<td>Data 2</td>
<td>Data 3</td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>D-Fructose-6-Phosphate</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Tween 80</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>α-Hydroxy glutaric acid-γ-lactone</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>α-Hydroxy butyric acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>β-Methyl-D-glucoside</td>
<td>W</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Adonitol</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Maltotriose</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>2-Deoxy adenosine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Adenosine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Glycyl-L-aspartic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Citric acid</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>myo-Inositol</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>D-Threonine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Fumaric acid</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Bromo succinic acid</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td></td>
</tr>
<tr>
<td>Propionic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Mucic acid (galactaric acid)</td>
<td>-</td>
<td>W</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Glycolic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Glyoxylic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>D-Cellobiose</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Inosine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Glycylyl-L-glutamic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Tricarballylic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>L-Serine</td>
<td>-</td>
<td>W</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>L-Threonine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>L-Alanine</td>
<td>-</td>
<td>W</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>L-Alanyl-glycine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Acetooacetic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>N-Acetyl-β-D-mannosamine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Mono methyl succinate</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Methyl pyruvate</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>D-Malic acid</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td></td>
</tr>
<tr>
<td>L-Malic acid</td>
<td>W</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Glycyl-L-proline</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>p-Hydroxy phenyl acetic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>m-Hydroxy phenyl acetic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Tyramine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>D-Psicose</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>L-Lyxose</td>
<td>-</td>
<td>W</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Glucuronamide</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Pyruvic acid</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>L-Galactonic acid-γ-lactone</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>D-Galacturonic acid</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Phenylethyl-amine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>2-Aminoethanol</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Chondroitin sulfate C</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>β-Cyclodextrin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>α-Cyclodextrin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>γ-Cyclodextrin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Dextrin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Gelatin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Glycogen</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Inulin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Laminarin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Mannan</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Pectin</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>N-Acetyl-D-galactosamine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>N-Acetyl-neuraminic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>β-3-D-Allose</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Amygdalin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>D-Arabinose</td>
