The phytopathogenic nature of Dickeya aquatica 174/2 and the dynamic early evolution of Dickeya pathogenicity

Alexandre Duprey, Najwa Taïb, Simon Leonard, Tiffany Garin, Jean-Pierre Flandrois, William Nasser, Céline Brochier-Armanet, Sylvie Reverchon

The phytopathogenic nature of *Dickeya aquatica* 174/2 and the dynamic early evolution of *Dickeya* pathogenicity

Alexandre Duprey[a,$,&], Najwa Taib[b,$,#], Simon Leonard[a], Tiffany Garin[b], Jean-Pierre Flandrois[b], William Nasser[a], Céline Brochier-Armanet[b]*, Sylvie Reverchon[a]*

[$]: equally contributing authors

[a] Univ Lyon, Université Claude Bernard Lyon 1, INSA-Lyon, CNRS, UMR5240, Microbiologie, Adaptation et Pathogénie, 10 rue Raphaël Dubois, F-69622, Villeurbanne, France

[b] Univ Lyon, Université Claude Bernard Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Évolutive, 43 bd du 11 novembre 1918, F-69622, Villeurbanne, France

*Corresponding authors

E-mail: Sylvie.reverchon-pescheux@insa-lyon.fr, Celine.brochier-armanet@univ-lyon1.fr

UMR5240 CNRS-UCBL-INSA, 10 rue Raphaël Dubois, F69622 Villeurbanne Cedex, France

Tel: (33) 472 43 26 95; Fax: (33) 472 43 26 86

Running title: *Dickeya aquatica* 174/2 complete genome

Keywords: Genome evolution, systematics, adaptation, virulence, plant cell wall degrading enzymes

Present address: Hub bioinformatique et biostatistique; Unité de Biologie Evolutive de la cellule microbienne, Département de Microbiologie, Institut Pasteur, Paris, France

& Present address: Department of Microbial Pathogenesis, Yale School of Medicine, 295 Congress Avenue, New Haven, CT 06536, USA

Originality-Significance statement

Although the reach of large-scale comparative studies has spread exponentially over the years, the phytopathogenic *Dickeya* group remains overlooked. In this work, we sequence the complete genome of *Dickeya aquatica* type strain, a species isolated from water that was first assumed to be non-phytopathogenic. We show that the proteome of *D. aquatica* contains a wide number of proteins involved in *Dickeya* virulence, including plant cell wall degrading enzymes, suggesting that this species could be in fact pathogenic. Using experimental approaches, we confirm this prediction and uncover the particular affinity of *D. aquatica* for acidic fruits. In-depth phylogenomic analyses reveal that *Dickeya* species display a great degree of genetic plasticity in the pathogenicity determinants, explaining how this bacterial group was able to colonize a wide variety of plants growing in different climates. These observations greatly advance our understanding of how bacteria adapt to new ecological niches.

Summary

Dickeya is a genus of phytopathogenic enterobacterales causing soft rot in a variety of plants (e.g. potato, chicory, maize). Among the species affiliated to this genus, *Dickeya aquatica*, described in 2014, remained particularly mysterious because it had no known host. Furthermore, while *D. aquatica* was proposed to represent a deep-branching species among *Dickeya* genus, its precise phylogenetic position remained elusive.

Here, we report the complete genome sequence of the *D. aquatica* type strain 174/2. We demonstrate the affinity of *D. aquatica* strain 174/2 for acidic fruits such as tomato and cucumber, and show that exposure of this bacterium to acidic pH induces twitching motility. An in-depth phylogenomic analysis of all available *Dickeya* proteomes pinpoints *D. aquatica* as the second deepest branching lineage within this genus and reclassifies two lineages that likely correspond to new genomospecies (gs.): *Dickeya* gs. poaceaephila (*Dickeya* sp NCPPB 569) and *Dickeya* gs. undicola (*Dickeya* sp 2B12), together with a new putative genus, tentatively named *Prodigiosinella*. Finally, from comparative analyses of *Dickeya* proteomes we infer the complex evolutionary history of this genus, paving the way to study the adaptive patterns and processes of *Dickeya* to different

environmental niches and hosts. In particular, we hypothetize that the lack of xylanases and xylose degradation pathways in *D. aquatica* could reflects adaptation to aquatic charophyte hosts which, in contrast to land plants, do not contain xyloglucans.

Introduction

Enterobacterales represent one of the most studied orders of *Gammaproteobacteria*. According to current systematics, *Enterobacterales* are divided into eight families: *Enterobacteriaceae, Yersiniaceae, Thorselliaceae, Hafniaceae, Morganellaceae*, *Budviciaceae, Erwiniaceae*, and *Pectobacteriaceae. Enterobacterales* are widespread, being found in very different environments such as soils, fresh water, ocean, sediments, and many of them are associated with plants and animals, including insects and humans (Brenner and Farmer III, 2005). *Enterobacterales* include also important model organisms such as *Escherichia coli*, human pathogens such as *Salmonella, Shigella,* and *Yersinia* (Dekker and Frank, 2015), and plant pathogens such as *Erwiniaceae (e.g. Erwinia*, *Pantoea, Phaseolibacter)* and *Pectobacteriaceae* (*e.g. Pectobacterium, Dickeya, Brenneria, Lonsdalea)* (Hauben et al., 1998; Samson et al., 2005). These phytopathogens share virulence genes with zoopathogens such as type III secretion systems (T3SS) that inject effector proteins in eukaryotic cells to suppress the host innate immune defence (Buttner, 2016). They also produce specialized plant virulence factors such as pectinases found in pectinolytic bacteria (Hugouvieux-Cotte-Pattat et al., 2014).

Among pectinolytic enterobacterales, *Dickeya* is the causative agent of soft rot in a wide variety of plants including economically important crops (e.g. potato, chicory, maize, rice, tomato, sugar beet, pineapple, banana) and many ornamental plants (Ma et al., 2007). This causes substantial production losses amounting, for instance for potato, to tens of millions of Euros/year in Europe (Toth et al., 2011). The *Dickeya* genus was first described by Samson et al. (2005), who initially distinguished six species: *Dickeya dadantii, Dickeya dieffenbachiae, Dickeya chrysanthemi, Dickeya paradisiaca, Dickeya zeae*, and *Dickeya dianthicola*. Subsequently*, Dickeya dieffenbachiae* has been reclassified as *D. dadantii* subsp. *dieffenbachiae* (Brady et al., 2012). More recently, three additional *Dickeya* species have been described: *D. solani*, a species isolated

from potatoes and hyacinth (Van der Wolf et al., 2014; Slawiak et al., 2009), *Dickeya aquatica* from freshwater rivers (Parkinson et al., 2014), and *Dickeya fangzhongdai* from pear trees displaying symptoms of bleeding canker in China (Tian et al., 2016). Thus, the *Dickeya* genus now comprises eight species with distinctive phenotypic features (Supplementary Table S1). Virulence mechanisms have been extensively studied in *Dickeya dadantii*. During infection, the bacterium enters the host using natural openings and multiplies in intercellular spaces without causing major damage. Then, it suddenly induces the production of aggression factors, such as pectinases, that break down the plant cell wall pectin, causing the macroscopic symptom called soft rot (Reverchon and Nasser, 2013). *Dickeya* populations were initially considered to be restricted to tropical and subtropical plant hosts and areas (Perombelon, 1990). This assumption was called into question following the identification of *D. dianthicola* strains from potato plants in Western Europe (Janse and Ruissen, 1988) and the first isolation of *D. solani* strains from potato plants in France, Finland, Poland, the Netherlands, and Israel (Czajkowski et al., 2011; van der Wolf et al., 2014). Strains of *D. solani* have also been isolated from hyacinth in Netherlands and it was proposed a possible transfer from hyacinth to potato, via contaminated irrigation waters (Sławiak *et al.*, 2009). Finally, it was postulated that the dissemination of the pathogen in Europe has occurred via the international trade of potato seeds (van der Wolf et al., 2014). This highlights the high capacity of adaptation and dissemination of *Dickeya* to new geographic areas and to new hosts.

According to a large phylogenomic analysis of 895 single copy gene families, *D. paradisiaca* was pinpointed as the first diverging species within *Dickeya*, while other species formed two groups, referred hereafter as to clusters I and II (Zhang et al., 2016). Cluster I encompassed *D. zeae* and *D. chrysanthemi*, while in cluster II, *D. solani* and *D. dadantii* formed the sister-lineage of *D. dianthicola* (Zhang et al., 2016). Species corresponding to these two clusters display different behaviours. For example, in temperate climate, *D. chrysanthemi* and *D. zeae* were frequently isolated from water and were rarely associated to potato infections, contrarily to *D. dianthicola* and *D. solani* (Potrykus et al. 2016). It is noteworthy, that in this study, strains M074 and M005, annotated as *D. chrysanthemi* and *D. solani*, respectively, branch in-between *D. dianthicola* and the *D. solani* / *D. dadantii* group within Cluster II, thus far away from the type strains of their

species (Zhang et al., 2016). *D. fangzhongdai* branched at the base of the Cluster II in a tree based on seven housekeeping genes (Tian et al., 2016). Finally, according to a tree based on six housekeeping genes, *D. aquatica* emerge just after the divergence of *D. paradisiaca* but before the divergence of Cluster I and Cluster II (Parkinson et al., 2014). Yet, the associated bootstrap values were weak (Parkinson et al., 2014), meaning that the relative order of emergence of *D. paradisiaca* and *D. aquatica* remained to determine. Among *Dickeya*, *D. aquatica* is remarkable because all the three known strains 174/2^{T}, 181/2, and Dw0440 have been isolated from waterways and, in contrast to other species, have no known vegetal host (Supplementary Table S1) (Parkinson et al., 2014; Ma et al., 2007). These features and its early-branching position within *Dickeya* make *D. aquatica* very interesting to study the origin and the early steps of the diversification of *Dickeya*, and in particular, the emergence of their virulence factors and capacity to infect plants.

These important questions were the focus of this study. For this purpose, we sequence the complete genome of the *D. aquatica* 174/2 type strain. Based on the phylogenetic analysis of 1,341 single copy core protein families, we show that *D. aquatica* represents the second deepest branching lineage within *Dickeya*, reclassify two lineages that likely correspond to new genomospecies (gs.): *Dickeya* gs. poaceaephila and *Dickeya* gs. undicola, and identify a new lineage, tentatively called *Prodigiosinella*, that likely represents the closest relative of *Dickeya*. We also highlight the presence of different virulence factors, suggesting that *D. aquatica* 174/2 could be pathogenic. Stimulated by these findings, we explore the pathogenic potential of *D. aquatica* 174/2. Surprisingly, we identify the acidic fruits tomato (pH 4.8) and cucumber (pH 5.1) as potential hosts for *D. aquatica*, reflecting the potential of this aquatic species to flexibly adapt to a new ecological niche provided by acidic fruits with a high water content. Accordingly, we show that this strain displays a specific induction of twitching motility under acidic conditions. Finally, using phylogenomic approaches, we trace back the origin and evolution of key factors associated with virulence and host specificity in *Dickeya*, including *D. aquatica*. From this study we infer the evolution of gene repertoires along the diversification of *Dickeya* and highlight a remarkable general tendency toward proteome reduction in all *Dickeya* species.

Results and discussion

General genomic features of D. aquatica *174/2 type strain*

The *D. aquatica* 174/2^T genome consists in a single circular chromosome of 4,501,560 base pairs in size, with a GC content of 54.6%. The sequence was submitted to European Nucleotide Archive (accession GCA_900095885). This genome contains a total of 4,202 genes including 4,080 protein-coding DNA sequences (CDS), 22 ribosomal RNA-coding genes organized into seven operons, 76 tRNA-coding genes, and 23 non-coding RNA genes identified by sequence similarity with known functional RNAs entries in the RFAM database (Daub et al., 2015) (Figure 1). These features are typical of *Dickeya* species (Supplementary Table S2). The origin of chromosome replication (*oriC* position 1-8 bp) was predicted between *mnmG/gidA* and *mioC* as observed in *Escherichia coli* (Wolanski et al., 2014) and other enterobacterales, including *Dickeya* (Glasner et al., 2011; Zhou et al., 2015; Khayi et al., 2016). The terminus of replication is predicted between positions 2,220,802 and 2,220,829 bp with the *D. aquatica* 174/2^T Dif site (GGT**T**CGCATAATGTATATTATGTTAAAT) differing from the *E. coli* K-12 Dif site by only one nucleotide substitution (GGT**G**CGCATAATGTATATTATGTTAAAT). The distance from *oriC* to the predicted terminus was nearly equal for the two halves of the chromosome. It corresponds to a region near an inflection point in a slight strand-specific nucleotide compositional bias (Figure 1). The protein coding density is 85% of the genome with a slight preference for the leading strand utilization (57%). This density of predicted ORFs (slightly less than one per kilobase) is typical for *Enterobacterales*, including *Dickeya* species (Supplementary Table S2) (Glasner et al., 2011; Zhou et al., 2015; Khayi et al., 2016). Thus, despite its different ecological niche, *D. aquatica* shows genomic organisation very similar to other species of the *Dickeya* genus.

***Deciphering the early evolution of* Dickeya**

The massive release of *Dickeya* genomic sequences (7 complete genomes and 41 draft genomes) in public databases (Supplementary Table S3), in addition to the complete genome of *D. aquatica* 174/2^T reported in this study, provided an interesting resource to explore the evolutionary history of this genus. We investigated the phylogeny of *Dickeya* using 51 ribosomal proteins (rprots), as

these were shown to be well suited to study the systematics of prokaryotes, especially *Proteobacteria* (Yutin et al., 2012; Ramulu et al., 2014), on the one hand, and 1,341 core protein families (core-pf), on the other hand. The maximum likelihood rprots tree, rooted with *Pectobacterium* (*Pectobacteriaceae*) and *Serratia* (*Yersiniaceae*), recovered the monophyly of *Dickeya* (Bootstrap Value (BV) = 100%, Supplementary Figure S1). Surprisingly, *Serratia* strains, used as outgroup to root the phylogeny of *Dickeya* together with *Pectobacterium*, did not form a monophyletic group, due to the robust clustering of *Serratia* sp. ATCC 39006 with *Dickeya* (BV = 99%). A similar grouping was also observed in trees based on the RNA components of the small and large subunits of the ribosome (SSU and LSU rRNA, respectively) (Supplementary Figure S2), suggesting that strain ATCC 39006 represents the sister-lineage of *Dickeya*, and was wrongly affiliated to the *Serratia*, based on its capacity to synthesize prodigiosin, a red pigment secondary metabolite with antimicrobial, anticancer, and immunosuppressant properties, characteristic of *Serratia* species (Thomson et al., 2000). The sister-relationship of ATCC 39006 and *Dickeya* is consistent with the close evolutionary link between the pectinolytic gene clusters (*pel*, *paeY*, *pemA*) shared by these taxa (Duprey et al., 2016b), and with the presence, in *Serratia* sp. ATCC 39006, of an homologue of the Vfm quorum sensing system previously reported as specific to *Dickeya* species (Supplementary Table S4 and Figure S3) (Nasser et al., 2013). The phylogenetic analysis of the genes involved in the biosynthesis of the prodigiosin showed that strain ATCC 39006 genes emerge in a cluster gathering sequences from unrelated taxa: *Serratia* (*Yersiniaceae*), *Hahella* (*Hahellaceae*), and *Streptomyces* (*Actinobacteria*), suggesting that the prodigiosin gene cluster spread among these lineages (including ATCC 39006 strain) through horizontal gene transfer (HGT) (Supplementary Figure S3). Interestingly, strain ATCC 39006, like the early-diverging *D. paradisiaca* species, was deprived of the *indABC* gene cluster responsible for the characteristic production of blue-pigmented indigoidine in *Dickeya* spp. (Supplementary Table S4) (Reverchon et al., 2002; Lee and Yu, 2006). Phylogenetic analyses showed that *Dickeya* IndA, IndB, and IndC proteins are closed to sequences from unrelated bacteria (Supplementary Figure S3), suggesting that indigoidine genes underwent HGT and have been secondarily acquired in *Dickeya* after the divergence with *D. paradisiaca*. The evolutionary link between strain ATCC 39006 and *Dickeya* is

also consistent with comparison of their proteomes. In fact, strain ATCC 39006 share in average more proteins families with *Dickeya* (2,554/4,221 – 60.5%), compared to *Pectobacterium* (2,387/4,221 - 56.5%) or other *Serratia* species (2,162/4,221 - 51.2%) (Supplementary Table S5). Yet, the large evolutionary distance between ATCC 39006 strain and *Dickeya* (Supplementary Figure S1), its lower GC content (49.2% vs 52.6 - 56.9% in *Dickeya*) (Supplementary Table S3), and the lower number of protein families shared between ATCC 39006 and *Dickeya* (60.5%) compared to between *Dickeya* species (64.7% - 89.6%, Supplementary Table S5), suggested that strain ATCC 39006 does not belong to *Dickeya* and represents rather a close but distinct lineage, possibly a new genus, that we propose to call *Prodigiosinella*, with strain ATCC 39006 being reclassified as *Prodigiosinella confusarubida*. By using leBIBI^{QBPP} phylogenetic SSU rRNA-based positioning tool (Flandrois et al., 2015), we detected two *Erwinia* sp. Strains: MK01 (SSU accession number at the NCBI: AY690711.1) and MK09 (SSU accession number at the NCBI: AY690717.1), isolated from the rhizosphere of *Phragmites communis*, that were very closely related to ATCC 39006 and could belong to *P. confusarubida*.