<td>-</td>
<td>-</td>
<td>W</td>
<td></td>
</tr>
<tr>
<td>D-Arabitol</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>L-Arabitol</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Arbutin</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>2-Deoxy-D-ribose</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
I-Erythritol
D-Fucose
3-0-β-D-Galacto-pyranosyl-D-arabinose
Gentiobiose
L-Glucose
Lactitol
D-Melezitose
Maltitol
a-Methyl-D-glucoside
β-Methyl-D-galactoside
3-Methyl Glucose
β-Methyl-D-glucuronic acid
α-Methyl-D-mannoside
β-Methyl-D-xyloside
Palatinose

D-Raffinose
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
Melibonic acid
Oxalic acid
Oxalomalic acid
Quinic acid
D-Ribono-1,4-lactone
Sebacic acid
Sorbic acid
Succinamic acid
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
<table>
<thead>
<tr>
<th>Compound</th>
<th>-</th>
<th>-</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Phenylalanine</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>L-Pyroglutamic acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>L-Valine</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D,L-Carnitine</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sec-Butylamine</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D,L-Octopamine</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Putrescine</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dihydroxy Acetone</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2,3-Butanediol</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2,3-Butanedione</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3-Hydroxy 2-Butanone</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Protein</th>
<th>Protein annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>56 transposase for IS1001 element</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>54 4-oxalocrotonate tautomerase</td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>150 UPF0306 protein YhbP</td>
<td></td>
</tr>
<tr>
<td>134</td>
<td>271 ABC transporter, permease protein 2 (cluster 5, nickel/peptides/opines)</td>
<td></td>
</tr>
<tr>
<td>277</td>
<td>366 Predicted cell-wall-anchored protein SasA (LPXTG motif)</td>
<td></td>
</tr>
<tr>
<td>280</td>
<td>509 PilV-like protein</td>
<td></td>
</tr>
<tr>
<td>286</td>
<td>473 Conjugal transfer protein traA</td>
<td></td>
</tr>
<tr>
<td>287</td>
<td>21 Mobile element protein</td>
<td></td>
</tr>
<tr>
<td>694</td>
<td>230 Expansin-YoAl</td>
<td></td>
</tr>
<tr>
<td>797</td>
<td>29 Mobile element protein</td>
<td></td>
</tr>
<tr>
<td>884</td>
<td>21 Mobile element protein</td>
<td></td>
</tr>
<tr>
<td>885</td>
<td>42 Type I restriction-modification system, restriction subunit R (EC 3.1.21.3)</td>
<td></td>
</tr>
<tr>
<td>915</td>
<td>91 Integrate</td>
<td></td>
</tr>
<tr>
<td>982</td>
<td>234 ATPase</td>
<td></td>
</tr>
<tr>
<td>984</td>
<td>258 putative membrane protein</td>
<td></td>
</tr>
<tr>
<td>985</td>
<td>697 Coupling protein VirD4, ATPase required for T-DNA transfer</td>
<td></td>
</tr>
<tr>
<td>991</td>
<td>343 Putative ATP-binding protein</td>
<td></td>
</tr>
<tr>
<td>993</td>
<td>412 Inc1 plasmid conjugative transfer pilus-tip adhesin protein PilV</td>
<td></td>
</tr>
<tr>
<td>994</td>
<td>218 Type-IV secretion leader peptidase/N-methyltransferase</td>
<td></td>
</tr>
<tr>
<td>995</td>
<td>163 Inc1 plasmid conjugative transfer putative membrane protein PilT</td>
<td></td>
</tr>
<tr>
<td>996</td>
<td>193 Inc1 plasmid conjugative transfer prepilin PilS</td>
<td></td>
</tr>
<tr>
<td>997</td>
<td>370 Inc1 plasmid conjugative transfer inner membrane protein PilR</td>