According to these results and to decrease computation time, the core-pf phylogeny was inferred using *P. confusarubida* ATCC 39006 as outgroup. The maximum likelihood core-pf and rprots trees displayed similar topologies (Figure 2 and Supplementary Figure S1), excepted regarding *D. chrysanthemi*, *D. sp.* NCPPB 569, and *D. solani* (see below). Worth to note, the rprots tree was overall less resolved than the core-pf tree (i.e. the former displayed lower BV than the latter). Both trees supported the monophyly of *D. paradisiaca*, *D. aquatica*, *D. zeae, D. dianthicola, D. dadantii*, and *D. solani* (BV ≥ 84%). The monophyly of *D. chrysanthemi* was recovered with core-pf (BV = 100%) but not with rprots. In fact, while the relationships among *D. chrysanthemi* strains were strongly supported in the core-pf tree, they were mostly unresolved (i.e. associated to weak BV) with rprots, meaning that these markers did not contain enough phylogenetic signal to resolved the *D. chrysanthemi* phylogeny. A composite group composed of strains annotated as *D. sp.*, *D. solani*, and *D. chrysanthemi*: M074, M005, B16, MK7, S1, NCPPB 3274, and ND14b (formerly misidentified as *Cedecea neteri* and transferred recently to *D. solani*) was present in both trees (BV ≥ 95%). By using leBIBI^{QBPP} tool (Flandrois et al., 2015), we identified strains B16, MK7, and S1 as

D. fangzhongdai, suggesting that the whole clade could correspond to *D. fangzhongdai* (Alič et al., 2018).

Cluster I (*D. zeae* and *D. chrysanthemi*) and cluster II (*D. dianthicola*, *D. solani*, *D. dadantii*, *D. fangzhongdai* and *D. undicola*) were monophyletic (core-pf: both BV ≥ 100% and rprots: both BV ≥ 80%, respectively). Within cluster II, *D. fangzhongdai* diverged first (BV = 100%), while *D. solani* represented either the sister-group of *D. dadantii* (core-pf: BV = 100%) or *D. dianthicola* (rprots: BV = 98%). The conflicting position of *D. solani* was supported by high BV in both trees, indicating a real inconsistence between the two trees. To go further, approximately unbiased tests (Shimodaira 2002) performed on individual core-pf and rprots by using the core-pf and the rprots maximum likelihood topologies (Supplementary Table S6). These tests showed that 1,063 out of 1,341 (79.3%) core-pf protein families reject the rprots topology, while 853 out of 1,341 (63.6%) core-pf protein families do not reject the core-pf topology. In contrast, 31.4% of the rprots do not reject the core-pf topology and 29.4% do not reject the rprots topology. Thus, a large majority of core-pf protein families favour the core-pf topology, while an equal number of rprots supports either one or the other topology. Accordingly, the rprots topology and in particular the grouping of *D. solani* with *D. dianthicola* could be questioned.

Regarding the first divergences within *Dickeya*, both trees pinpointed *D. paradisiaca* and *D. aquatica*, as early diverging lineages (core-pf: BV = 100%, rprots: BV = 86%), with *D. paradisiaca* emerging first (core-pf: BV = 100%, rprots: BV = 65%). Finally, the core-pf placed robustly *D. sp.* NCPPB 569, a strain isolated from sugarcane in Australia (Supplementary Table S3), at the base of Cluster I (BV = 100%), while its position was unresolved in the rprots tree (i.e. unsupported grouping with *D. aquatica*, BV = 46%). Irrespective of its position, the evolutionary distances between strain NCPPB 569 and the eight *Dickeya* species were roughly similar, suggesting NCPPB 569 could correspond to an independent lineage within *Dickeya*. This observation was in agreement with a previous work based on the phylogenetic analysis of *recA* (Parkinson et *al*. 2009), and led us to propose that strain NCPPB 569 could represent a distinct genomospecies, we proposed to name *Dickeya* gs. poaceaephila. This hypothesis was strengthened by the fact that

strain NCPPB 569 could be distinguished from other species according to shared protein families (Figure 2 and Supplementary Figure S4). Among cluster II, *Dickeya* sp. 2B12 deserved attention. In fact, this strain branched as a distant sister-lineage of *D. fangzhongdai* (core-pf: BV = 100% and rprots: BV = 95%). This suggested that this strain could also represent a new genomospecies that we have tentatively called *Dickeya* gs. undicola.

Altogether, the analysis of more than one thousand core protein families and ribosomal proteins allows to resolve most of the speciation events within *Dickeya* and provides a solid framework to investigate the origin and the evolution of important *Dickeya* biological features. As expected, both trees were largely consistent, even if the rprots tree was globally less resolved, in particular regarding *D. chrysanthemi* and *Dickeya* gs. poaceaephila. The main inconsistency concerned the relative order of divergence of *D. dadantii*, *D. dianthicola*, and *D. solani* within cluster II. Approximately unbiased tests suggested that rprots could contain a mix phylogenetic signal, and thus that the rprots tree could be less reliable than the core-pf tree. Determining the origin of the conflicting signal in rprots would require more investigations that are beyond the scope of the study. Because *D. solani* is a relatively late diverging lineage within *Dickeya*, we anticipated that this will not impact significantly inferences on the ancient *Dickeya* evolution. Nevertheless, in all analyses (e.g. evolution of virulence related systems, inference of ancestral gene repertoires, see below), we used the species tree based on core-pf, but by considering both possible placements for *D. solani*: either as the sister-group of *D. dadantii* or *D. dianthicola*. Regarding *D. aquatica*, the core-pf tree suggested that this species represent the second diverging lineage with *Dickeya*, while its position was unresolved according to rprots. Our analyses also reclassified strain ATCC 39006 as *Prodigiosinella confusarubida,* which could represent the closest relative of *Dickeya* genus. Better characterization of the biology and diversity of *Prodigiosinella* will provide important clues about their evolutionary links with *Dickeya* and the emergence of *Dickeya* as a genus.

Virulence-associated phenotypes of D. aquatica

While most *Dickeya* species have known vegetal hosts, all the identified strains of *D. aquatica* (174/2^T, 181/2 and Dw0440) have been isolated from waterways (Supplementary Table S1)

(Parkinson et al., 2014). To our knowledge, their pathogenicity has never been demonstrated. The *in silico* survey of the *D. aquatica* 174/2^T genome for transcriptional regulator binding sites revealed the presence of conserved key regulators of virulence (Supplementary Table S7). More precisely, all the regulators (KdgR, PecS, PecT, CRP, Fis, H-NS, GacA-GacS, RsmA-RsmB, MfbR) controlling virulence gene expression in *D. dadantii* (Reverchon et al., 2016) are present in *D. aquatica* 174/2^T. The global regulators FNR and CRP displayed the highest number (approximately 80) of targets, followed by Fur, CpxR, ArcA, Lrp and KdgR each with more than 30 targets. The activator CRP and the repressor KdgR have a key role in *Dickeya* virulence (Duprey et al., 2016a), as they tighly control the pectin degradation pathway and are involved in coupling central metabolism to pectinase gene expression (Nasser et al., 1997), while Fur acts as a repressor of pectinase genes (Franza et al., 2002). In addition, Fur represses genes involved in the metabolism of iron, a metal that plays an important role as a virulence regulatory signal in *Dickeya* (Franza et al., 2002).

The presence of regulators controlling the virulence in *D. aquatica* 174/2^T was puzzling and prompted us to investigate its pathogenicity. For this, we compared the pathogenic potential of *D. aquatica* 174/2^T and *D. dadantii* 3937 on various infection models such as chicory leaves, potato tubers, cucumbers and tomato fruits (Figure 3). *D. aquatica* showed little efficiency in rotting potato and chicory. Indeed, only a small rotten area was observed near the entry point, opposed to *D. dadantii,* which spread inside the tuber or the leaf (Figure 3). Surprisingly, *D. aquatica* appeared particularly efficient on tomatoes and cucumbers, showing a generalised infection and fruit opening after 42 hours. By contrast, *D. dadantii* was much less efficient in infecting tomatoes and cucumbers. Importantly, tomatoes and cucumbers are acidic fruits (pH 4.8 and pH 5.1, respectively), whereas potato tubers and chicory leaves display higher pH (around 6.0 and 6.5, respectively), suggesting that *D. aquatica* could be more efficient on acidic fruits. Yet, neither *D. aquatica* nor *D. dadantii* are able to induce pathogenic symptoms on very acidic fruits such as pineapple (pH 4) or kiwi (pH 3) (Figure 3B). Previous studies have shown that some unrelated *Dickeya* species (i.e. *D. chrysanthemi*, *D. dianthicola*) are also capable of infecting tomatoes (Supplementary Table S1), supporting the hypothesis that recurrent adaptations to similar hosts

occurred independently during the evolution of *Dickeya* genus.

Stress resistance phenotypes of D. aquatica

Osmotic stress

Depending on the plant host, *Dickeya* species encounter various stresses during the infectious process (Reverchon et al., 2016). We therefore assessed stress resistance of *D. aquatica* 174/2^T (Figure 4). This strain proved to be very sensitive to osmotic stress, as it displayed a 50% growth rate reduction on 0.3 M NaCl while *D. dadantii* was only slightly affected (20% growth rate reduction). This effect was even more pronounced on 0.5 M NaCl with a growth rate reduction of 90 % for *D. aquatica* and 43% for *D. dadantii* (Figure 4A). This sensitivity to osmotic stress could be linked to the lack of two osmoprotectant biosynthetic pathways in *D. aquatica*: glycine betaine biosynthesis (*betA-betB* gene cluster) and phosphoglycosyl glycerate biosynthesis (*pggS-pggP* gene cluster) (Jiang et al., 2016) (Supplementary Table S4). The *pggSP* gene cluster was present in *P. confusarubida* and all *Dickeya* species except *D. aquatica*. The phylogeny of the two genes was consistent with the phylogeny of species (Supplementary Figure S3), indicating that their absence in *D. aquatica* results from a specific and secondary loss in this taxon. In contrast, the *betABI* gene cluster was absent in *D. paradisiaca, D. aquatica*, *D. dianthicola* and Cluster I species (Supplementary Table S4). Their phylogenies suggested that the *betABI* gene cluster could have been present in the ancestor of *P. confusarubida* and *Dickeya*, and secondarily and independently lost in the *Dickeya* lineages mentioned above (Supplementary Figure S3). Finally, the *ousA* gene encoding the major osmoprotectant uptake system in *D. dadantii* is missing in *D. aquatica*, as well as in most other *Dickeya* species, suggesting either a recent acquisition by a few species or multiple losses of an ancestral system (Supplementary Table S4). Interestingly, *D. solani* and *D. dianthicola*, both lacking the *ousA* gene, are also more sensitive to osmotic stress than *D. dadantii* (Figure 4B).

Oxidative stress

D. aquatica 174/2^T was also very sensitive to oxidative stress, displaying longer lag time than *D. dadantii* in presence of 75 µM H_2O_2 (Figure 4C). This is consistent with the lack of the periplasmic

superoxide dismutase SodC, whose gene was likely acquired in *Dickeya* after the divergence of *D. paradisiaca* and *D. aquatica* (Supplementary Figure S3), and the lack of the Fe-S cluster assembly SUF system (Supplementary Table S4). Iron-sulfur (Fe-S) clusters are fundamental to numerous biological processes in most organisms, but these protein cofactors can be prone to damage by various oxidants (e.g., O_2, reactive oxygen species, and reactive nitrogen species) (Roche et al., 2013). In addition, the release of free iron in the bacterial cytoplasm amplifies the oxidative stress through the Fenton reaction, producing the highly toxic and reactive hydroxyl radical $OH^{\bullet}$. Most gammaproteobacteria have two Fe-S cluster biogenesis systems SUF and ISC (Roche et al., 2013). For example, *Escherichia coli* cells switch from the ISC to the SUF system under oxidative stress as OxyR, a sensor of oxidative stress, acts as an activator of *suf* operon expression (Lee et al., 2004). Both the SUF and ISC systems were shown to be required for *D. dadantii* growth in the plant host environment that is continually changing in terms of iron availability and redox conditions (Nachin et al., 2001; Expert et al., 2008; Rincon-Enriquez et al., 2008). While both the ISC and SUF systems were likely present in the ancestor of *Dickeya* and *P. confusarubida*, phylogeny and taxonomic distribution suggested that SUF genes were secondarily lost in *D. paradisiaca*, *D. aquatica*, *D*. gs. poaceaephila, and *D. zeae* (Supplementary Figure S3). The absence of both the SUF system and SodC in *D. aquatica* could be an important factor limiting its host range (Supplementary Figure S3). Finally, the presence of the DNA-binding protein Dps, which has a protective function against a wide range of stresses, in *P. confusarubida* and all *Dickeya* excepted *D. paradisiaca* suggested a specific loss in this latter lineage (Supplementary Figure S3).

Regarding other oxidative stress sources, the NorWVR system responsible for nitrite oxide detoxification was absent in the first diverging *Dickeya* species (Supplementary Table S4). The phylogenies of the corresponding proteins suggested secondary losses in *P. confusarubida*, *D. paradisiaca*, and *D. aquatica*, as these proteins were present in *Pectobacterium*, a closely related genus of *Dickeya* (Supplementary Figure S3). Finally, the ascorbate degradation pathway encoded by the *ula* genes was present in the *D. paradisiaca, D. aquatica, D. zeae*, and *D. chrysanthemi* species, suggesting that the pathway was also present in the ancestor of *Dickeya* and secondarily lost in *D.* gs. poaceaephila and in the ancestor of Cluster II (Supplementary Figure S3). Ascorbic

acid, the major and probably the only antioxidant buffer in the plant apoplast, becomes oxidised during pathogen attack (Pignocchi and Foyer, 2003). Modification of the apoplastic redox state modulates receptor activity and signal transduction to regulate plant defence and growth (Pignocchi and Foyer, 2003). The capacity of some *Dickeya* species to catabolize ascorbate could be a strategy to weaken the plant defence.

Acidic stress

Regarding pH sensitivity, both *D. aquatica* 174/2^T and *D. dadantii* showed optimum growth at pH 7.0. *D. aquatica* 174/2^T displayed no significant growth rate reduction down to pH 5.3 while *D. dadantii* was slightly affected (20% growth rate reduction at pH 5.3 compared to optimum pH 7.0) (Figure 4D). Below pH 4.9, the growth of *D. aquatica* abruptly diminished compared to that of *D. dadantii* indicating that *D. aquatica* growth was much more impacted by low pH (Figure 4D). Bacterial response to acid stress involves evasion of cell damage and adaptation of the enzymatic profile by reducing reactions producing protons and promoting those consuming protons (Bearson et al., 1997). Concerning the protection of proteins against pH damage, *Dickeya* strains, including *D. aquatica* strain 174/2^T, encode the lysine-rich protein Asr acid-inducible periplasmic chaperone, known to protect the proteins by sequestering protons due to its high basic amino acids composition (Seputiene et al., 2003). In *Dickeya*, the corresponding gene is located downstream the *rstAB* gene cluster, known to control the *asr* gene expression (Ogasawara et al., 2007). Yet, due to their atypical amino acid composition these proteins were often wrongly annotated as histone proteins (e.g. in *D. dadantii* NCPPB 898, *D. dianthicola* NCPPB 3534, *D. dianthicola* NCPPB 453, *D. dianthicola* GBBC 2039, and *D. chrysanthemi* ATCC 11663). In terms of metabolic adaptation, a major strategy for bacterial pH homeostasis consists in using decarboxylases to remove cytoplasmic protons (Krulwich et al., 2011). The glutamate-, arginine-, and lysine-inducible decarboxylases classically found in enteric bacteria (Foster, 2004) are absent in *Dickeya* species. However they contain some organic acid decarboxylases such as oxalate and malonate decarboxylases (Supplementary Table S4). Some *Dickeya* species, including *D. dadantii,* have two oxalate decarboxylation pathways: the frc-oxc pathway, being involved in the acid tolerance response in *E. coli* (Fontenot et al., 2013) and the OxdD decarboxylase pathway. Most *Dickeya*

388 species harbour at least one of these oxalate decarboxylation pathways, except *D. aquatica*, *D.* gs.
389 poaceaephila, *D. chrysanthemi*, and a few *D. zeae* strains, which were deprived of both pathways
390 (Supplementary Table S4 and Supplementary Figure S3). The lack of the two oxalate-
391 decarboxylation pathways in *D. aquatica* could contribute to its sensitivity to acidic pH. Finally, in
392 contrast with *P. confusarubida* and most *Dickeya* species, *D. aquatica* and a few *D. zeae* strains
393 were also devoid of malonate decarboxylation pathway (mdcABCDEFGHR) (Maderbocus *et al.*,
394 2017), which consumes protons (Supplementary Table S4), suggesting secondary gene losses in
395 these strains, a hypothesis that was confirmed by phylogenetic analyses (Supplementary Figure
396 S3).

397 **Twitching motility**

398 Interestingly, colony shape changes were observed when *D. aquatica* was grown at low pH in
399 presence of malic acid (Figure 5). More precisely, colonies became wider (2-3 folds increase in
400 diameter compared to unstressed *D. aquatica*) with a morphology characteristic of twitching motility
401 (Henrichsen, 1972). Correspondingly, we detected genes encoding a complete type IV pilus
402 assembly responsible for twitching motility in *D. aquatica* genome and named them *pil* genes
403 according the *Pseudomonas aeruginosa* nomenclature (Maier and Wong, 2015). Among the
404 collection of twenty-six *Dickeya* strains representing different species, only eight showed twitching
405 at acidic pH (Figure 5). Interestingly, this phenotype appeared strain-dependent rather than
406 species-dependent, as *D. solani*, strain RNS07.7.3B showed twitching motility at acidic pH, while
407 *D. solani* strain PP09019 did not (Figure 5). As the *pil* genes, inherited by *Dickeya* from the
408 common ancestor shared with *P. confusarubida,* were conserved in all species (Supplementary
409 Table S4 and Figure S3), twitching was probably linked to strain-specific induction of these genes
410 under acidic pH. It is noteworthy that twitching motility is strongly influenced by changes in the
411 environment (Henrichsen, 1975). In plant pathogens, the contributions of type IV pilus to virulence
412 have been investigated mainly in vascular pathogens, such as *Ralstonia* and *Xylella*, where they
413 were proposed to contribute to bacterial colonization and spread in the xylem through cell
414 attachment, biofilm formation, and twitching motility (Burdman et al., 2011). However, type IV
415 piliation was also shown to be important for initial adhesion and colonization of leaves in a few non-

vascular bacteria such as *Xanthomonas oryzae pv. oryzicola, Pseudomonas syringae pv. tabaci* and *Pseudomonas syringae pv. syringae* (Burdman et al., 2011). The importance of type IV pilus in *Dickeya* pathogenicity is therefore an interesting question for the future studies.

Antimicrobial peptides

In response to infection, plants produce antimicrobial peptides (AMPs) to limit pathogen propagation. To overcome AMPs, bacteria remodel their envelope, and more precisely, they modify their LPS to decrease interaction with positively charged AMPs. Among the genes involved in LPS modification, the operons *arnABCDEFT* and *dltXABCD* turned out to be ancestral and conserved in all *Dickeya* species (Supplementary Figure S3), while other genes (i.e. *eptAB, pagP, lpxO* and *lpxT*) displayed different taxonomic distributions and evolutionary histories (Supplementary Table S4 and Figure S3). For example, *D. aquatica* 174/2^T is lacking *pagP*, *lpxO* and *lpxT* genes possibly contributing to a greater sensitivity to AMPs.

To conclude, our data indicate that, *D. aquatica* 174/2^T has an unanticipated phytopathogenic capacity similar to that of other *Dickeya* species. Particular stress resistance profiles and induction of twitching at acidic pH may contribute to its restricted host range. Based on this observation, we decided to compare the *D. aquatica* 174/2^T and other *Dickeya* species proteomes with a special focus on the virulence determinants including plant cell wall degrading enzymes, secretion systems, iron metabolism, plant adhesion elements and secondary metabolism.

Distribution of plant cell wall degrading enzymes in Dickeya

The plant cell wall is a complex and dynamic meshwork of polymers (cellulose, hemicellulose, pectin, structural glycoproteins) (Pauly and Keegstra, 2016). Among these polymers, pectin is the most complex and includes both linear regions composed of polygalacturonan and ramified regions (RGI and RGII, respectively). RGI contains a rhamnogalacturonan backbone and various lateral chains such as galactan, arabinan and galacturonan (Caffall and Mohnen, 2009). RGII contains a short galacturonan backbone, carrying four side chains, with a diversity of rare monosaccharides (O'Neill et al., 2004). The carboxylic groups of D-galacturonate residues are methyl-esterified to various degrees (up to 80%) and these residues are, to a lesser extent, acetylated at the C2 and/or

C3 positions. Feruloyl esters are a type of modification commonly found in arabinan and galactan chains of ramified regions (Ishii, 1997). The virulence of *Dickeya* is correlated with their ability to synthesize and secrete plant cell wall degrading enzymes, including a full set of pectinases (Hugouvieux-Cotte-Pattat et al., 2014), xylanases and xylosidases (Keen et al., 1996) proteases PrtA, PrtB, PrtC, PrtG (Wandersman et al., 1987), and the cellulase Cel5Z (Py et al., 1991). The presence of these virulence factors varies depending on the species (Matsumoto et al., 2003; Duprey et al., 2016b). To obtain a comprehensive view of this phenomenon, we explored the "Plant cell wall degradosome" of *Dickeya* species including *D. aquatica* (**Figure** 6).

The pectinasome

The *Dickeya* pectinasome includes multiple pectate lyases (PelA, PelB, PelC, PelD, PelE, PelI, PelL, PelN, PelW, PelX, PelZ, Pel10), pectin lyases (PnlG, PnlH), polygalacturonases (PehK, PehN, PehV, PehW, PehX), pectin methyl esterases (PemA, PemB), pectin acetyl esterases (PaeX, PaeY), feruloyl esterases (FaeD, FaeT), rhamnogalacturonate lyases (RhiE, RhiF), and one periplasmic endogalactanase (GanA) (Figure 6) (Hugouvieux-Cotte-Pattat et al., 2014). *D.* gs. poaceaephila NCPPB 569 was the genospecies with the poorest pectinase content. According to their taxonomic distribution and phylogeny, most of these proteins could be inferred in the ancestor of *Dickeya* and in its sister lineage *P. confusarubida*. Regarding the *pelAED* cluster, while *P. confusarubida* has a single gene, the ancestor of *Dickeya* had two copies, *pelA* and an undifferentiated *pelDE* pectate lyase-coding gene (Duprey et al., 2016b), suggesting that a duplication event occurred in the stem of *Dickeya*. This undifferentiated *pelDE* has undergone a duplication event after the emergence of *D. paradisiaca*, giving rise to *pelD* and *pelE* (Supplementary Figure S5) (Duprey et al., 2016b). *D. aquatica* has then lost *pelE*, while *D. poaceaephila* has lost both *pelA* and *pelE*. The *pelA* gene was also lost in *D. dianthicola*, while *pelE* was lost in some *D. chrysanthemi* strains (Figure 6, Supplementary Figure S5). The *pelBC* cluster was also likely present in the *Dickeya* ancestor, and conserved in most *Dickeya* species. However, *pelB* was lost in *D.* gs. poaceaephila, and *D. dadantii subsp. dieffenbachiae* NCPPB 2976 (Supplementary Figure S3). The pectate lyase PelI was present in *Pectobacterium* and in all

Dickeya, except *D. paradisiaca*, and in *P. confusarubida*, suggesting that it could be ancestral in *Dickeya*. Similarly, the phylogeny of pectin methyl esterase PemB indicated that the protein was present in *P. confusarubida* and *Pectobacterium*, suggesting an ancestral presence in *Dickeya*. Accordingly, its absence in *D. paradisiaca*, *D. aquatica*, *D.* gs. poaceaephila, and *D. zeae* strains likely reflected secondary losses. The phylogeny of this protein also suggested that the *pemB* gene found in *D. chrysanthemi* was acquired by HGT from Cluster II species (Supplementary Figure S3). Regarding the polygalacturonase PehN, a similar scenario could be inferred except that the corresponding gene was also lost in *P. confusarubida*, and that a few *D. zeae* strains seemed to have reacquired a PehN from Cluster II species by HGT (Supplementary Figure S3). Regarding the *pehVWX* cluster, the three genes were derived from a single *pehX* gene that was present in the ancestor of *P. confusarubida* and *Dickeya*. A gene duplication event occurred in the ancestor of *D. dianthicola*, *D. solani* and *D. dadantii* leading to PehV. A second event led to the divergence of PehW in the ancestor of *D. solani* and *D. dadantii* (Supplementary Figure S5). The polygalacturonase PehK was likely present in the ancestor of all *Dickeya* and secondarily lost in *D.* gs. poaceaephila*, D. solani*, and *D. dianthicola* (Supplementary Figure S3). The pectin lyase PnlG was likely present in the ancestor shared by *Dickeya* and *P. confusarubida*, and secondarily lost in *D. aquatica*, *D.* gs. poaceaephila, *D. chrysanthemi*, *D. dianthicola* and in some *D. zeae* strains (Supplementary Figure S3). Finally, the rare pectate lyase Pel10 and pectin lyase PnlH displayed patchy taxonomic distributions (Figure 6). Their phylogenies suggested that they spread through HGT in *Dickeya* (Supplementary Figure S3).

The saturated and unsaturated digalacturonates resulting from pectin degradation by pectinases are converted into monogalacturonate and 5-keto-4-deoxyuronate by the oligogalacturonate lyase Ogl, which is present in all *Dickeya* species (Figure 6). The phylogenies of the *gan* gene cluster responsible for degradation of galactan chains in pectin-ramified regions and the rhamnogalacturonate lyase RhiE involved in degradation of RGI pectin-ramified regions indicated they were likely present in the ancestor of *P. confusarubida* and *Dickeya*, and secondarily lost in the basal *D. paradisiaca* species, in *D.* gs. poaceaephila, and in a few other strains (Figure 6). The ferulate esterases FaeT and FaeD were absent in *P. confusarubida* and were acquired in the basal

D. paradisiaca and *D. aquatica* species. The FaeT enzyme was then secondarily lost in *D.* gs. poaceaephila, and *D. dadantii subsp dieffenbachiae* (Supplementary Figure S3). Altogether, the variability observed in the pectinasome among *Dickeya* species and even among different strains likely reflects the dynamic evolution of the corresponding gene families involving gene acquisitions and losses. More compellingly, this variability indicates that the pectinasome cannot be used to distinguish among various species.

Cellulases and Xylanases

Two cellulases Cel5Z and CelY were likely ancestral in *Dickeya* species except in *D.* gs. poaceaephila, which has lost Cel5Z. While Cel5Z was involved in cellulose degradation, CelY belonged to the *bcs* gene cluster responsible for cellulose fiber formation (Prigent-Combaret et al., 2012). Xylanases (XynA, XynB) and xylosidases (XynC, XynD) cleave xylan and xyloglucan, which belong to the hemicelluloses. The ß 1,4 xylan is mainly present in plant cell wall of monocots (Pena et al., 2016) and is further decorated, often by acetyl, arabinosyl, and glucuronosyl side-chain substitutions. Notably, the xylan substitution patterns depend on the plant species and are distinct in gymnosperms and angiosperms (Busse-Wicher et al., 2016). Xyloglucan is a β-1,4 glucan that can be substituted with a diverse array of glycosyl and nonglycosyl residues. The type and order of xyloglycan substituents depend on the plant species (Pauly and Keegstra, 2016). Xyloglucan polymers fall into one of two general types. In one type, three out of four backbone glucosyl residues are xylosylated, leading to an XXXG-type xyloglycan, which is predominant in most dicots. Another type of xyloglycan exhibits reduced xylosylation in that only two out of the four or more backbone glucosyl residues are xylosylated, resulting in the XXGGn-type xyloglucan, which is present in early land plants such as liverworts, mosses, lycophytes, and ferns of the order Polypodiales. This type of xyloglucan seems to be absent from the gymnosperms and angiosperms, with the exception of the grasses (Poales) and plants from the order Solanales, such as potato, tobacco and tomato (Pauly and Keegstra, 2016). Thus, it is tempting to hypothesize that the xylanase and xylosidase content of *Dickeya* species could be correlated with their plant host. Interestingly, all *Dickeya* species as well as *P. confusarubida* contained at least one XynA, XynB,

XynC, or XynD coding gene (Figure 6). The phylogeny of XynA indicated that the corresponding gene was likely present in the ancestor of *P. confusarubida* and *Dickeya*, and secondarily lost in *D. dianthicola*, *D*. gs. undicola, *D. chrysanthemi*, *D*. gs. poaceaephila, and *D. aquatica* (Supplementary Figure S3). The XynB phylogeny suggested a secondary acquisition in *Dickeya* after the divergence of *P. confusarubida*, *D. paradisiaca*, and *D. aquatica*, followed by losses in *D. dianthicola*, *D. chrysanthemi*, and *D*. gs. undicola (Supplementary Figure S3). While the mode of action of the xylanase XynB was not studied, XynA is a glucuronoxylanase (CAZy family GH30) hydrolyzing the xylan backbone adjacent to each glucuronosyl side-chain (Urbanikova et al., 2011). *D. aquatica, D. chrysanthemi, D*. gs. undicola*,* and *D. dianthicola* were deprived of both XynA and XynB (Figure 6) suggesting that these species would preferentially infect dicots as xylan is mainly present in plant cell wall of monocots. However, this is probably a tendency since *D. chrysanthemi* strain Ech1591 was isolated from maize, which is a monocot. The phylogeny of Xylosidase XynD suggested that the corresponding gene has been acquired by HGT in *D. aquatica* and *D. chrysanthemi*, (Supplementary Figure S3). A secondary acquisition via HGT could also be hypothesized for XynC (Supplementary Figure S3), as it is absent in *P. confusarubida* and the basal *Dickeya* species (i.e. *D. paradisiaca*, and *D. aquatica*), as well as in *D.* gs. poaceaephila, *D. zeae* and some *D. chrysanthemi* strains. Strikingly, the xylose degradation pathway (XylABFGHR), present in *P. confusarubida* and in all *Dickeya*, was absent in *D. aquatica*, indicating clearly a specific loss in this species (Supplementary Figure S3). The absence of the xylanases XynA and XynB, and xylosidase XynC as well as the xylose degradation pathway in *D. aquatica* could contribute to its restricted host range.

Proteases

Finally, the proteases PrtA, PrtB, PrtC, and PrtG, resulting from specific duplications that occurred during the diversification of *Dickeya*, and the associated type I protease secretion system PrtDEF were absent in *P. confusarubida* and the basal branching *D. paradisiaca*. Yet, *Pectobacterium* harbour closely related homologues of PrtDEF and a single protease-coding gene closely related to *Dickeya* PrtA, PrtB, PrtC and PrtG. This suggested that the whole system could have been

present in the common ancestor they shared with *Dickeya*, and then secondarily lost in *P. confusarubida* and *D. paradisiaca* (Supplementary Figure S3). PrtG was specifically lost in *D. chrysanthemi, D. dianthicola D*. gs. undicola and *D.* gs. poaceaephila. This later genomospecies conserved only one protease PrtC whereas *D. dianthicola* strains RNS04-9 and NCPPB453 contained a *prtA* pseudogene and retained two proteases PrtC, PrtB (Figure 6).

Other factors

In addition to plant cell wall degrading enzymes, *Dickeya* use several other factors to colonize plant tissue and enhance the progression of disease. Such factors include the extracellular necrosis inducing protein NipE and the two paralogous proteins AvrL and AvrM. NipE and AvrL are conserved in *Pectobacterium* and in most *Dickeya* species, except in *D. paradisiaca*, *D.* gs. poaceaephila, and *P. confusarubida*, suggesting secondary losses in these lineages (Supplementary Figure S3). AvrM was also absent in *D. zeae*, some *D. chrysanthemi*, *D*. gs. undicola, and *D. dianthicola*. Interestingly *D. zeae* strains isolated from rice were the only *Dickeya* to be devoid of Avr proteins (Supplementary Table S4). Altogether, our data indicate that the host range specificity of the various *Dickeya* species is probably linked to the particular combination of plant cell wall degrading enzymes and accessory toxins they produce.

Distribution of secretion systems in Dickeya species

For all *Dickeya* species, possession of secretion systems allowing them to actively secrete virulence factors is of crucial importance. Unsurprisingly, different protein secretion systems (T1SS to T6SS) are present in *Dickeya* species (Supplementary Table S4). As previously mentioned, the type I protease secretion system PrtDEF was present in the ancestor of all *Dickeya* species and lost in the basal *D. paradisiaca* species and in *P. confusarubida*. All *Dickeya* species are equipped with the Out specific T2SS responsible for the secretion of most pectinases and the cellulase Cel5Z. The phylogeny of the components of the Out system indicated it was likely present in the ancestor shared by *Pectobacterium*, *P. confusarubida*, and *Dickeya*, and was conserved during the diversification of *Dickeya* (Supplementary Figure S3). A second Stt specific T2SS, allows secretion

584 of the pectin lyase PnlH (Ferrandez and Condemine, 2008). Accordingly, the taxonomic
585 distributions of the pectin lyase PnlH and the Stt T2SS components were similar, being present in
586 *D. dianthicola, D. chrysanthemi,* and some *D. dadantii* strains, (Supplementary Table S4 and
587 Figure 6). In *D.* gs. poaceaephila, the Stt T2SS was also present but not associated with PnlH,
588 thus its function in this strain remains to be determined (Supplementary Table S4 and Figure 6).
589 Interestingly, a T3SS is present in *Pectobacterium*, *P. confusarubida* and in all *Dickeya* species,
590 except *D. paradisiaca* and *D.* gs. poaceaephila deprived of the T3SS and the associated DspE
591 effector, and thus suggesting secondary losses (Supplementary Table S4). Therefore, these latter
592 species are probably unable to suppress the plant immune response (see below). Interestingly,
593 phylogenies of these proteins disclosed a close relationship with two bacterial phytopathogens,
594 *Erwinia* and *Pseudomonas syringae*, suggesting that HGT occurred among these lineages
595 (Supplementary Figure S3). In fact, the effector DspE belongs to the AvrE superfamily of Type III
596 effectors (T3Es) (Degrave et al., 2015). The AvrE family is the only family of T3Es present in all
597 type III-dependent, agriculturally important phytobacterial lineages that belong to the unrelated
598 *Enterobacteriales*, *Xanthomonadales*, *Pseudomonadales* and *Ralstonia* taxa. This indicates that
599 HGT of these effectors occurred in the ancestors of these important plant pathogen lineages
600 (Jacobs et al., 2013). Recent studies indicated that AvrE-type effectors alter the sphingolipid
601 pathway *in planta* by inhibiting the serine palmitoyl transferase (Siamer *et al.*, 2014). The
602 sphingolipid biosynthetic pathway is induced during the plant hypersensitive response that blocks
603 pathogen attack at the site of infection (Berkey *et al.*, 2012). Therefore, inhibition of this pathway
604 delays hypersensitive response-dependent cell death and allows bacterial development *in planta*
605 (Degrave et al., 2015). In *Dickeya* species, the T3SS genes are in synteny with the *plcA* gene
606 encoding a phospholipase. These genes were probably acquired during the same event since *D.*
607 *paradisiaca* and *D.* gs. poaceaephila that were deprived of the T3SS are also deprived of PlcA
608 (Figure 6).
609 All *Dickeya* species and *P. confusarubida* were found to possess a two-partner secretion system
610 (T5SS) CdiB-CdiA mediating bacterial intercellular competition. Their phylogenies clearly indicate
611 an ancestral presence in both lineages (Supplementary Figure S3). CdiB is a transport protein that

exports and presents CdiA proteins on the cell surface (Willett et al., 2015). The Cdi system is involved in contact-dependent growth inhibition (CDI) by delivering the C-terminal toxin domain of CdiA (CdiA-CT) to target bacteria (Aoki et al., 2010). Some *Dickeya* strains are equipped with two CdiA proteins, for example, *D. dadantii* 3937, which produces two different CdiA-CT toxins: the first one being a tRNase and the second one harbouring DNase activity (Aoki et al., 2010; Ruhe et al., 2013). Each Cdi system also encodes a specific CdiI antitoxin that interacts with the cognate CdiA-CT toxin and prevents auto-inhibition (Willett *et al.*, 2015).

The most striking feature in *D. aquatica* was the absence of both type IV (T4SS) and type VI (T6SS) secretion systems, a trait that was shared with *D. paradisiaca, D.* gs. poaceaephila and *P. confusarubida*. Yet, the presence of closely related T4SS and T6SS in *Pectobacterium* and other *Dickeya* species suggests that both systems were present in the ancestor of *Dickeya* and *Prodigiosinella*, and secondarily lost in the three mentioned species (Supplementary Figure S3). T4SS systems were used to transport a variety of biomolecules (DNA or proteins) across the bacterial envelope (Chandran Darbari and Waksman, 2015). Most T4SS detected in *Dickeya* species are associated with conjugal transfer proteins and thus, correspond likely to conjugative T4SS that transferred DNA (de la Cruz et al., 2010; Ilangovan et al., 2015). This process is instrumental in bacterial adaptation to environmental changes (Thomas and Nielsen, 2005). T6SS is used for interaction with the host and for inter-bacterial competition (Poole et al., 2011). Unfortunately, while effectors associated to the *Dickeya* T6SS carry C-terminal nuclease domains that degrade target cell DNA, little is known concerning their function in virulence (Ryu, 2015), limiting the interpretation of its absence in *D. aquatica*.