<td></td>
</tr>
<tr>
<td>999</td>
<td>167 Inc1 plasmid pilus assembly protein PilP</td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>565 Inc1 plasmid conjugative transfer lipoprotein PilN</td>
<td></td>
</tr>
<tr>
<td>1003</td>
<td>403 Inc1 plasmid conjugative transfer protein PilL</td>
<td></td>
</tr>
<tr>
<td>1006</td>
<td>160 Putative uncharacterized protein STY4534</td>
<td></td>
</tr>
<tr>
<td>1016</td>
<td>586 Protein with ParB-like nuclease domain in PFGI-1-like cluster</td>
<td></td>
</tr>
<tr>
<td>1021</td>
<td>296 Chromosome partitioning ATPase in PFGI-1-like cluster, ParA-like</td>
<td></td>
</tr>
<tr>
<td>1028</td>
<td>37 Mobile element protein</td>
<td></td>
</tr>
<tr>
<td>1037</td>
<td>321 DNA methylase N-4/N-6</td>
<td></td>
</tr>
<tr>
<td>1042</td>
<td>78 Mobile element protein</td>
<td></td>
</tr>
<tr>
<td>1043</td>
<td>247 Transposase</td>
<td></td>
</tr>
<tr>
<td>1046</td>
<td>160 Arsenical pump-driving ATPase (EC 3.6.3.16) TEMP</td>
<td></td>
</tr>
<tr>
<td>1047</td>
<td>122 Arsenical resistance operon trans-acting repressor ArsD</td>
<td></td>
</tr>
<tr>
<td>1134</td>
<td>40 Transposase, IS3/IS911 family</td>
<td></td>
</tr>
<tr>
<td>1259</td>
<td>338 CiaB PROTEIN</td>
<td></td>
</tr>
<tr>
<td>1706</td>
<td>377 Putative exported protein precursor</td>
<td></td>
</tr>
<tr>
<td>1898</td>
<td>49 Mobile element protein</td>
<td></td>
</tr>
<tr>
<td>2085</td>
<td>61 Uncharacterized MFS-type transporter</td>
<td></td>
</tr>
<tr>
<td>2151</td>
<td>91 Transcriptional regulator, AsnC family</td>
<td></td>
</tr>
<tr>
<td>2328</td>
<td>130 SSU ribosomal protein S11p (S14e)</td>
<td></td>
</tr>
<tr>
<td>2361</td>
<td>107 YkfI toxin protein</td>
<td></td>
</tr>
<tr>
<td>2362</td>
<td>112 YafW protein (antitoxin to YkfI)</td>
<td></td>
</tr>
<tr>
<td>2363</td>
<td>158 UPF0758 family protein</td>
<td></td>
</tr>
<tr>
<td>2364</td>
<td>274 UPF0380 proteins YafZ and homologs</td>
<td></td>
</tr>
<tr>
<td>2403</td>
<td>21 Mobile element protein</td>
<td></td>
</tr>
<tr>
<td>2421</td>
<td>229</td>
<td>YjbF outer membrane lipoprotein</td>
</tr>
<tr>
<td>2423</td>
<td>713</td>
<td>Uncharacterized lipoprotein YjhH</td>
</tr>
<tr>
<td>2435</td>
<td>480</td>
<td>2-methylcitrate dehydratase (EC 4.2.1.79)</td>
</tr>
<tr>
<td>2437</td>
<td>402</td>
<td>Acyltransferase 3</td>
</tr>
<tr>
<td>2439</td>
<td>733</td>
<td>ATP-dependent helicase</td>
</tr>
<tr>
<td>2440</td>
<td>439</td>
<td>ATP/GTP-binding protein</td>
</tr>
<tr>
<td>2441</td>
<td>468</td>
<td>Mobile element protein</td>
</tr>
<tr>
<td>2442</td>
<td>21</td>
<td>Mobile element protein</td>
</tr>
<tr>
<td>2451</td>
<td>45</td>
<td>Assimilatory nitrate reductase large subunit (EC 1.7.99.4)</td>
</tr>
<tr>
<td>2588</td>
<td>49</td>
<td>COG5499: Predicted transcription regulator containing HTH domain</td>
</tr>
<tr>
<td>2781</td>
<td>357</td>
<td>Undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase (EC 2.7.8.33)</td>
</tr>
<tr>
<td>2782</td>
<td>379</td>
<td>Putative polysaccharide export protein YccZ precursor</td>
</tr>
<tr>
<td>2783</td>
<td>123</td>
<td>Low molecular weight protein-tyrosine-phosphatase (EC 3.1.3.48) =&gt; Etp</td>
</tr>
<tr>
<td>2784</td>
<td>723</td>
<td>Tyrosine-protein kinase (EC 2.7.10.2)</td>
</tr>
<tr>
<td>2800</td>
<td>402</td>
<td>Radical SAM domain protein</td>
</tr>
<tr>
<td>2822</td>
<td>257</td>
<td>Branched-chain amino acid ABC transporter, amino acid-binding protein (TC 3.A.1.4.1)</td>
</tr>
<tr>
<td>2825</td>
<td>21</td>
<td>Mobile element protein</td>
</tr>
<tr>
<td>2838</td>
<td>53</td>
<td>Homoserine kinase (EC 2.7.1.39)</td>
</tr>
<tr>
<td>3124</td>
<td>56</td>
<td>Transposase for IS1001 element</td>
</tr>
<tr>
<td>3175</td>
<td>129</td>
<td>Transposase IS3/IS911</td>
</tr>
<tr>
<td>3388</td>
<td>32</td>
<td>Transcriptional regulator, GntR family domain</td>
</tr>
<tr>
<td>3389</td>
<td>40</td>
<td>DNA-binding transcriptional regulator, MocR family</td>
</tr>
<tr>
<td>3404</td>