Distribution of plant adherence elements in Dickeya species

A chaperone–usher pilus assembly pathway, associated with type I fimbriae (*fimEAICDFGHB*), was present in *D. aquatica* but not in any other *Dickeya* species (Supplementary Table S4). Phylogenetic analyses suggested an acquisition from *Morganellaceae* or *Enterobacteriaceae*, through HGT (Supplementary Figure S3). The adhesin FimH, a two-domain protein at the tip of type I fimbriae is known to recognize mannoside structures and to be responsible for adhesion to

both animal epithelial cells and plant surface (Haahtela et al., 1985; Sauer et al., 2000). Therefore, we can hypothesize that type I fimbriae could be involved in adherence of *D. aquatica* to plant surface. Interestingly, in *Xylella fastidiosa,* the Fim system is known to be an antagonist of twitching motility caused by type IV pili (De La Fuente et al., 2007). In *D. aquatica*, the mutually exclusive production of type I fimbriae and type IV pilus could be linked to the specific pH regulation of type IV pilus. When *D. aquatica* penetrates into the intercellular apoplast, which is an acidic compartment, twitching motility would be induced to favour bacteria dissemination in plant tissues, while the type I fimbriae would be no longer required during the colonization. This regulation of twitching would thus contribute to the efficiency of *D. aquatica* for infecting tomatoes, cucumbers, and probably other acidic fruits.

While the presence of type I fimbriae is a specific feature of *D. aquatica* among *Dickeya*, the Flp/Tad pilus, involved in plant surface adherence (Nykryri et al., 2013), is restricted to *D. chrysanthemi* likely as the consequence of a HGT from another proteobacterium (Supplementary Table S4 and Supplementary Figure S3). An operon encoding a multi-repeat adhesin (Dda3937_01477) associated to a T1SS secretion pathway was found in the genome of *D. dadantii* (Supplementary Table S4). This protein contains multiple cadherin-homologous domains and is likely involved in plant adhesion. Indeed, in *Pectobacterium atrosepticum*, such a multi-repeat adhesin secreted by a type I pathway was shown to be required for binding to the host plant (Perez-Mendoza et al., 2011). This adhesion and its secretion system were absent in the basal *Dickeya* species as well as in *D. zeae* and *D. chrysanthemi.* Phylogenetic analyses indicated an acquisition in the ancestor of Cluster II, followed by a secondary loss in *D. dianthicola* (Supplementary Figure S3).

From this analysis, it appears that the different *Dickeya* species retained distinct strategies to adhere to plant surfaces and that HGT played an essential role in the acquisition of the involved genes.

Distribution of iron assimilation systems in Dickeya species

Iron acquisition by *Dickeya* is required for the systemic progression of maceration symptoms in the

668 plant hosts (Enard et al., 1988, Dellagi et al., 2005, Franza et al., 2005). To chelate iron from the
669 surroundings, most *Dickeya* species synthesize and excrete two siderophores: the
670 hydroxycarboxylate achromobactin encoded by the *acsABCDEF* cluster (Munzinger et al., 2000),
671 and the catecholate chrysobactin encoded by the *cbsABCEFHP* genes (Persmark et al., 1989).
672 The *Dickeya* strain EC16 produces dichrysobactin and linear/cyclic trichrysobactin in addition to
673 the monomeric siderophore chrysobactin (Sandy and Butler, 2011). These siderophores form a
674 complex with Fe(III) designated as ferric-siderophore (Franza and Expert, 2013). The ferric-
675 siderophores are specifically recognized by outer membrane transporters (Acr for ferric-
676 achromobactin; Fct for ferric-chrysobactin). These transporters are gated-channels energized by
677 the cytoplasmic membrane-generated proton motive force transduced by the TonB protein and its
678 auxiliary proteins ExbB and ExbD (Franza and Expert, 2013). Two pairs of ExbB and ExbD
679 proteins are present in most *Dickeya* species. Transport of a ferric-siderophore across the inner
680 membrane involves a specific ABC permease (CbrABCD for ferric achromobactin; CbuBCDG for
681 ferric chrysobactin). Interestingly, the achromobactin genes (*acs* gene cluster) and related
682 transport system (*cbr* gene cluster) were absent in *P. confusarubida* and *D. paradisiaca*,
683 suggesting they were acquired by HGT in *Dickeya* after the divergence of these two lineages.
684 *Dickeya* genes are closely related to *Pseudomonas fulva* and *P. syringae* sequences, suggesting
685 an HGT between these plant-associated bacteria (Supplementary Figure S3). By contrast the
686 chrysobactin genes (*cbs* gene cluster) and related transport system (*cbu* gene cluster) were likely
687 present in the ancestor of *Dickeya* and *P. confusarubida* (Supplementary Figure S3), and then
688 specifically lost in *D. dadantii* subspecies *dieffenbachiae.* In addition to ferric-siderophores, various
689 other iron uptake systems are present in *Dickeya*. The ferrous iron transport systems FeoAB and
690 EfeUOB can be inferred as ancestral in all *Dickeya* species and *P. confusarubida* (Supplementary
691 Figure S3). In contrast, the taxonomic distribution and the phylogeny of the YfeABCD permease
692 that can import both iron and manganese, suggested an acquisition through HGT by *D.*
693 *chrysanthemi* and Cluster II species, except *D*. gs. undicola (Supplementary Table S4 and
694 Supplementary Figure S3). The haem transport Hmu system was present in most *Dickeya*. Its
695 phylogeny suggested an ancestral presence in *Dickeya*, followed by secondary losses in *D.*

dianthicola, *D.* gs. undicola, *D. zeae*, and *D. aquatica* (Supplementary Figure S3, Supplementary Table S4). Although variable combinations of iron assimilation systems exist in *Dickeya* species, at least four systems were present in each species. This multiplicity underscores the fact that competition for this essential metal is critical for the outcome of the plant-*Dickeya* interaction.

Biosynthesis of secondary metabolites in Dickeya species

In addition to siderophores, some *Dickeya* species produce secondary metabolites such as the phytotoxin zeamine and the antifungal compound oocydin via non-ribosomal peptide synthases (NRPS) and polyketide synthases (PKS) (Zhou et al., 2011a; Matilla et al., 2012). To evaluate the diversity of secondary metabolites produced by the *Dickeya* genus, we screened the eight complete genomes (*D. paradisiaca* Ech703, *D. aquatica* 174/2, *D. zeae* EC1, *D. zeae* Ech586, *D. chrysanthemi* Ech1591, *D. solani* IPO2222, *D. fangzhongdai* N14b*, D. dadantii* 3937*)* and the three partial genomes *(D. dianthicola* RNS04.9, *D.* gs. poaceaephila Ech569, *D.* gs. undicola 2B12) for gene clusters encoding NRPS or/and PKS. Then, we analysed the evolutionary history of these genes clusters among the 49 *Dickeya* genomes. The *oocBCDEFGJKLMNOPQRSTUVW* gene cluster coding for oocydin biosynthesis proteins was present in *D. paradisiaca, D. zeae* strains isolated from rice*, D. chrysanthemi* subspecies *chrysanthemi, D. solani, D. dianthicola*, and *D. fangzhongdai* strain NCPPB 3274 (Supplementary Table S4). This could suggest an ancestral presence in *Dickeya* accompanied by losses in *D. aquatica*, *D*. gs. poaceaephila, *D. dadantii*, *D*. gs. undicola, most *D. fangzhongdai* strains, and some Cluster I strains (Supplementary Figure S3). However, the hypothesis of acquisition and spreading through HGT within *Dickeya* could not be excluded. The *zmsABCDEFGIJKLMNPQRS* gene cluster directing zeamine biosynthesis was restricted to *D. zeae* strains isolated from rice, *D. fangzhongdai* and *D. solani* (Supplementary Table S4), suggesting secondary acquisition by HGT (Supplementary Figure S3). In addition, genes involved in coronafic acid biosynthesis, a phytotoxin classically produced by *Pseudomonas syringae* (Bender et al., 1999), were present only in *D*. gs. poaceaephila and *D. dadantii* subspecies *dieffenbachiae* (Supplementary Table S4), suggesting specific acquisition via HGT. We detected four additional gene clusters encoding NRPS and PKS, (i) cluster 1 was specific to *D.*

paradisiaca and *P. confusarubida*, (ii) cluster 2 was specific to *D. paradisiaca*, (iii) cluster 3 was specific to *D. aquatica*, suggesting recent acquisitions by these species, while (iv) cluster 4 was more widely distributed, being detected in *D. aquatica, D.* gs. poaceaephila*, D. fangzhongdai, D. solani*, some *D. dadantii* strains, and *D. zeae* except the strains isolated from rice (Supplementary Table S4). To conclude, each *Dickeya* species was characterized by a specific combination of large gene clusters possibly involved in the production and secretion of toxic secondary metabolites. These clusters were likely acquired from unrelated bacteria through HGT and could have been selected based on the constraints imposed by host or environmental factors. For example, the *D. zeae* strains can be subdivided in two groups, the strains isolated from rice, which produce both zeamine and oocydin, while the other strains infecting other crops produce the fourth-type metabolite encoded by cluster 4. This difference between *D. zeae* strains was used by Zhou et *al*. (2015) to define the distinct pathovar linked to rice as *D. zeae* subsp. *oryzae.*

Evolution of Dickeya gene repertoires

The 49 *Dickeya* proteomes used in this study contained in average 4,022 proteins, ranging in size from 3,533 (*D*. gs. poaceaephila) up-to 4,352 (*D. fangzhongdai* NCPPB 3274) proteins, representing a difference of 819 proteins (Supplementary Table S3). Comparison of the 197,073 proteins contained in the 49 *Dickeya* proteomes led to the delineation of 11,566 protein families (Figure 7). These protein families correspond to the pan-proteome of *Dickeya*. Among these protein families, 13.9% (1,604) were present at least in one copy in all *Dickeya* proteomes (Figure 7) and defined the core-proteome of this genus. Yet, the size of both pan- and core-proteomes could be slightly underestimated because the proteomes of some strains were deduced from draft genomes. Nevertheless, this meant that in average, ~39.9% of the proteins of any *Dickeya* proteome belonged to the core proteome, while a given *Dickeya* proteome encompassed only ~34.8% of the pan-proteome of this genus. Random taxonomic sampling-based rarefaction curves indicated that sequencing more *Dickeya* genomes will probably not change significantly the estimated size of the core-proteome, while it appears that the pan-proteome is far from being fully disclosed (Figure 7B). This highlights the high diversity and plasticity of the gene repertoires in

752 *Dickeya*. This observation coupled to the great diversity of the virulence, stress resistance,
753 metabolism, and secretion systems imply that none of the *Dickeya* strains could be regarded as a
754 representative model for this genus. Core protein families could be punctually lost in a given strain,
755 while some protein families can be present transiently in a few strains. Thus, it is also relevant to
756 consider the persistent (i.e. protein families present in more than 90% of the strains) and volatile
757 proteomes (i.e. protein families present in less than 10% of the strains) (Touchon et al., 2009). In
758 *Dickeya*, most protein families could be classified either as persistent (2,714 protein families,
759 23.5%) or volatile (6,426 protein families, 55.6%) (Figure 7A). Unsurprisingly, the persistent
760 proteome was enriched in proteins with known functions, while volatile proteome encompassed
761 mostly hypothetical proteins, prophage elements, and transposases. It is tempting to consider the
762 genes encoding for the volatile proteome as a reservoir of functional innovations, yet the adaptive
763 potential of these genes remains a matter of debate (Touchon et al., 2009).
764 Among the 11,566 protein families inferred in *Dickeya*, 3,452 corresponded to strain specific
765 families (i.e. being present in a single proteome) (Figure 7A and Supplementary Table S8). The
766 number of strain specific protein families ranged from 336 in *D. aquatica* 174/2^{T} and 304 in *D*. gs.
767 Poaceaephila, down to zero in *Dickeya solani* strains MK16, PPO9134 and RNS0773B
768 (Supplementary Table S8). To determine the origin of these strain specific protein families, we
769 used them as seeds to query with BLASTP (e-value cut-off 10^{-4}) a local database gathering 3,104
770 complete prokaryotic proteomes, including the 49 *Dickeya* and *P. confusarubida* ATCC 39006
771 proteomes. Results indicated that 1,051 (30.4%) *Dickeya* strain specific protein families displayed
772 best hit in one of the other 48 *Dickeya* strains, meaning that those sequences were wrongly
773 considered as strain specific because they did not satisfy the coverage and identity parameters
774 used to delineate the protein families. This was not surprising because some protein families could
775 be fast-evolving, meaning that applying uniform parameters can fail to delineate correctly these
776 protein families and lead to an overestimation of the strain specific protein families. In contrast,
777 1,625 (47.1%) *Dickeya* strain specific protein families displayed best hits in non-*Dickeya*
778 proteomes, meaning that the corresponding genes were likely acquired by HGT from non-*Dickeya*
779 donors. The taxonomic distribution of the corresponding sequences pinpointed *Proteobacteria*

780 (especially *Enterobacteriales*), and to a less extent *Firmicutes* as major donors (Supplementary
781 Table S8 B-C). Yet, the contribution of these two phyla is likely overestimated due to their
782 overrepresentation in sequence databases compared to other lineages. Accordingly, these results
783 should be interpreted as general trends but additional data would be required to precisely estimate
784 the real contribution of *Firmicutes* and *Proteobacteria*. Finally, 776 (22.5%) *Dickeya* strain specific
785 protein families displayed no significant hits or no hits at all, indicating that the corresponding
786 genes were truly strain specific or corresponded to annotation errors (false positives).
787 At the species level proteome size variation ranged from 435 (*D. aquatica*) to 120 (*D. paradisiaca*)
788 proteins (Supplementary Table S8). The taxonomic distribution of the species-specific protein
789 families displayed overall similar pictures, with most of them being present in all strains of the
790 species (Supplementary Figure S6). This revealed a relative homogeneity of proteomes within
791 species. Interestingly, a few strains diverged from this general trend, such as strain NCPBB 3274
792 within *D. fangzhongdai, D. aquatica* 174/2^{T}, *D. chrysanthemi* NCPPB 402, *D. dadantii* NCPPB
793 2976, and *D. solani* RNS 0512A. This is consistent with some previous studies. For instance, *D.*
794 *dadantii* NCPPB 2976 is part of the *dieffenbachiae* subspecies and has been shown to be clearly
795 different from the other *D. dadantii* subsp *dadantii* strains based on ANI values (Zhang et al.,
796 2016). Among, *D. solani*, strain RNS 0512A was proposed to define a novel *D. solani* sub-group
797 based on the high number and wide distribution of nucleotide variations compared to other *D.*
798 *solani* strains (Khayi et al., 2015).
799 Using COUNT (Csuros, 2010) on the 12,660 protein families built with SILIX and the topology of
800 the core-pf tree, we inferred the ancestral protein repertoires at each node of the *Dickeya* core-pf
801 phylogeny (Figure 8). COUNT provided similar results for different *Dickeya* species, irrespectively
802 of the position of *D. solani*, as either the sister-group of *D. dadantii* or *D. dianthicola*
803 (Supplementary Table S9), as suggested by core-pf and rprots phylogenetic analyses (see above).
804 The main difference lies in the numbers of gains and losses at the base of *Dickeya*.
805 We inferred 4,627 protein families in the ancestor of *Dickeya*, while in average 3,921 protein
806 families are contained in present-day *Dickeya* proteomes. This corresponds to a global loss of 18%
807 of the protein families. Interestingly, loss of protein families dominated over gains and affected all

Dickeya species (Figure 8). Highest protein losses were observed on the stems leading to *D*. gs. poaceaephila and *D. aquatica* and to a lesser extent in *D. paradisiaca*, *D*. gs. undicola, and *D. dianthicola*. These losses were only partially compensated by protein family gains. Surprisingly, more gains were observed in *D. aquatica* 174/2^T, compared to the two other *D. aquatica* strains (Figure 8). We assume that this was not due to biases in the annotation process by RAST, because very similar results were obtained when using PROKKA (Seeman, 2014). In fact, this may reflect the fact that the genomes of DW 0440 and CSL RW240 strains were not completed, being reported as draft genomes. The general trends observed were robust irrespectly of the postion of *D. solani* relatively to *D. dadantii* and *D. dianthicola* (Supplementary Table S9).

Characterization of the D. aquatica mobilome

Mobile Genetic Elements (MGEs) are the main actors of the HGT and include plasmids, viruses (phages and prophages) and transposons (Jackson et al., 2011). They are often localised within genomic islands on chromosomes. The mobilome of a strain is the repertoire of all the genes associated with MGEs. Using both PHAST and IslandViewer, we detected seven phage elements and ten genomic islands in *D. aquatica* 174/2^T (Figure 1, Supplementary Table S10). Most of the genomic islands contained genes of transposases, integrases or mobile elements that were likely remnants of HGT. They also contain 105 of the 336 ORFAN genes detected in *D. aquatica* 174/2^T strain. Among the seven detected prophages, P2, P6 and P7 were related to transposable Mu-phage. These were also present in some *D. zeae* strains isolated from river as well as *D. dianthicola* strains (Supplementary Table S10). P3 and P5 were specific to some *D. aquatica* strains, while P1, a defective prophage with only few conserved genes, and P4 were widespread in all *Dickeya* species (Supplementary Table S10). Among the ten genomic islands detected in *D. aquatica* 174/2^T, seven (GI1, GI2, GI4, GI5, GI6, GI7, GI9) were mainly composed of mobile elements and small hypothetical proteins, GI2 also contained a type III restriction-modification system, whereas GI4 included a type I restriction-modification system and a toxin-antitoxin system (Supplementary Table S10). Similarly, GI6 contained a toxin-antitoxin system as well as an isolated non-ribosomal peptide synthase, which was also found in *D.* gs. poaceaephila and *D. zeae*

836 (Supplementary Table S10). GI7 contained some metabolic proteins, including the previously
837 mentioned cluster 3 encoding NRPS and PKS, as well as transporters, notably a cobalt/nickel ABC
838 transporter (Supplementary Table S10). Excluding the mobile elements, GI1, GI2, GI4, GI5, GI6
839 GI7, and GI9 were specific to *D. aquatica* strains, even if a few genes composing these GIs can be
840 punctually detected in some others strains (Supplementary Table S10). The three other genomic
841 islands (GI3, GI8, GI10) were metabolic islands (Supplementary Table S10). GI3 has been laterally
842 transferred between *Erwinia pyrofolia* and *D. aquatica* (Supplementary Figure S3). GI8 that
843 included proteins related to fatty acid metabolism, was conserved in *D. zeae* strains isolated from
844 rice or originated from China (Supplementary Table S10). GI10 contained proteins related to the
845 complete carbapenem biosynthetic pathway CarABCDE and the associated resistance proteins
846 CarFG (Supplementary Table S10). *P. confusarubida* contained genes coding for CarABCDE as
847 well as CarFG, in agreement with its capacity to synthesize the carbapenem antibiotic (carbapen-
848 2-em-3-carboxylic acid) (Thomson et al., 2000; Coulthurst et al., 2005). This cluster was conserved
849 in *D. undicola, D. dadantii* subspecies *dieffenbachiae*, some *D. chrysanthemi* strains and in the *D.*
850 *zeae* CSL-RW192 strain isolated from water (Supplementary Table S10). *D. paradisiaca* only
851 contained the *carFG* resistance genes but was deprived of the biosynthetic genes (Supplementary
852 Table S9). The phylogeny of CarABCDE and CarFG shows that relationships among the *Dickeya*
853 sequences are inconsistent with the species phylogeny, suggesting a secondary acquisition
854 through HGT in this genus (Supplementary Figure S3).
855 Overall, phages and genomic islands from *D. aquatica* are mostly species specific suggesting a
856 wide genomic plasticity in the *Dickeya* genus.
857
858 *Concluding remarks*
859 In this work, we sequenced the complete genome of *D. aquatica* 174/2 and showed that unlike
860 initially supposed, this bacterium, similarly to other *Dickeya* species, is a phytopathogen. Who is
861 the natural host of *D. aquatica* and why is this bacterium found in aquatic environments rather than
862 associated with plants as are its close relatives? Althought the data at hand are not sufficient to
863 answer these questions, we speculate that *D. aquatica* can be pathogenic for aquatic plants such

as charophytes, which are a family of complex-structured algae living in a variety of wetland and freshwater habitats including those, from which the *D. aquatica* strains have been isolated. Charophytes are thought to be the closest ancestor of land plants. The cells of these algae are surrounded by polysaccharide-based cell walls. However, their cell walls are thin and cannot be distinguished as primary or secondary cell walls. Furthermore, they lack xyloglucans, which are common in most land plants (Sarkar et al., 2009). Intriguingly, *D. aquatica* lacks the xylanases and xylose degradation pathways and we hypothetize that this could reflect an evolutionary adaptation to charophyte hosts.

The comparison of *D. aquatica* 174/2 and *Dickeya* strain proteomes available in public databases showed that the *Dickeya* genus displays a remarkable diversity featuring many unique protein families and emphasizing that our knowledge of this genus is still limited. The real size of the pan-proteome is an open question encouraging further exploratory studies on these bacteria, the success of which will depend on the collection of new strains isolated from various environments and the sequencing of the genomes of newly identified representatives of the genus. Remarkably, we observed an enormous degree of genetic plasticity in the pathogenicity determinants that enable various *Dickeya* species to colonize a wide range of plant hosts. Furthermore, in this work, we have postulated the existence of a sister genus of *Dickeya*, *Prodigiosinella*, and characterized two new *Dickeya* genomospecies. The reconstruction of ancestral genomes allowed us to gain new insights into the evolutionary history of this genus and highlighted an evolutionary trajectory dominated by the loss of protein families.

Experimental procedures

Genome sequencing, assembly and annotation

DNA for sequencing of the *D. aquatica* type strain (174/2) genome was extracted from overnight broth culture using Promega bacterial genomic DNA kit. PacBio sequencing to > 350X coverage was performed by Eurofins Genomics (https://www.eurofinsgenomics.eu/). Reads were assembled using CANU (Koren et al., 2017). The annotation was performed automatically with RAST (Aziz et al., 2008), then expertly reviewed using MAGE (Vallenet et al., 2013) and literature data. The

892 expert review allowed to assign 2,457 gene names, correct 547 annotations and add 5 CDS
893 missed by RAST. The replication origin (oriC) was predicted by OriFinder
894 (http://tubic.tju.edu.cn/Ori-Finder) (Gao and Zhang, 2008). The presence of mobile genetic
895 elements in the *D. aquatica* 174/2 genome was investigated by the following online tools:
896 IslandViewer (http://pathogenomics.sfu.ca/islandviewer) (Dhillon et al., 2015) for the GI regions,
897 CRISPRDetect (http://brownlabtools.otago.ac.nz/CRISPRDetect/predict_crispr_array.html) (Biswas
898 et al., 2016) for CRISPR arrays, whereas putative prophage sequences were identified by PHAST
899 and PHASTER analysis (http://phast.wishartlab.com/) (Zhou et al., 2011b; Arndt et al., 2016). The
900 genome sequence has been submitted to EMBL database under accession number
901 GCA_900095885 (http://www.ebi.ac.uk/).
902
903 *Virulence assays on various hosts*
904 Bacterial cultures were grown in M63G minimal medium (M63 + 0.2% w/v glucose) (Miller 1972)
905 and diluted to a given OD_{600} depending on the host: 0.2 (chicory) or 1 (potato, cucumber, tomato,
906 pineapple and kiwi). For chicory, 5 µL of bacterial suspension were injected into a 2 cm incision at
907 the center of the leaf. For potato, cucumber and tomato, 5, 100 and 200 µL of bacterial suspension
908 were injected into the vegetable, respectively. 200 µL of bacterial suspension were also injected
909 into pineapple and kiwi fruits. Plants were incubated at 30°C with 100% humidity for 18 h (chicory)
910 or 42 h (potato, cucumber and tomato) or 78 h (pineapple and kiwi). The soft rot mass is used to
911 quantify virulence.
912
913 *Stress resistance assays*
914 Bacteria were cultured at 30°C in 96 well plates using M63G (M63 + 0.2% w/v glucose) pH 7.0 as
915 minimal medium. Bacterial growth (OD_{600nm}) was monitored for 48 h using an Infinite® 200 PRO -
916 Tecan instrument. Resistance to osmotic stress was analysed using M63G enriched in 0.05 to 0.5
917 M NaCl. Resistance to oxidative stress was analysed in the same medium by adding H_2O_2
918 concentrations ranging from 25 to 200 µM. The pH effect was analyzed using the same M63G
919 medium buffered with malic acid at different pH ranging from 3.7 to 7.0.

Proteome database construction

We built a local database (DickeyaDB) gathering the proteomes of *Serratia* sp. ATCC 39006, 48 *Dickeya* available at the NCBI (by january 2017), and *D. aquatica* type strain 174/2 (Supplementary Table S3). A second database (prokaDB) containing 3,104 proteomes of prokaryotes, including the 50 proteomes of DickeyaDB, was also built.

Assembly of Dickeya and Serratia ATCC 39006 protein families

Dickeya and *Serratia* ATCC 39006 protein families were assembled with SILIX version 1.2.9 (Miele et al., 2011). More precisely, pairwise comparisons of protein sequences contained in DickeyaDB were performed using the BLASTP program version 2.2.26 with default parameters (Altschul et al., 1997). Proteins in a pair providing HSP (High-scoring Segment Pairs) with identity over 60% and covering at least 80% of the protein lengths were gathered in the same family. This led to the assembly of 12,660 protein families, among which 1,493 were present in at least one copy in all DickeyaDB proteomes. In contrast, considering the 49 *Dickeya* proteomes without taking into account *Serratia* ATCC 39006 led to the assembly of 11,566 protein families, among which 1,604 were present at least in one copy in all *Dickeya* and *Serratia* ATCC 39006 proteomes, and 1,420 in exactly one copy.

Inferrence of reference phylogenies of Dickeya

Reference phylogenies of *Dickeya* were inferred using ribosomal proteins (rprots) on the one hand and core protein families (core-pf) on the other hand. The rprots phylogenetic tree was rooted with sequences from *Serratia* ATCC 39006, together with three *Pectobacterium* species (*Pectobacterium carotovorum* PC1, *Pectobacterium atrosepticum* SCRI1043, and *Pectobacterium wasabia* WPP163) and five additional *Serratia* species (*Serratia marcescens* FGI94, *Serratia fonticola* DSMZ4576, *Serratia liquefaciens* ATCC27592, *Serratia proteamaculans* 568, and *Serratia plymuthica* AS9), while the core-pf phylogenetic tree was rooted with *Serratia* ATCC 39006 (see results).

948 Rprots sequences were extracted from the DickeyaDB using the engine of the riboDB database
949 (Jauffrit et al., 2016). Briefly, the riboDB engine allows retrieving rprots sequences through a
950 double approach combining reciprocal best-blast-hits and hidden Markov model (HMM) profiles
951 searches.
952 Starting from the 1,420 core-pf protein families containing exactly one copy in each proteome of
953 the DickeyaDB and 51 rprots, we applied several quality controls. First, within protein families
954 extremely short sequences (<30% of the median length of the family) were discarded. Then, we
955 used FastTreeMP (Price et al., 2010), and PhyloMCOA (de Vienne et al., 2012) to detect and
956 discard outlier sequences, on the basis of nodal and patristic distances. At the end of the quality
957 control process, 1,341 protein families present in more than 35 (70%) out of the 50 considered
958 proteomes and 51 rprots were kept. For each of these protein families, multiple alignments were
959 built using the CLUSTAL-Omega-1.1.0 program (Sievers et al., 2011) and trimmed using
960 GBLOCKS (Castresana, 2000) with parameters set to a minimal trimming. The trimmed multiple
961 alignments corresponding to 1,341 core-pf on the one hand and the 51 rprots on the other hand
962 have been combined using Seaview-4.5.4 (Gouy et al., 2010) to build the core-pf and the rprots
963 supermatrices containing 414,696 and 6,295 amino acid positions, respectively.
964 Maximum likelihood phylogenies of these supermatrices have been inferred with IQ-TREE-1.5.3
965 (Nguyen *et al.*, 2015) with a C60 profile mixture of the Le and Gascuel evolutionary model (Le and
966 Gascuel 2008) and a gamma distribution with four site categories ($\Gamma 4$) to model the heterogeneity
967 of evolutionary rates across sites, as proposed by the model testing tool (Kalyaanamoorthy *et al.*,
968 2017) available in IQ-TREE. The robustness of the inferred maximum likelihood trees was
969 estimated with the non-parametric bootstrap procedure implemented in IQ-TREE-1.5.3 for the
970 rprots supermatrix (100 replicates of the original alignments) and the ultrafast bootstrap approach
971 for the core-pf supermatrix (1,000 replicates).
972
973 *Phylogenetic analysis of single markers*
974 947 SSU rRNA and 125 LSU rRNA complete sequences from *Dickeya*, *Serratia* (including strain
975 ATCC 39006), and *Pectobacterium* available in public databases were retrieved and aligned with

MAFFT v7.222. The resulting multiple alignments were trimmed using BMGE-1.1 (default parameters). The phylogeny of the 947 SSU rRNA sequences was inferred using FastTree-2.1.9 (Price et al., 2010), with the GTR + gamma + cat 4 model, while the LSU rRNA tree was inferred with IQ-TREE with the TIM3+F+I+Γ4 model according to the BIC criterion, as suggested by the propose model tool available in IQ-TREE.

Homologues of proteins of interest were identified in the prokaDB with BLASTP. The first 75 HSP hits with evalue smaller than 10^{-4} were kept. The retrieved sequences, together with the seed were aligned using MAFFT v7.222, alignments were trimmed using BMGE-1.1 (default parameters). Maximum likelihood phylogenies were built using IQTREE-1.5.3 (Nguyen *et al.*, 2015) with the LG+Γ4+I+F model. In order to identify the origin and evolution of these proteins in *Dickeya*, their phylogenies were compared and reconciliated with the *Dickeya* species phylogenies, by considering the two alternative positions of *D. solani*: as either sister-group of *D. dadantii* or *D. dianthicola*.

Identification of the D. fangzhongdai *phylogenetic cluster*

To determine whether some sequenced strains are related to *D. fangzhongdai,* we used all the *D. fangzhongdai* genes available in public databases SSU-rDNA (KT992690.1), *dnaX* (KT992713.1)*, fusA* (KT992697.1)*, purA* (KT992705.1)*, recA* (KT992693.1), *gapA* (KT992701.1) and *rplB* (KT992709.1) genes, we extracted the corresponding genes from the DickeyaDB database and then used these sequences as input for phylogenetic analyses using leBIBI^{QBPP} phylogenetic positioning tool (Flandrois et al., 2015). We found that the strains B16, MK7 and S1 were systematically affiliated with *D. fangzhongdai*.

Ancestral gene content

The program COUNT (Csuros, 2010) was used for gene families evolutionary reconstruction in *Dickeya* species using the topology of the core-pf tree as reference. All the generated 12,660 families were submitted to COUNT, which can perform ancestral genome reconstruction by posterior probabilities in a phylogenetic birth-and-death model. Rates were optimized using a gain–

loss–duplication model and three discrete gamma categories capturing rate variation across families, with other parameters set at default and allowing different gain–loss and duplication–loss rates for different branches. One hundred rounds of optimization were computed. COUNT was run twice: first with *D. solani* as sister-group of *D. dadantii* and then as sister-group of *D. dianthicola*.

Acknowledgements

This work was supported by the Investissement d'Avenir grant (ANR-10-BINF-01-01), by the ANR Combicontrol grant (ANR-15-CE21-0003-01), by a grant from the FR BioEnviS and using the computing facilities of the Computing Cluster of the LBBE/PRABI. We would like to thank Georgi Muskhelishvili for critical reading the manuscript and Frederic Jauffrit as well as Mailys Dumet for providing the extraction of ribosomal proteins.

Conflict of interest statement

The authors declare that no conflicting interests exist.

References

Alič Š, Naglic T, Llop P, Toplak N, Koren S, Ravnikar M & Dreo T (2015) Draft Genome Sequences of *Dickeya* sp. Isolates B16 (NIB Z 2098) and S1 (NIB Z 2099) Causing Soft Rot of Phalaenopsis Orchids. *Genome announcements* **3**.

Alič Š, Van Gijsegem F, Pédron J, Ravnikar M & Dreo T (2018) Diversity within the novel *Dickeya fangzhongdai* sp., isolated from infected orchids, water and pears. *Plant Pathology* **67***, Issue 7,* September 2018: 1612-1620.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.

Aoki, S.K., Diner, E.J., de Roodenbeke, C.T., Burgess, B.R., Poole, S.J., Braaten, B.A. et al. (2010) A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. *Nature* **468**: 439-442.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016) PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* **44**: W16-21.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A. et al. (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Bearson, S., Bearson, B., and Foster, J.W. (1997) Acid stress responses in enterobacteria. *FEMS Microbiol*

Lett **147**: 173-180.

Bender, C.L., Alarcon-Chaidez, F., and Gross, D.C. (1999) *Pseudomonas syringae* phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiol Mol Biol Rev* **63**: 266-292.

Berkey, R., Bendigeri, D., and Xiao, S. (2012) Sphingolipids and plant defense/disease: the "death" connection and beyond. *Front Plant Sci* **3**: 68.

Bertani I, Passos da Silva D, Abbruscato P, Piffanelli P and Venturi V (2013) Draft Genome Sequence of the Plant Pathogen *Dickeya zeae* DZ2Q, Isolated from Rice in Italy. *Genome announcements* **1**.

Biswas, A., Staals, R.H., Morales, S.E., Fineran, P.C., and Brown, C.M. (2016) CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics* **17**: 356.

Brady, C.L., Cleenwerck, I., Denman, S., Venter, S.N., Rodriguez-Palenzuela, P., Coutinho, T.A., and De Vos, P. (2012) Proposal to reclassify *Brenneria quercina* (Hildebrand and Schroth 1967) Hauben et al. 1999 into a new genus, *Lonsdalea* gen. nov., as *Lonsdalea quercina* comb. nov., descriptions of *Lonsdalea quercina* subsp. *quercina* comb. nov., *Lonsdalea quercina* subsp. *iberica* subsp. nov. and *Lonsdalea quercina* subsp. *britannica* subsp. nov., emendation of the description of the genus *Brenneria*, reclassification of *Dickeya dieffenbachiae* as *Dickeya dadantii* subsp. *dieffenbachiae* comb. nov., and emendation of the description of *Dickeya dadantii*. *Int J Syst Evol Microbiol* **62**: 1592-1602.

Brenner, D.J., and Farmer III, J.J. (2005) Family I. *Enterobacteriaceae*. In *Bergey's Manual of Systematic Bacteriology*. D. J. Brenner, N. R. Krieg, J. T. Staley, G. M. Garrity, D. R. Boone, P. Vos et al. (eds). New York, US: Springer, pp. 587-607.

Burdman, S., Bahar, O., Parker, J.K., and De La Fuente, L. (2011) Involvement of Type IV Pili in Pathogenicity of Plant Pathogenic Bacteria. *Genes (Basel)* **2**: 706-735.

Busse-Wicher, M., Li, A., Silveira, R.L., Pereira, C.S., Tryfona, T., Gomes, T.C. et al. (2016) Evolution of Xylan Substitution Patterns in Gymnosperms and Angiosperms: Implications for Xylan Interaction with Cellulose. *Plant Physiol* **171**: 2418-2431.

Buttner, D. (2016) Behind the lines-actions of bacterial type III effector proteins in plant cells. *FEMS Microbiol Rev*.

Caffall, K.H., and Mohnen, D. (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr Res* **344**: 1879-1900.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**: 540-552.

Chan KG, Kher HL, Chang CY, Yin WF and Tan KH (2015) Analysis of Pectate Lyase Genes in *Dickeya chrysanthemi* Strain L11, Isolated from a Recreational Lake in Malaysia: a Draft Genome Sequence Perspective. *Genome announcements* **3**.

Chandran Darbari, V., and Waksman, G. (2015) Structural Biology of Bacterial Type IV Secretion Systems. *Annu Rev Biochem* **84**: 603-629.

Conway, J.R., Lex, A., and Gehlenborg, N. (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**: 2938-2940.

Coulthurst, S.J., Barnard, A.M., and Salmond, G.P. (2005) Regulation and biosynthesis of carbapenem antibiotics in bacteria. *Nat Rev Microbiol* **3**: 295-306.