<td>216</td>
<td>Transcriptional regulator KPN_02146, AcrR family</td>
</tr>
<tr>
<td>3417</td>
<td>72</td>
<td>transposase for IS1001 element</td>
</tr>
<tr>
<td>3418</td>
<td>179</td>
<td>transposase for IS1001 element</td>
</tr>
<tr>
<td>3420</td>
<td>141</td>
<td>Oligopeptide transport ATP-binding protein OppD (TC 3.A.1.5.1)</td>
</tr>
<tr>
<td>3421</td>
<td>54</td>
<td>transposase for IS1001 element</td>
</tr>
<tr>
<td>3422</td>
<td>109</td>
<td>Flagellar hook-length control protein FliK</td>
</tr>
<tr>
<td>3434</td>
<td>276</td>
<td>5-deoxy-glucuronate isomerase (EC 5.3.1.1) IolB, myo-insositol catabolism</td>
</tr>
<tr>
<td>3435</td>
<td>297</td>
<td>Inosose dehydratase (EC 4.2.1.44) IolE, myo-insositol catabolism</td>
</tr>
<tr>
<td>3445</td>
<td>154</td>
<td>DNA gyrase inhibitory protein</td>
</tr>
<tr>
<td>3447</td>
<td>504</td>
<td>Malonate-semialdehyde dehydrogenase [inositol] (EC 1.2.1.18) IolA, myo-insositol catabolism</td>
</tr>
<tr>
<td>3509</td>
<td>122</td>
<td>Methyl-accepting chemotaxis protein I (serine chemoreceptor protein)</td>
</tr>
<tr>
<td>3529</td>
<td>429</td>
<td>L-rhamnonate transporter (predicted by genome context)</td>
</tr>
<tr>
<td>3626</td>
<td>123</td>
<td>Putative regulatory protein</td>
</tr>
<tr>
<td>3629</td>
<td>233</td>
<td>(adenine-N6-)-methyltransferase homolog</td>
</tr>
<tr>
<td>3637</td>
<td>168</td>
<td>Regulatory protein CII bacteriophage 186</td>
</tr>
<tr>
<td>3638</td>
<td>100</td>
<td>Phage regulatory protein</td>
</tr>
<tr>
<td>3686</td>
<td>65</td>
<td>Integrase</td>
</tr>
<tr>
<td>3732</td>
<td>86</td>
<td>hypothetical protein formerly called flagellar hook-length control protein FliK</td>
</tr>
<tr>
<td>3733</td>
<td>85</td>
<td>hypothetical protein formerly called flagellar hook-length control protein FliK</td>
</tr>
<tr>
<td>3735</td>
<td>71</td>
<td>hypothetical protein formerly called flagellar hook-length control protein FliK</td>
</tr>
<tr>
<td>3977</td>
<td>280</td>
<td>ATPase</td>
</tr>
<tr>
<td>3981</td>
<td>71</td>
<td>probable membrane protein STY4566</td>
</tr>
<tr>
<td>3986</td>
<td>138</td>
<td>putative lipoprotein</td>
</tr>
<tr>
<td>3987</td>
<td>940</td>
<td>Type IV secretory pathway, VirB4 components</td>
</tr>
<tr>
<td>3988</td>
<td>129</td>
<td>conserved hypothetical protein</td>
</tr>
<tr>
<td>3990</td>
<td>67</td>
<td>corresponds to STY4575 from Accession AL513382: Salmonella typhi CT18</td>
</tr>
<tr>
<td>3991</td>
<td>48</td>
<td>corresponds to STY4575 from Accession AL513382: Salmonella typhi CT18</td>
</tr>
<tr>
<td>3993</td>
<td>47</td>
<td>Type IV secretory pathway, VirB4 components</td>
</tr>
<tr>
<td>4002</td>
<td>119</td>
<td>transcriptional regulator, MerR family</td>
</tr>
<tr>
<td>4003</td>
<td>516</td>
<td>putative membrane protein</td>
</tr>
<tr>
<td>Page</td>
<td>Column 1</td>
<td>Column 2</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>4022</td>
<td>260</td>
<td>VrlR-like protein</td>
</tr>
<tr>
<td>4023</td>
<td>1170</td>
<td>DEAD/DEAH box helicase domain protein</td>
</tr>
<tr>
<td>4025</td>
<td>581</td>
<td>Bipolar DNA helicase HerA</td>
</tr>
<tr>
<td>4235</td>
<td>189</td>
<td>D-alanyl-D-alanine dipeptidase (EC 3.4.13.22)</td>
</tr>
<tr>
<td>4275</td>
<td>39</td>
<td>Maltodextrin ABC transporter, substrate-binding protein MdxE</td>
</tr>
<tr>
<td>4297</td>
<td>304</td>
<td>Siderophore a chromobactin ABC transporter, substrate-binding protein</td>
</tr>
<tr>
<td>4316</td>
<td>88</td>
<td>beta-glucosidase (EC 3.2.1.21); 6-phospho-beta-glucosidase (EC 3.2.1.86)</td>
</tr>
<tr>
<td>4317</td>
<td>156</td>
<td>beta-glucosidase (EC 3.2.1.21); 6-phospho-beta-glucosidase (EC 3.2.1.86)</td>
</tr>
</tbody>
</table>
Table S4. Genes present in *M. paradisiaca* CFBP 4178<sup>T</sup> but absent in strain *M. keenii* A3967<sup>T</sup> (CFBP 8732<sup>T</sup>);

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.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Protein Number</th>
<th>Protein Length</th>
<th>Protein annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>926</td>
<td>DNA helicase</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>466</td>
<td>DNA-cytosine methyltransferase (EC 2.1.1.37)</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>342</td>
<td>Uncharacterized protein RhuM</td>
<td></td>
</tr>
<tr>
<td>489</td>
<td>41</td>
<td>Mobile element protein</td>
<td></td>
</tr>
<tr>
<td>577</td>
<td>187</td>
<td>Acetyltransferase, GNAT family</td>
<td></td>
</tr>
<tr>
<td>716</td>
<td>50</td>
<td>Cell division protein FtsQ</td>
<td></td>
</tr>
<tr>
<td>760</td>
<td>51</td>
<td>Alcohol dehydrogenase (EC 1.1.1.1)</td>
<td></td>
</tr>
<tr>
<td>796</td>
<td>627</td>
<td>Retron-type RNA-directed DNA polymerase (EC 2.7.7.49)</td>
<td></td>
</tr>
<tr>
<td>807</td>
<td>185</td>
<td>CRISPR-associated protein, Csy4 family</td>
<td></td>
</tr>
<tr>
<td>808</td>
<td>337</td>
<td>CRISPR-associated protein, Csy3 family</td>
<td></td>
</tr>
<tr>
<td>809</td>
<td>317</td>
<td>CRISPR-associated protein, Csy2 family</td>
<td></td>
</tr>
<tr>
<td>810</td>
<td>442</td>
<td>CRISPR-associated protein, Csy1 family</td>
<td></td>
</tr>
<tr>
<td>811</td>
<td>1096</td>
<td>CRISPR-associated helicase Cas3</td>
<td></td>
</tr>
<tr>
<td>812</td>
<td>334</td>
<td>CRISPR-associated protein Cas1</td>
<td></td>
</tr>
<tr>
<td>813</td>
<td>166</td>
<td>Sensory box histidine kinase</td>
<td></td>
</tr>
<tr>
<td>982</td>
<td>38</td>
<td>Mobile element protein</td>
<td></td>
</tr>
<tr>
<td>1069</td>
<td>401</td>
<td>Macrolide-efflux protein</td>
<td></td>
</tr>
<tr>
<td>1070</td>
<td>224</td>
<td>Thymidylate kinase (EC 2.7.4.9)</td>
<td></td>
</tr>
<tr>
<td>1071</td>
<td>381</td>
<td>Biotin synthase-related enzyme</td>
<td></td>
</tr>
<tr>
<td>1073</td>
<td>188</td>
<td>Cell division trigger factor (EC 5.2.1.8)</td>
<td></td>
</tr>
<tr>
<td>1250</td>
<td>363</td>
<td>ABC transporter, substrate-binding protein (cluster 10, nitrate/sulfonate/bicarbonate)</td>
<td></td>
</tr>
<tr>
<td>1488</td>
<td>427</td>
<td>Raffinose and melibiose permease, RafT</td>
<td></td>
</tr>
<tr>
<td>1489</td>
<td>710</td>
<td>alpha-galactosidase (EC 3.2.1.22), RafA, raffinose and melibiose catabolism</td>
<td></td>
</tr>
<tr>
<td>1504</td>
<td>457</td>
<td>Maltodextrin ABC transporter, substrate-binding protein MdxE</td>
<td></td>
</tr>
<tr>
<td>1508</td>
<td>288</td>
<td>Metal-dependent hydrolase</td>
<td></td>
</tr>
<tr>
<td>1566</td>
<td>448</td>
<td>N-acetylgalactosamine-regulated outer membrane porin</td>
<td></td>
</tr>
<tr>
<td>1592</td>
<td>104</td>
<td>Type I restriction-modification system, restriction subunit R (EC 3.1.21.3)</td>
<td></td>
</tr>
<tr>
<td>1596</td>
<td>421</td>
<td>Polyketide biosynthesis 3-hydroxy-3-methylglutaryl-ACP synthase PksG</td>
<td></td>
</tr>
<tr>
<td>1598</td>
<td>84</td>
<td>Acyl carrier protein</td>
<td></td>
</tr>
<tr>
<td>1601</td>
<td>435</td>
<td>Monoxygenase, flavin-binding family</td>
<td></td>
</tr>
<tr>
<td>1604</td>
<td>6877</td>
<td>Modular polyketide synthase</td>
<td></td>
</tr>
<tr>
<td>1609</td>
<td>5613</td>
<td>Modular polyketide synthase</td>
<td></td>
</tr>
<tr>
<td>1610</td>
<td>5613</td>
<td>Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39); Enoyl-[acyl-carrier-protein] reductase [FMN]</td>
<td></td>
</tr>
<tr>
<td>1611</td>
<td>275</td>
<td>Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39)</td>
<td></td>
</tr>
<tr>
<td>1612</td>
<td>216</td>
<td>Modular polyketide synthase</td>
<td></td>
</tr>
<tr>
<td>1614</td>
<td>161</td>
<td>Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39)</td>
<td></td>
</tr>
<tr>
<td>1803</td>
<td>188</td>
<td>Transcriptional regulator, AcrR family</td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td>268</td>
<td>Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39)</td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>232</td>
<td>Putative preQ0 transporter</td>
<td></td>
</tr>
<tr>
<td>2041</td>
<td>352</td>
<td>Oxidoreductase</td>