1077 Csuros, M. (2010) Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood.
1078 *Bioinformatics* **26**: 1910-1912.
1079 Czajkowski, R., Pérombelon, M.C.M., van Veen, J.A., and van der Wolf, J.M. (2011) Control of blackleg and
1080 tuber soft rot of potato caused by *Pectobacterium* and *Dickeya* species: a review. *Plant Pathol* **60**: 999–
1081 1013.
1082 Daub, J., Eberhardt, R.Y., Tate, J.G., and Burge, S.W. (2015) Rfam: annotating families of non-coding RNA
1083 sequences. *Methods Mol Biol* **1269**: 349-363.
1084 de la Cruz, F., Frost, L.S., Meyer, R.J., and Zechner, E.L. (2010) Conjugative DNA metabolism in Gram-
1085 negative bacteria. *FEMS Microbiol Rev* **34**: 18-40.
1086 De La Fuente, L., Burr, T.J., and Hoch, H.C. (2007) Mutations in type I and type IV pilus biosynthetic genes
1087 affect twitching motility rates in *Xylella fastidiosa*. *J Bacteriol* **189**: 7507-7510.
1088 de Vienne, D.M., Ollier, S., and Aguileta, G. (2012) Phylo-MCOA: a fast and efficient method to detect outlier
1089 genes and species in phylogenomics using multiple co-inertia analysis. *Mol Biol Evol* **29**: 1587-1598.
1090 Degrave, A., Siamer, S., Boureau, T., and Barny, M.A. (2015) The AvrE superfamily: ancestral type III
1091 effectors involved in suppression of pathogen-associated molecular pattern-triggered immunity. *Mol Plant*
1092 *Pathol* **16**: 899-905.
1093 Dekker, J., and Frank, K. (2015) *Salmonella, Shigella*, and *Yersinia*. *Clin Lab Med* **35**: 225-246.
1094 Dellagi, A., Rigault, M., Segond, D., Roux, C., Kraepiel, Y., Cellier, F. et al. (2005) Siderophore-mediated
1095 upregulation of Arabidopsis ferritin expression in response to *Erwinia chrysanthemi* infection. *Plant J* **43**:
1096 262-272.
1097 Dhillon, B.K., Laird, M.R., Shay, J.A., Winsor, G.L., Lo, R., Nizam, F. et al. (2015) IslandViewer 3: more
1098 flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res* **43**: W104-
1099 108.
1100 Duprey, A., Muskhelishvili, G., Reverchon, S., and Nasser, W. (2016a) Temporal control of *Dickeya dadantii*
1101 main virulence gene expression by growth phase-dependent alteration of regulatory nucleoprotein
1102 complexes. *Biochim Biophys Acta* **1859**: 1470-1480.
1103 Duprey, A., Nasser, W., Leonard, S., Brochier-Armanet, C., and Reverchon, S. (2016b) Transcriptional Start
1104 Site Turnover in the Evolution of Bacterial Paralogous Genes: the *pelE-pelD* Virulence Genes in *Dickeya*.
1105 *FEBS J*. **283**(22):4192-4207.
1106 Enard, C., Diolez, A., and Expert, D. (1988) Systemic virulence of *Erwinia chrysanthemi* 3937 requires a
1107 functional iron assimilation system. *J Bacteriol* **170**: 2419-2426.
1108 Expert, D., Boughammoura, A., and Franza, T. (2008) Siderophore-controlled iron assimilation in the
1109 enterobacterium *Erwinia chrysanthemi*: evidence for the involvement of bacterioferritin and the Suf iron-
1110 sulfur cluster assembly machinery. *J Biol Chem* **283**: 36564-36572.
1111 Ferrandez, Y., and Condemine, G. (2008) Novel mechanism of outer membrane targeting of proteins in
1112 Gram-negative bacteria. *Mol Microbiol* **69**: 1349-1357.
1113 Fineran PC, Iglesias Cans MC, Ramsay JP*, et al.* (2013) Draft Genome Sequence of *Serratia* sp. Strain
1114 ATCC 39006, a Model Bacterium for Analysis of the Biosynthesis and Regulation of Prodigiosin, a
1115 Carbapenem, and Gas Vesicles. *Genome announcements* **1**.

1116 Flandrois JP, Perrière G, and Gouy M. (2015) leBIBIQBPP: a set of databases and a webtool for automatic
1117 phylogenetic analysis of prokaryotic sequences. *BMC Bioinformatics*. **16:** 251.
1118 Fontenot, E.M., Ezelle, K.E., Gabreski, L.N., Giglio, E.R., McAfee, J.M., Mills, A.C. et al. (2013) YfdW and
1119 YfdU are required for oxalate-induced acid tolerance in *Escherichia coli* K-12. *J Bacteriol* **195**: 1446-1455.
1120 Foster, J.W. (2004) *Escherichia coli* acid resistance: tales of an amateur acidophile. *Nat Rev Microbiol* **2**:
1121 898-907.
1122 Franza, T., and Expert, D. (2013) Role of iron homeostasis in the virulence of phytopathogenic bacteria: an
1123 'a la carte' menu. *Mol Plant Pathol* **14**: 429-438.
1124 Franza, T., Mahe, B., and Expert, D. (2005) *Erwinia chrysanthemi* requires a second iron transport route
1125 dependent of the siderophore achromobactin for extracellular growth and plant infection. *Mol Microbiol*
1126 **55**: 261-275.
1127 Franza, T., Michaud-Soret, I., Piquerel, P., and Expert, D. (2002) Coupling of iron assimilation and
1128 pectinolysis in *Erwinia chrysanthemi* 3937. *Mol Plant Microbe Interact* **15**: 1181-1191.
1129 Gao, F., and Zhang, C.T. (2008) Ori-Finder: a web-based system for finding oriCs in unannotated bacterial
1130 genomes. *BMC Bioinformatics* **9**: 79.
1131 Garlant L, Koskinen P, Rouhiainen L, Laine P, Paulin L, Auvinen P, Holm L and Pirhonen M (2013) Genome
1132 Sequence of *Dickeya solani*, a New soft Rot Pathogen of Potato, Suggests its Emergence May Be
1133 Related to a Novel Combination of Non-Ribosomal Peptide/Polyketide Synthetase Clusters. *Diversity* **5**:
1134 824-842.
1135 Glasner, J.D., Yang, C.H., Reverchon, S., Hugouvieux-Cotte-Pattat, N., Condemine, G., Bohin, J.P. et al.
1136 (2011) Genome sequence of the plant-pathogenic bacterium *Dickeya dadantii* 3937. *J Bacteriol* **193**:
1137 2076-2077.
1138 Golanowska M, Galardini M, Bazzicalupo M, Hugouvieux-Cotte-Pattat N, Mengoni A, Potrykus M, Slawiak M
1139 and Lojkowska E (2015) Draft Genome Sequence of a Highly Virulent Strain of the Plant Pathogen
1140 *Dickeya solani*, IFB0099. *Genome announcements* **3**.
1141 Goto, M. (1979) Bacterial foot rot of rice caused by a strain of *Erwinia chrysanthemi*. *Phytopathology* **69**:
1142 213-216.
1143 Gouy, M., Guindon, S., and Gascuel, O. (2010) SeaView version 4: A multiplatform graphical user interface
1144 for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**: 221-224.
1145 Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New Algorithms
1146 and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0.
1147 *Systematic Biology*, **59**(3), 307–321.
1148 Haahtela, K., Tarkka, E., and Korhonen, T.K. (1985) Type 1 fimbria-mediated adhesion of enteric bacteria to
1149 grass roots. *Appl Environ Microbiol* **49**: 1182-1185.
1150 Hauben, L., Moore, E.R., Vauterin, L., Steenackers, M., Mergaert, J., Verdonck, L., and Swings, J. (1998)
1151 Phylogenetic position of phytopathogens within the Enterobacteriaceae. *Syst Appl Microbiol* **21**: 384-397.
1152 He, Z., Zhang, H., Gao, S., Lercher, M.J., Chen, W.H., and Hu, S. (2016) Evolview v2: an online visualization
1153 and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res*. **44** (Web
1154 Server issue): W236-W241.
1155 Henrichsen, J. (1972) Bacterial surface translocation: a survey and a classification. *Bacteriol Rev* **36**: 478-

503.
Henrichsen, J. (1975) The influence of changes in the environment on twitching motility. *Acta Pathol Microbiol Scand B* **83**: 179-186.
Hugouvieux-Cotte-Pattat, N., Condemine, G., and Shevchik, V.E. (2014) Bacterial pectate lyases, structural and functional diversity. *Environ Microbiol Rep* **6**: 427-440.
Ilangovan, A., Connery, S., and Waksman, G. (2015) Structural biology of the Gram-negative bacterial conjugation systems. *Trends Microbiol* **23**: 301-310.
Ishii, T. (1997) Structure and functions of feruloylated polysaccharides. *Plant Sci* **127**: 111-127.
Jackson, R.W., Vinatzer, B., Arnold, D.L., Dorus, S., and Murillo, J. (2011) The influence of the accessory genome on bacterial pathogen evolution. *Mob Genet Elements* **1**: 55-65.
Jacobs, J.M., Milling, A., Mitra, R.M., Hogan, C.S., Ailloud, F., Prior, P., and Allen, C. (2013) *Ralstonia solanacearum* requires PopS, an ancient AvrE-family effector, for virulence and To overcome salicylic acid-mediated defenses during tomato pathogenesis. *MBio* **4**: e00875-00813.
Janse, J.D., and Ruissen, M.A. (1988) Characterization and classification of *Erwinia chrysanthemi* strains from several hosts in The Netherlands. *Phytopathology* **78**: 800-808.
Jauffrit, F., Penel, S., Delmotte, S., Rey, C., de Vienne, D.M., Gouy, M. et al. (2016) RiboDB Database: A Comprehensive Resource for Prokaryotic Systematics. *Mol Biol Evol* **33**: 2170-2172.
Jiang, X., Zghidi-Abouzid, O., Oger-Desfeux, C., Hommais, F., Greliche, N., Muskhelishvili, G. et al. (2016) Global transcriptional response of *Dickeya dadantii* to environmental stimuli relevant to the plant infection. *Environ Microbiol*. **18** (11):3651-3672.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods* **14**: 587-589.
Keen, N.T., Boyd, C., and Henrissat, B. (1996) Cloning and characterization of a xylanase gene from corn strains of *Erwinia chrysanthemi*. *Mol Plant Microbe Interact* **9**: 651-657.
Khayi, S., Blin, P., Chong, T.M., Chan, K.G., and Faure, D. (2016) Complete genome anatomy of the emerging potato pathogen *Dickeya solani* type strain IPO 2222T. *Stand Genomic Sci* **11**: 87.
Khayi, S., Blin P, Pédron J, Chong TM, Chan KG, Moumni M, Hélias V, Van Gijsegem F, Faure D (2015) Population genomics reveals additive and replacing horizontal gene transfers in the emerging pathogen *Dickeya solani BMC Genomics* **16**:788
Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**: 722-736.
Krulwich, T.A., Sachs, G., and Padan, E. (2011) Molecular aspects of bacterial pH sensing and homeostasis. *Nat Rev Microbiol.* **9**(5):330-43.
Le, S. Q., and Gascuel, O. (2008) An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, **25**(7), 1307–1320.

Lee, J.H., Yeo, W.S. and Roe, J.H. (2004) Induction of the *sufA* operon encoding Fe-S assembly proteins by superoxide generators and hydrogen peroxide: involvement of OxyR, IHF and an unidentified oxidant-responsive factor. *Mol Microbiol* **51**: 1745-1755.

1196 Lee, Y.A., and Yu, C.P. (2006) A differential medium for the isolation and rapid identification of a plant soft
1197 rot pathogen, *Erwinia chrysanthemi*. *J Microbiol Methods* **64**: 200-206.
1198 Letunic, I. and Bork, P. (2016) Interactive Tree Of Life (iTOL) v3: an online tool for the display and annotation
1199 of phylogenetic and other trees. *Nucleic Acids Res* **44 (W1)**: W242-245.
1200 Li, B., Shi, Y., Ibrahim, M., Liu, H., Shan, C., Wang, Y., Kube, M., Xie, G.L. and Sun, G. (2012) Genome
1201 sequence of the rice pathogen *Dickeya zeae* strain ZJU1202. *J Bacteriol* **194**: 4452-4453.
1202 Ma, B., Hibbing, M.E., Kim, H.S., Reedy, R.M., Yedidia, I., Breuer, J. et al. (2007) Host range and molecular
1203 phylogenies of the soft rot enterobacterial genera *Pectobacterium* and *Dickeya*. *Phytopathology* **97**: 1150-
1204 1163.
1205 Maderbocus, R., Fields, B.L., Hamilton, K., Luo, S., Tran, T.H., Dietrich, L.E.P. and Tong, L. (2017) Crystal
1206 structure of a *Pseudomonas* malonate decarboxylase holoenzyme hetero-tetramer. *Nature*
1207 *communications* **8**: 160.
1208 Maier, B., and Wong, G.C. (2015) How Bacteria Use Type IV Pili Machinery on Surfaces. *Trends Microbiol*
1209 **23**: 775-788.
1210 Matilla, M.A., Stockmann, H., Leeper, F.J., and Salmond, G.P. (2012) Bacterial biosynthetic gene clusters
1211 encoding the anti-cancer haterumalide class of molecules: biogenesis of the broad spectrum antifungal
1212 and anti-oomycete compound, oocydin A. *J Biol Chem* **287**: 39125-39138.
1213 Matsumoto, H., Jitareerat, P., Baba, Y., and Tsuyumu, S. (2003) Comparative study of regulatory
1214 mechanisms for pectinase production by *Erwinia carotovora* subsp. *carotovora* and *Erwinia chrysanthemi.*
1215 *Mol Plant Microbe Interact* **16**: 226-237.
1216 Miele, V., Penel, S., and Duret, L. (2011) Ultra-fast sequence clustering from similarity networks with SiLiX.
1217 *BMC Bioinformatics* **12**: 116.
1218 Miller JH (1972) *Experiments in Molecular Genetics*. New York.
1219 Munzinger, M., Budzikiewicz, H., Expert, D., Enard, C., and Meyer, J.M. (2000) Achromobactin, a new citrate
1220 siderophore of *Erwinia chrysanthemi*. *Z Naturforsch [C]* **55**: 328-332.
1221 Nachin, L., El Hassouni, M., Loiseau, L., Expert, D., and Barras, F. (2001) SoxR-dependent response to
1222 oxidative stress and virulence of *Erwinia chrysanthemi*: the key role of SufC, an orphan ABC ATPase. *Mol*
1223 *Microbiol* **39**: 960-972.
1224 Nasser, W., Dorel, C., Wawrzyniak, J., Van Gijsegem, F., Groleau, M.C., Deziel, E., and Reverchon, S.
1225 (2013) Vfm a new quorum sensing system controls the virulence of *Dickeya dadantii*. *Environ Microbiol*
1226 **15**: 865-880.
1227 Nasser, W., Robert-Baudouy, J., and Reverchon, S. (1997) Antagonistic effect of CRP and KdgR in the
1228 transcription control of the *Erwinia chrysanthemi* pectinolysis genes. *Mol Microbiol* **26**: 1071-1082.
1229 Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015) IQ-TREE: a fast and effective
1230 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**: 268-274.
1231 Nykyri, J., Mattinen, L., Niemi, O., Adhikari, S., Koiv, V., Somervuo, P. et al. (2013) Role and regulation of
1232 the Flp/Tad pilus in the virulence of *Pectobacterium atrosepticum* SCRI1043 and *Pectobacterium*
1233 *wasabiae* SCC3193. *PLoS One* **8**: e73718.

1234 Ogasawara, H., Hasegawa, A., Kanda, E., Miki, T., Yamamoto, K., and Ishihama, A. (2007) Genomic SELEX
1235 search for target promoters under the control of the PhoQP-RstBA signal relay cascade. *J Bacteriol.*
1236 **189**(13):4791-9.
1237 O'Neill, M.A., Ishii, T., Albersheim, P., and Darvill, A.G. (2004) Rhamnogalacturonan II: structure and
1238 function of a borate cross-linked cell wall pectic polysaccharide. *Annu Rev Plant Biol* **55**: 109-139.
1239 Parkinson, N., DeVos, P., Pirhonen, M., and Elphinstone, J. (2014) *Dickeya aquatica* sp. nov., isolated from
1240 waterways. *Int J Syst Evol Microbiol* **64**: 2264-2266.
1241 Parkinson, N., Stead, D., Bew, J., Heeney, J., Tsror Lahkim, L. and Elphinstone, J. (2009) *Dickeya* species
1242 relatedness and clade structure determined by comparison of *recA* sequences. *Int J Syst Evol Microbiol*
1243 **59**: 2388-2393.
1244 Pauly, M., and Keegstra, K. (2016) Biosynthesis of the Plant Cell Wall Matrix Polysaccharide Xyloglucan.
1245 *Annu Rev Plant Biol* **67**: 235-259.
1246 Pedron, J., Mondy, S., des Essarts Y.R., Van Gijsegem, F., and Faure, D. (2014) Genomic and metabolic
1247 comparison with *Dickeya dadantii* 3937 reveals the emerging *Dickeya solani* potato pathogen to display
1248 distinctive metabolic activities and T5SS/T6SS-related toxin repertoire. *BMC Genomics* **15**: 283.
1249 Pena, M.J., Kulkarni, A.R., Backe, J., Boyd, M., O'Neill, M.A., and York, W.S. (2016) Structural diversity of
1250 xylans in the cell walls of monocots. *Planta* **244**: 589-606.
1251 Perez-Mendoza, D., Coulthurst, S.J., Humphris, S., Campbell, E., Welch, M., Toth, I.K., and Salmond, G.P.
1252 (2011) A multi-repeat adhesin of the phytopathogen, *Pectobacterium atrosepticum*, is secreted by a Type
1253 I pathway and is subject to complex regulation involving a non-canonical diguanylate cyclase. *Mol*
1254 *Microbiol* **82**: 719-733.
1255 Perombelon, M.C. (1990) *Ecology and pathology of soft rot Erwinias: an overview*. Budapest: Akadamiai
1256 Kiado.
1257 Persmark, M., Expert, D., and Neilands, J.B. (1989) Isolation, characterization, and synthesis of
1258 chrysobactin, a compound with siderophore activity from *Erwinia chrysanthemi*. *J Biol Chem* **264**: 3187-
1259 3193.
1260 Pignocchi, C., and Foyer, C.H. (2003) Apoplastic ascorbate metabolism and its role in the regulation of cell
1261 signalling. *Curr Opin Plant Biol* **6**: 379-389.
1262 Poole, S.J., Diner, E.J., Aoki, S.K., Braaten, B.A., t'Kint de Roodenbeke, C., Low, D.A., and Hayes, C.S.
1263 (2011) Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition
1264 (CDI) and rearrangement hotspot (Rhs) systems. *PLoS genetics* **7**: e1002217.
1265 Potrykus, M., Golanowska, M., Sledz, W., Zoledowska, S., Motyka, A., Kolodziejska, A. Butrymowicz,J., and
1266 Lojkowska, E. (2016) Biodiversity of *Dickeya* spp. isolated from potato plants and water sources in
1267 temperate climate. *Plant Disease* **100** (2): 408-417
1268 Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2--approximately maximum-likelihood trees for
1269 large alignments. *PLoS One* **5**: e9490.
1270 Prigent-Combaret, C., Zghidi-Abouzid, O., Effantin, G., Lejeune, P., Reverchon, S., and Nasser, W. (2012)
1271 The nucleoid-associated protein Fis directly modulates the synthesis of cellulose, an essential component
1272 of pellicle-biofilms in the phytopathogenic bacterium *Dickeya dadantii*. *Mol Microbiol* **86**: 172-186.
1273 Pritchard, L., Humphris, S., Baeyen, S., Maes, M., Van Vaerenbergh, J., Elphinstone, J., Saddler, G., and