<td></td>
</tr>
<tr>
<td>2042</td>
<td>281</td>
<td>putative prolly aminopeptidase</td>
<td></td>
</tr>
<tr>
<td>2045</td>
<td>401</td>
<td>cobalt dependent X-Pro dipeptidase</td>
<td></td>
</tr>
<tr>
<td>2062</td>
<td>217</td>
<td>Transcriptional regulator, AcrR family</td>
<td></td>
</tr>
<tr>
<td>2086</td>
<td>64</td>
<td>Sodium-dependent phosphate transporter</td>
<td></td>
</tr>
<tr>
<td>2250</td>
<td>138</td>
<td>FIG00904844: hypothetical protein</td>
<td></td>
</tr>
<tr>
<td>2257</td>
<td>104</td>
<td>Death on curing protein, Doc toxin</td>
<td></td>
</tr>
<tr>
<td>Gene ID</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2348</td>
<td>Tail-specific protease precursor (EC 3.4.21.102)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2359</td>
<td>Replication gene B protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2365</td>
<td>Putative phage tail protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2366</td>
<td>Putative phage tail protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2455</td>
<td>Putative membrane protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2457</td>
<td>Adenine phosphoribosyltransferase (EC 2.4.2.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2550</td>
<td>Hypothetical MFS-type transporter protein YcaD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2668</td>
<td>Integrase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2685</td>
<td>Prophage protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2686</td>
<td>Prophage protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2687</td>
<td>DNA primase (EC 2.7.7.-)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2798</td>
<td>ISSod13, transposase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2835</td>
<td>Formyl-coenzyme A transferase (EC 2.8.3.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2838</td>
<td>Formyl-coenzyme A transferase (EC 2.8.3.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2839</td>
<td>Formyl-coenzyme A transferase (EC 2.8.3.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2939</td>
<td>tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase (EC 2.1.1.61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2966</td>
<td>Transcriptional regulator, TetR family</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2972</td>
<td>Nitrogenase (iron-iron) delta chain (EC 1.18.6.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2974</td>
<td>DNA recombination protein RmuC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2977</td>
<td>AnfO protein, required for Mo- and V-independent nitrogenase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2978</td>
<td>Collagenase and related proteases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3037</td>
<td>Inosine-5’-monophosphate dehydrogenase (EC 1.1.1.205) / CBS domain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3348</td>
<td>Phage integrase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3360</td>
<td>TriL protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3362</td>
<td>CcdA protein (antitoxin to CcdB)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3438</td>
<td>CcdB toxin protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3440</td>
<td>FIGO06126. DNA helicase, restriction/modification system component YeeB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3441</td>
<td>FIGO45374. Type II restriction enzyme, methylase subunit YeeA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3442</td>
<td>Phage DNA invertase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3452</td>
<td>Putative integrase protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3454</td>
<td>Nodulation protein nolO (EC 2.1.3.-)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3461</td>
<td>Polyketide synthase?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3464</td>
<td>Glyoxalase family protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3465</td>
<td>Polyketide synthase modules and related proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3472</td>
<td>Split AAA-ATPase protein PA0787</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3484</td>
<td>Acyl-coenzyme A:6-aminopenicillanic-acetyltransferase 40 kDa form (EC 2.3.1.164)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3499</td>
<td>Thioesterase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3501</td>
<td>Putative membrane protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3505</td>
<td>CcdA protein (antitoxin to CcdB)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3506</td>
<td>CcdB toxin protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3512</td>
<td>Putative cryptic D-serine deaminase (EC 4.3.1.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3516</td>
<td>Arsenate reductase (EC 1.20.4.1) glutaredoxin-coupled, glutaredoxin-like family</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3518</td>
<td>Putative membrane protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3520</td>
<td>Arsenical-resistance protein ACR3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3521</td>
<td>Arsenical pump-driving ATPase (EC 3.6.3.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3557</td>
<td>Predicted permeases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3698</td>
<td>Group 1 glycosyl transferase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3755</td>
<td>RNA polymerase ECF-type sigma factor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3756</td>
<td>Fe2+-dicitrate sensor, membrane component</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3852</td>
<td>TriL protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>3853</td>
<td>73</td>
<td>CcdA protein (antitoxin to CcdB)</td>
<td></td>
</tr>
<tr>
<td>3854</td>
<td>102</td>
<td>CcdB toxin protein</td>
<td></td>
</tr>
<tr>
<td>3855</td>
<td>333</td>
<td>Retron-type RNA-directed DNA polymerase (EC 2.7.7.49)</td>
<td></td>
</tr>
<tr>
<td>3941</td>
<td>83</td>
<td>Potassium uptake protein TrkH</td>
<td></td>
</tr>
<tr>
<td>3946</td>
<td>364</td>
<td>ABC transporter, ATP-binding protein</td>
<td></td>
</tr>
<tr>
<td>3985</td>
<td>69</td>
<td>UbiD family decarboxylase, Lactobacillus brevis type</td>
<td></td>
</tr>
<tr>
<td>3993</td>
<td>52</td>
<td>Replicative helicase RepA</td>
<td></td>
</tr>
<tr>
<td>4000</td>
<td>219</td>
<td>Transposase</td>
<td></td>
</tr>
<tr>
<td>4003</td>
<td>682</td>
<td>Replicative helicase RepA</td>
<td></td>
</tr>
<tr>
<td>4163</td>
<td>233</td>
<td>Homolog of eukaryotic DNA ligase III</td>
<td></td>
</tr>
<tr>
<td>4179</td>
<td>79</td>
<td>Mobile element protein</td>
<td></td>
</tr>
<tr>
<td>4181</td>
<td>38</td>
<td>Mobile element protein</td>
<td></td>
</tr>
<tr>
<td>4205</td>
<td>70</td>
<td>DNA-binding transcriptional regulator, MocR family / aminotransferase domain</td>
<td></td>
</tr>
<tr>
<td>4264</td>
<td>174</td>
<td>Transposase</td>
<td></td>
</tr>
<tr>
<td>4265</td>
<td>75</td>
<td>Mobile element protein</td>
<td></td>
</tr>
<tr>
<td>4317</td>
<td>182</td>
<td>endolysin</td>
<td></td>
</tr>
<tr>
<td>4325</td>
<td>371</td>
<td>Zn peptidase with DNA binding</td>
<td></td>
</tr>
<tr>
<td>4335</td>
<td>92</td>
<td>Cox</td>
<td></td>
</tr>
<tr>
<td>4337</td>
<td>52</td>
<td>C protein</td>
<td></td>
</tr>
<tr>
<td>4338</td>
<td>328</td>
<td>Integrase</td>
<td></td>
</tr>
<tr>
<td>4521</td>
<td>224</td>
<td>putative lipoprotein</td>
<td></td>
</tr>
<tr>
<td>4522</td>
<td>118</td>
<td>putative lipoprotein</td>
<td></td>
</tr>
<tr>
<td>4523</td>
<td>220</td>
<td>putative lipoprotein</td>
<td></td>
</tr>
<tr>
<td>4534</td>
<td>363</td>
<td>4-hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27)</td>
<td></td>
</tr>
<tr>
<td>4543</td>
<td>223</td>
<td>Toxin HigB / Protein kinase domain of HipA</td>
<td></td>
</tr>
<tr>
<td>4544</td>
<td>84</td>
<td>HipB protein, Antitoxin HigA</td>
<td></td>
</tr>
</tbody>
</table>
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.