1274 Toth, I. (2013a) Draft Genome Sequences of Four *Dickeya dianthicola* and Four *Dickeya solani* Strains.
1275 *Genome announcements* **1**.
1276 Pritchard, L., Humphris, S., Saddler, G.S,. Elphinstone, J.G., Pirhonen, M., and Toth, I.K. (2013b) Draft
1277 genome sequences of 17 isolates of the plant pathogenic bacterium *Dickeya*. *Genome announcements* **1**.
1278 Py, B., Bortoli-German, I., Haiech, J., Chippaux, M., and Barras, F. (1991) Cellulase EGZ of *Erwinia*
1279 *chrysanthemi:* structural organization and importance of His98 and Glu133 residues for catalysis. *Protein*
1280 *Eng* **4**: 325-333.
1281 Ramulu, H.G., Groussin, M., Talla, E., Planel, R., Daubin, V., and Brochier-Armanet, C. (2014) Ribosomal
1282 proteins: toward a next generation standard for prokaryotic systematics? *Mol Phylogenet Evol* **75**: 103-
1283 117.
1284 Raoul des Essarts, Y., Mondy, S., Helias, V. and Faure, D. (2015) Genome Sequence of the Potato Plant
1285 Pathogen *Dickeya dianthicola* Strain RNS04.9. *Genome announcements* **3**.
1286 Reverchon, S., and Nasser, W. (2013) *Dickeya* ecology, environment sensing and regulation of virulence
1287 program. *Environ Microbiol Rep* **5**: 622-636.
1288 Reverchon, S., Muskhelisvili, G., and Nasser, W. (2016) Virulence Program of a Bacterial Plant Pathogen:
1289 The *Dickeya* Model. *Prog Mol Biol Transl Sci* **142**: 51-92.
1290 Reverchon, S., Rouanet, C., Expert, D., and Nasser, W. (2002) Characterization of indigoidine biosynthetic
1291 genes in *Erwinia chrysanthemi* and role of this blue pigment in pathogenicity. *J Bacteriol* **184**: 654-665.
1292 Rincon-Enriquez, G., Crete, P., Barras, F., and Py, B. (2008) Biogenesis of Fe/S proteins and pathogenicity:
1293 IscR plays a key role in allowing *Erwinia chrysanthemi* to adapt to hostile conditions. *Mol Microbiol* **67**:
1294 1257-1273.
1295 Roche, B., Aussel, L., Ezraty, B., Mandin, P., Py, B., and Barras, F. (2013) Iron/sulfur proteins biogenesis in
1296 prokaryotes: formation, regulation and diversity. *Biochim Biophys Acta* **1827**: 455-469.
1297 Ruhe, Z.C., Low, D.A., and Hayes, C.S. (2013) Bacterial contact-dependent growth inhibition. *Trends*
1298 *Microbiol* **21**: 230-237.
1299 Ryu, C.M. (2015) Against friend and foe: type 6 effectors in plant-associated bacteria. *J Microbiol* **53**: 201-
1300 208.
1301 Samson, R., Legendre, J.B., Christen, R., Fischer-Le Saux, M., Achouak, W., and Gardan, L. (2005)
1302 Transfer of *Pectobacterium chrysanthemi* (Burkholder et al. 1953) Brenner et al. 1973 and *Brenneria*
1303 *paradisiaca* to the genus *Dickeya* gen. nov. as *Dickeya chrysanthemi* comb. nov. and *Dickeya*
1304 *paradisiaca* comb. nov. and delineation of four novel species, *Dickeya dadantii* sp. nov., *Dickeya*
1305 *dianthicola* sp. nov., *Dickeya dieffenbachiae* sp. nov. and *Dickeya zeae* sp. nov. *Int J Syst Evol Microbiol*
1306 **55**: 1415-1427.
1307 Sandy, M., and Butler, A. (2011) Chrysobactin siderophores produced by *Dickeya chrysanthemi* EC16. *J Nat*
1308 *Prod* **74**: 1207-1212.
1309 Sarkar, P., Bosneaga, E., and Auer, M. (2009) Plant cell walls throughout evolution: towards a molecular
1310 understanding of their design principles. *J Exp Bot* **60**: 3615-3635.
1311 Sauer, F.G., Barnhart, M., Choudhury, D., Knight, S.D., Waksman, G., and Hultgren, S.J. (2000) Chaperone-
1312 assisted pilus assembly and bacterial attachment. *Curr Opin Struct Biol* **10**: 548-556.
1313 Seeman, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30 (**14) : 2068-2069

1314 Seputiene, V., Motiejunas, D., Suziedelis, K., Tomenius, H., Normark, S., Melefors, O., and Suziedeliene, E.
1315 (2003) Molecular characterization of the acid-inducible *asr* gene of *Escherichia coli* and its role in acid
1316 stress response. *J Bacteriol* **185**: 2475-2484.
1317 Shimodaira H. (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol.* **51**(3):492-
1318 508.
1319 Siamer, S., Guillas, I., Shimobayashi, M., Kunz, C., Hall, M.N., and Barny, M.A. (2014) Expression of the
1320 bacterial type III effector DspA/E in *Saccharomyces cerevisiae* down-regulates the sphingolipid
1321 biosynthetic pathway leading to growth arrest. *J Biol Chem* **289**: 18466-18477.
1322 Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W. et al. (2011) Fast, scalable generation of
1323 high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.
1324 Sławiak, M., Beckhoven, J., Speksnijder, A., Czajkowski, R., Grabe, G. and van der Wolf, J.M. (2009)
1325 Biochemical and genetical analysis reveal a new clade of biovar 3 *Dickeya* spp. strains isolated from
1326 potato in Europe. *European Journal of Plant Pathology* **125**: 245-261.
1327 Tan, K.H., Sheng, K.Y., Chang, C.Y., Yin, W.F., and Chan, K.G. (2015) Draft Genome Sequence of a
1328 Quorum-Sensing Bacterium, *Dickeya* sp. Strain 2B12, Isolated from a Freshwater Lake. *Genome*
1329 *announcements* **3**.
1330 Thomas, C.M., and Nielsen, K.M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between
1331 bacteria. *Nat Rev Microbiol* **3**: 711-721.
1332 Thomson, N.R., Crow, M.A., McGowan, S.J., Cox, A., and Salmond, G.P. (2000) Biosynthesis of
1333 carbapenem antibiotic and prodigiosin pigment in *Serratia* is under quorum sensing control. *Mol Microbiol*
1334 **36**: 539-556.
1335 Tian, Y., Zhao, Y., Yuan, X., Yi, J., Fan, J., Xu, Z. et al. (2016) *Dickeya fangzhongdai* sp. nov., a plant-
1336 pathogenic bacterium isolated from pear trees (*Pyrus pyrifolia*). *Int J Syst Evol Microbiol* **66**: 2831-2835.
1337 Toth, I., van der Wolf, J.M., Saddler, G., Lojkowska, E., Hélias, V., Pirhonen, M. et al. (2011) *Dickeya*
1338 species: an emerging problem for potato production in Europe. *Plant Pathology* **60**: 385-399.
1339 Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C,
1340 Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard M, Dossat C, Karoui ME,
1341 Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguénec C, Lescat M, Mangenot S, Martinez-
1342 Jéhanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf CS, Schneider D, Tourret J,
1343 Vacherie B, Vallenet D, Médigue C, Rocha EP, Denamur E. (2009) Organised genome dynamics in the
1344 *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* **5**: e1000344.
1345 Urbanikova, L., Vrsanska, M., Morkeberg Krogh, K.B., Hoff, T., and Biely, P. (2011) Structural basis for
1346 substrate recognition by *Erwinia chrysanthemi* GH30 glucuronoxylanase. *FEBS J* **278**: 2105-2116.
1347 Vallenet, D., Belda, E., Calteau, A., Cruveiller, S., Engelen, S., Lajus, A. et al. (2013) MicroScope--an
1348 integrated microbial resource for the curation and comparative analysis of genomic and metabolic data.
1349 *Nucleic Acids Res* **41**: D636-647.
1350 van der Wolf, J.M., Nijhuis, E.H., Kowalewska, M.J., Saddler, G.S., Parkinson, N., Elphinstone, J.G. et al.
1351 (2014) *Dickeya solani* sp. nov., a pectinolytic plant-pathogenic bacterium isolated from potato (*Solanum*
1352 *tuberosum*). *IJSEM* **64**: 768-774.
1353 Varani, A.M., Siguier, P., Gourbeyre, E., Charneau, V., and Chandler, M. (2011) ISsaga is an ensemble of

web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol* **12**: R30.

Wandersman, C., Delepelaire, P., Letoffe, S., and Schwartz, M. (1987) Characterization of *Erwinia chrysanthemi* extracellular proteases: cloning and expression of the protease genes in *Escherichia coli*. *J Bacteriol* **169**: 5046-5053.

Wang, H.C., Minh, B.Q., Susko, E., and Roger, A.J. (2017) Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Syst Biol*.

Wheeler, T.J., and Eddy, S.R. (2013) nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**: 2487-2489.

Willett, J.L., Ruhe, Z.C., Goulding, C.W., Low, D.A., and Hayes, C.S. (2015) Contact-Dependent Growth Inhibition (CDI) and CdiB/CdiA Two-Partner Secretion Proteins. *J Mol Biol* **427**: 3754-3765.

Wolanski, M., Donczew, R., Zawilak-Pawlik, A., and Zakrzewska-Czerwinska, J. (2014) oriC-encoded instructions for the initiation of bacterial chromosome replication. *Front Microbiol* **5**: 735.

Yutin, N., Puigbo, P., Koonin, E.V., and Wolf, Y.I. (2012) Phylogenomics of prokaryotic ribosomal proteins. *PLoS One* **7**: e36972.

Zhang, J.X., Lin, B.R., Shen, H.F., and Pu, X.M. (2013) Genome Sequence of the Banana Pathogen *Dickeya zeae* Strain MS1, Which Causes Bacterial Soft Rot. *Genome announcements* **1**.

Zhang, Y., Fan, Q., and Loria, R. (2016) A re-evaluation of the taxonomy of phytopathogenic genera *Dickeya* and *Pectobacterium* using whole-genome sequencing data. *Syst Appl Microbiol.* **39**(4):252-259.

Zhou, J., Zhang, H., Wu, J., Liu, Q., Xi, P., Lee, J. et al. (2011a) A novel multidomain polyketide synthase is essential for zeamine production and the virulence of *Dickeya zeae*. *Mol Plant Microbe Interact* **24**: 1156-1164.

Zhou, J., Cheng, Y., Lv, M., Liao, L., Chen, Y., Gu, Y. et al. (2015) The complete genome sequence of *Dickeya zeae* EC1 reveals substantial divergence from other *Dickeya* strains and species. *BMC Genomics* **16**: 571.

Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., and Wishart, D.S. (2011b) PHAST: a fast phage search tool. *Nucleic Acids Res* **39**: W347-352.

Figure legends

Figure 1. Genomic organisation of *Dickeya aquatica* 174/2^T chromosome

The chromosome is represented as a wheel and the origin of replication (OriC) and terminus (Ter) are indicated. The circles from outside to inside represent protein-coding sequences (CDS) on the forward strand, CDS on the reverse strand, distribution of ribosomal RNA operons (four on the right replichore and three on the left replichore), distribution of tRNA genes, and then distribution of ncRNA. The blue areas correspond to phage elements detected using PHAST. The thin black lines correspond to four CRISPR arrays and the red areas represent the ten genomic islands predicted using IslandViewer. The next circle (black) indicates the GC content and the central circle (green/purple) shows the GC-skew. The window size of the GC content and GC-skew is 100 nucleotides. Figure 1 was drawn using Gview https://server.gview.ca/.

Figure 2. Phylogeny and proteome comparison of *Dickeya*

Maximum likelihood phylogeny (left) of 1,341 *Dickeya* single copy protein families (core protein families: 50 sequences, 414,696 amino acid positions). The tree was computed with IQ-TREE with the LG+C60 model and rooted using *Serratia* ATCC 39006. Numbers associated to branches correspond to ultrafast bootstrap values. The scale bar corresponds to evolutionary distance (i.e. the average number of the substitutions inferred per site). The table (right) corresponds to the S_{AB} association coefficient computed for each pair of strains as $S_{AB} = (100 \times 2N_{AB})/(N_A + N_B)$, in which N_A is the number of protein families present in strain A, N_B is the number of protein families present in strain B and N_{AB} is the number of protein families shared by strain A and strain B. This coefficient ranged from 0 when both strains do not share any gene family to 100 when all the families present in strain A were also present in strain B. The figure was generated using Evolview (He et al., 2016).

Figure 3. Virulence of *D. aquatica* on potato, chicory, cucumber and tomato. Bacterial cultures were grown in M63G minimal medium (M63 + 0.2% w/v glucose) and diluted to a given

OD_{600} depending on the host: 0.2 (chicory) or 1 (potato, cucumber and tomato). For chicory, 5 µL of bacterial suspension were injected into a 2 cm incision at the center of the leaf. For potato, cucumber and tomato, 5, 100 or 200 µL of bacterial suspension were injected into the vegetable, respectively. Plants were incubated at 30°C with 100% humidity for 18 h (chicory) or 42 h (potato, cucumber and tomato). **A)** Picture of representative specimens of infected plants after incubation. Note that the rotten area was removed for potato. **B)** Quantification of the soft rot mass. Data is represented as mean +/- SD of 6 replicates. No soft rot symptoms were detected after 78 h for pineapple and kiwi fruits infected with 200 µL of bacterial suspension at OD_{600} 1.

Figure 4. *D. aquatica* stress resistance. Bacteria were cultured at 30°C in 96 well plates using M63G (M63 + 0.2% w/v glucose) pH 7.0 as minimal medium. Bacterial growth (OD_{600nm}) was monitored for 48 h using an Infinite® 200 PRO - Tecan instrument. A and B) Resistance to osmotic stress was analysed using M63G enriched in 0.05 to 0.5 M NaCl (abscissa) and growth rates (ordinate) were determined. C) Resistance to oxidative stress was analysed in the same medium by adding H_2O_2 concentrations ranging from 25 to 200 µM (abscissa). The lag time (ordinate) is represented instead of the growth rate because after the degradation of H_2O_2 by bacterial catalases, the growth rates are similar. D) The pH effect on growth rate (ordinate) was analyzed using the same M63G medium buffered with malic acid at different pH ranging from 3.7 to 7.0 (abscissa).

Figure 5. Twitching motility induced under acidic condition in *Dickeya* species. Colony morphologies of various *Dickeya* strains grown on M63G pH 7.0 and M63G buffered with malic acid at pH 5 in agar plates. The strains corresponding to colony numbers and their twitching phenotypes are indicated on the right.

Figure 6. Distribution of plant cell wall degrading enzymes in *Dickeya*. The type of cell wall degrading enzyme and the number of homologues of each enzyme are indicated for each of the

strains. The phylogeny on the left corresponds the core-pf maximum likelihood tree as described in figure 2.

Figure 7: The core-, pan-, versatile-, and persistent-proteomes of *Dickeya* genus

The delineation of core-, pan-, versatile-, and persistant-proteomes was based on the taxonomic distribution of the 11,566 proteins families identified by analyzing the 49 proteomes of *Dickeya*. The core-proteome is defined by the protein families present at least in one copy in all 49 *Dickeya* proteomes, the pan-proteome is defined by the 11,566 protein families, while the versatile- and the persistent-proteomes are defined as the protein families present in less than 10% and in more than 90% of the 49 *Dickeya* proteomes, respectively.

(A) Distribution 11,566 protein families across the 49 *Dickeya* proteomes.

The 3,452 protein families present in a single *Dickeya* proteome (i.e. strain specific families) are on the left of the x-axis, while the 1,604 protein families defining the core-proteome are on the right of the x-axis.

(B) Estimation of the *Dickeya* core and pan-genomes

The graph shows the estimated sizes of the core- and pan-proteomes of *Dickeya* according to the number of considered strains. The curves were computed by calculating the core- and the pan-proteomes for an increasing number of strains randomly selected among the 49 *Dickeya* strains (100 replicates at each point). When all the 49 Dickeya strains were considered, the core- and the pan-proteomes encompasse 1,604 and 11,566 protein families.

Figure 8. Evolution of protein family repertoires along the *Dickeya* phylogeny

The number of protein family repertoires, gains, and losses inferred by COUNT are mapped on the topology of the core-pf reference phylogeny of *Dickeya*. Values mapped above the branches correspond to gains / losses, while values below branches correspond to the number of protein families inferred. Considering the alternative placement of *D. solani* as sister-group of *D. dianthicola* provided very similar results (see Supplementary Table S9).

Supplementary materials

Supplementary Figure S1: Maximum likelihood phylogenetic tree of *Dickeya* strains based on the Fr-protein supermatrix gathering 51 ribosomal proteins (LG+C60, 58 sequences, 6,295 amino acid positions)**.**

Numbers at branch correspond to bootstrap values (100 replicates of the original data set). The scale bar indicates the average number of substitutions per site. The figure was generated using Evolview (He et al., 2016).

Supplementary Figure S2: Phylogenetic position of "*Serratia* sp. ATCC 39006" based on LSU and SSU rDNA trees

Itrees of the *Dickeya* genus inferred with a collection of **(A)** 947 SSU rDNA and **(B)** 125 LSU rDNA sequences retrieved from public databases. The trees were rooted using *Pectobacterium* and *Serratia* sequences. The scale bars represent the estimated average number of substitution per site. Numbers at nodes represent ultrafast bootstrap values **(A)** and bootstrap values **(B)**. *Pectobacterium* sequences are in pink, *Serratia* sequences in purple, and *Dickeya* sequences in black. Worth to note, LSU and SSU rDNA sequences from *Serratia* sp. 39006 robustly branch with *Dickeya* sequences and not within the *Serratia* genus. The trees were drawn using the iToL webserver (Letunic and Bork, 2016).

Supplementary Figure S3 - Single gene phylogenies

Maximum likelihood phylogenetic trees of 912 proteins of interest. Numbers associated with each branch correspond to ultrafast bootstrap values (1000 replicates of the original data set). The scale bars indicate the average number of substitutions per site. For each tree, the name and the annotation of the seed is provided in red.

Supplementary Figure S4 – Non-metric multi-dimensional scaling (NMDS) plot of *Dickeya* genomes according to their gene content. A Bray distance similarity matrix was calculated

based on the presence/absence profiles of 8,115 gene families (present in at least 2 genomes) for the 49 analyzed genomes and used to generate NMDS coordinates for each strain. The shorter distance linking two genomes indicates higher similarities between these genomes. Genomes from different species are indicated in different colors.

Supplementary Figure S5: Phylogenies and genetic organisation of *pelAED* and *pehVWX* clusters.

Maximum likelihood trees of *pelAED* and *pehVWX* clusters together with genetic organisation of these clusters are shown. Numbers associated with each branch correspond to ultrafast bootstrap values (1000 replicates of the original data set). The scale bar indicates the average number of substitutions per site. For each tree, the name and the annotation of the seed is provided in red.

Supplementary Figure S6: The number of shared protein families between and within *Dickeya* species. Figures were generated using the package UpSetR (Conway et al. 2017).

Supplementary Table S1: Phenotypic differentiation of species within the genus *Dickeya*.

Table showing the phenotypic features of *Dickeya* species based on Samson et al. (2005), Parkinson et al. (2014), van der Wolf et al. (2014) and Tian et al. (2016).

Supplementary Table S2: Genomic features of the *Dickeya* strains with completely sequenced genomes

Supplementary Table S3: List of the 49 *Dickeya* genomes used in this study indicating the plant host or habitat and the geographical origin of each strain

Supplementary Table S4: Distribution of genes with a role in pathogenicity or in adaptation to plant niche in *Dickeya* species

Genes of *Dickeya* species were considered as present if identity of the encoded protein was higher

than 60% of full-length amino acids sequence. If local alignments were too short with regard to the length of similar sequences, we performed a nucleotide BLAST on full-length DNA sequences with similar thresholds (10^{-5} e-value, 80% identity of full length sequence). This allowed us to eliminate false genes. Note that because some genomes represent only drafts, false negatives may occur.

Supplementary Table S5: Protein family distribution within proteomes from *Dickeya*, *Pectobacterium* and *Serratia*.

Considered strains were *Dickeya aquatica* 174/2, *Dickeya dadantii* 3937, *Dickeya paradisiaca* NCPPB 2511, *Dickeya solani* IPO 2222, *Dickeya zeae* EC1, *Pectobacterium atrosepticum* SCRI1043, *Pectobacterium carotovorum* subsp carotovorum PC1, *Pectobacterium wasabiae* WPP163, *Serratia fonticola*, *Serratia liquefaciens* ATCC27592, *Serratia marcescens* FGI94, *Serratia plymuthica* AS9, *Serratia proteamaculans* 568 and *Serratia* sp ATCC39006. Families were built using SILIX version 1.2.9 with 60% of sequence identity and 80% of sequence coverage.

Supplementary Table S6: Results of the approximately unbiased tests performed on core-pf and rprots topologies using the 1,341 core-pf and the 51 rprots.

Supplementary Table S7: Distribution of Transcription factor-binding sites in *D. aquatica* 174/2 genome. Binding sites for 56 transcriptional regulators were predicted in *D. aquatica* using nhmmer (Wheeler and Eddy, 2013).

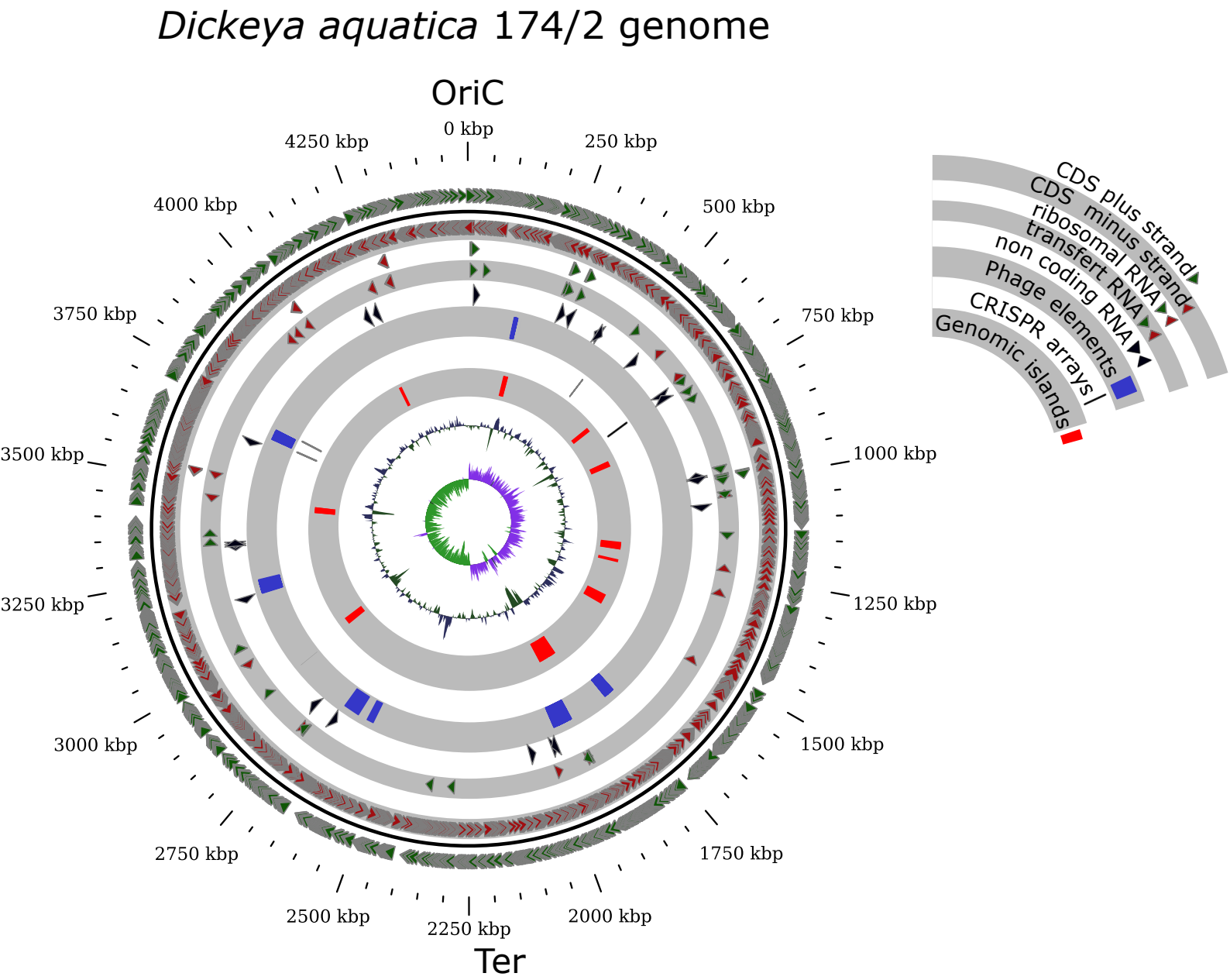
Supplementary Table S8: Protein family distribution within the proteomes of the 49 *Dickeya* strains.

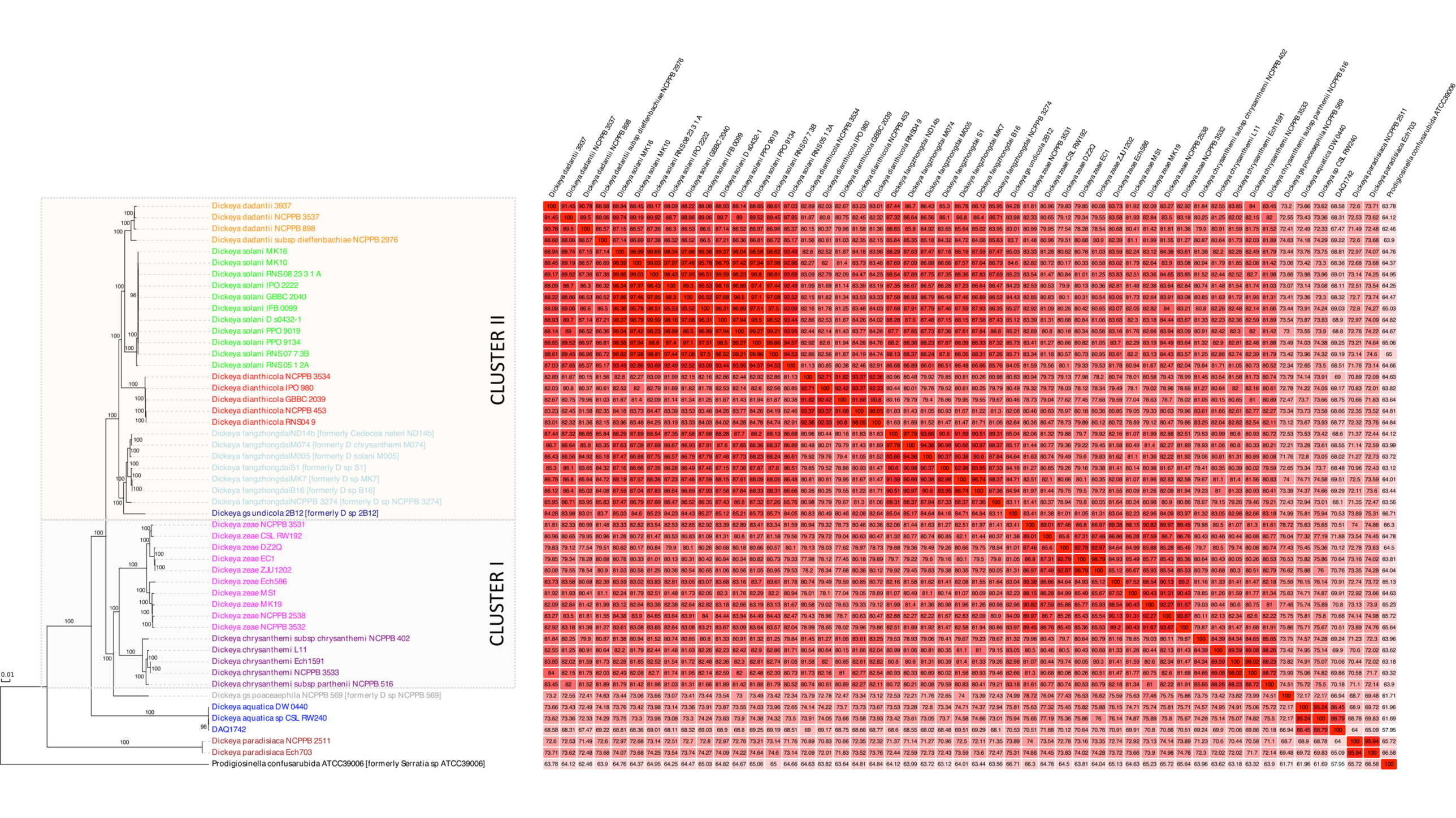
Families were computed with SILIX version 1.2.9 using 60% of sequence identity and 80% of sequence coverage.

Supplementary Table S9: Comparison of inferences of ancestral gene repertoires in *Dickeya* considering the two possible alternative positions for *D. solani*

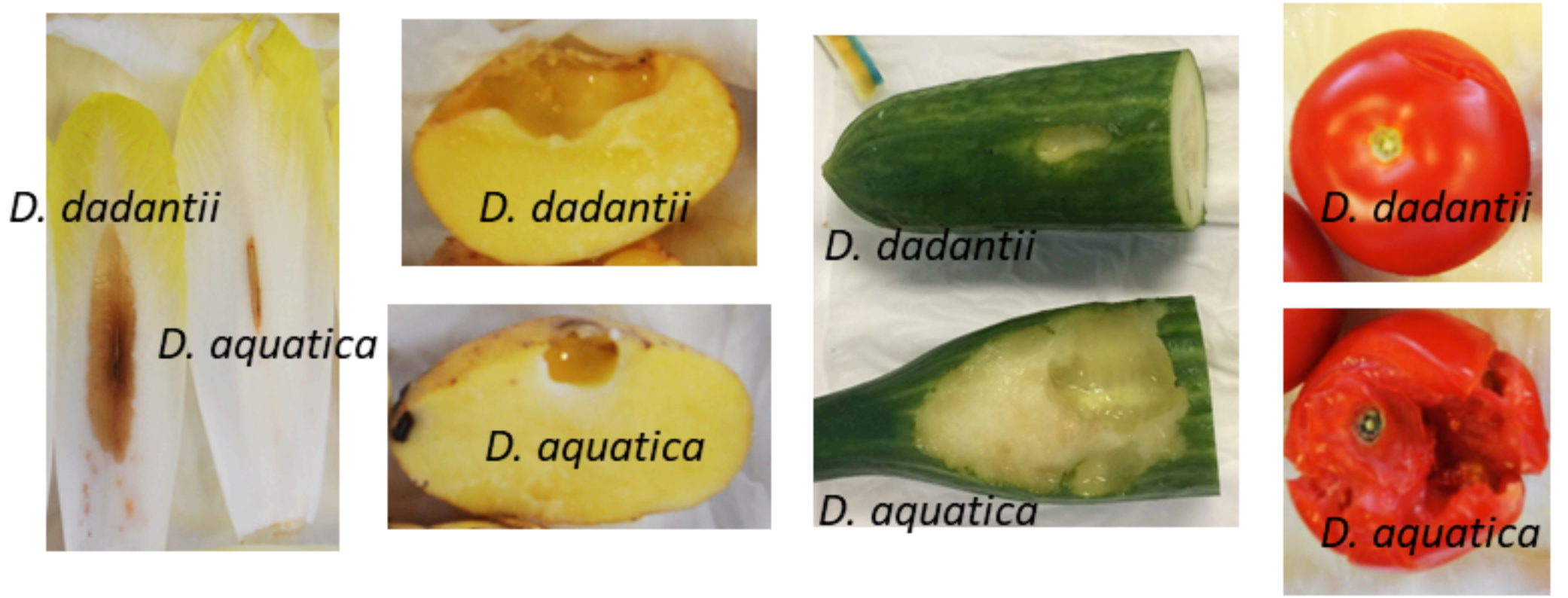
Supplementary Table S10: Characterization of the mobilome of *D. aquatica* 174/2 and its conservation in other *Dickeya* species

Genes of *Dickeya* species were considered as present if identity of the encoded protein was higher than 60% of full-length amino acids sequence. If local alignments were too short with regard to the length of similar sequences, we performed a nucleotide BLAST on full-length DNA sequences with similar thresholds (10^{-5} e-value, 80% identity of full length sequence). This allowed us to eliminate false genes. Note that because some genomes represent drafts, false negatives may occur.


Dickeya aquatica 174/2 genome
OriC
0 kbp
250 kbp
500 kbp
750 kbp
1000 kbp
1250 kbp
1500 kbp
1750 kbp
2000 kbp
2250 kbp
2500 kbp
2750 kbp
3000 kbp
3250 kbp
3500 kbp
3750 kbp
4000 kbp
4250 kbp
Ter
CDS plus strand
CDS minus strand
ribosomal RNA
transfert RNA
non coding RNA
Phage elements
CRISPR arrays
Genomic islands


CLUSTER II
CLUSTER I
Dickeya dadantii 3937
Dickeya dadantii NCPPB 3537
Dickeya dadantii NCPPB 898
Dickeya dadantii subsp dieffenbachiae NCPPB 2976
Dickeya solani MK16
Dickeya solani MK10
Dickeya solani RNS08 23 3 1 A
Dickeya solani IPO 2222
Dickeya solani GBBC 2040
Dickeya solani IFB 0099
Dickeya solani D s0432-1
Dickeya solani PPO 9019
Dickeya solani PPO 9134
Dickeya solani RNS07 7 3B
Dickeya solani RNS05 1 2A
Dickeya dianthicola NCPPB 3534
Dickeya dianthicola IPO 980
Dickeya dianthicola GBBC 2039
Dickeya dianthicola NCPPB 453
Dickeya dianthicola RNS04 9
Dickeya fangzhongdaiND14b [formerly Cedecea neteri ND14b]
Dickeya fangzhongdaiM074 [formerly D chrysanthemi M074]
Dickeya fangzhongdaiM005 [formerly D solani M005]
Dickeya fangzhongdaiS1 [formerly D sp S1]
Dickeya fangzhongdaiMK7 [formerly D sp MK7]
Dickeya fangzhongdaiB16 [formerly D sp B16]
Dickeya fangzhongdaiNCPPB 3274 [formerly D sp NCPPB 3274]
Dickeya gs undicola 2B12 [formerly D sp 2B12]
Dickeya zeae NCPPB 3531
Dickeya zeae CSL RW192
Dickeya zeae DZ2Q
Dickeya zeae EC1
Dickeya zeae ZJU1202
Dickeya zeae Ech586
Dickeya zeae MS1
Dickeya zeae MK19
Dickeya zeae NCPPB 2538
Dickeya zeae NCPPB 3532
Dickeya chrysanthemi subsp chrysanthemi NCPPB 402
Dickeya chrysanthemi L11
Dickeya chrysanthemi Ech1591
Dickeya chrysanthemi NCPPB 3533
Dickeya chrysanthemi subsp parthenii NCPPB 516
Dickeya gs poaceaephila NCPPB 569 [formerly D sp NCPPB 569]
Dickeya aquatica DW 0440
Dickeya aquatica sp CSL RW240
DAQ1742
Dickeya paradisiaca NCPPB 2511
Dickeya paradisiaca Ech703
Prodigiosinella confusarubida ATCC39006 [formerly Serratia sp ATCC39006]
0.01

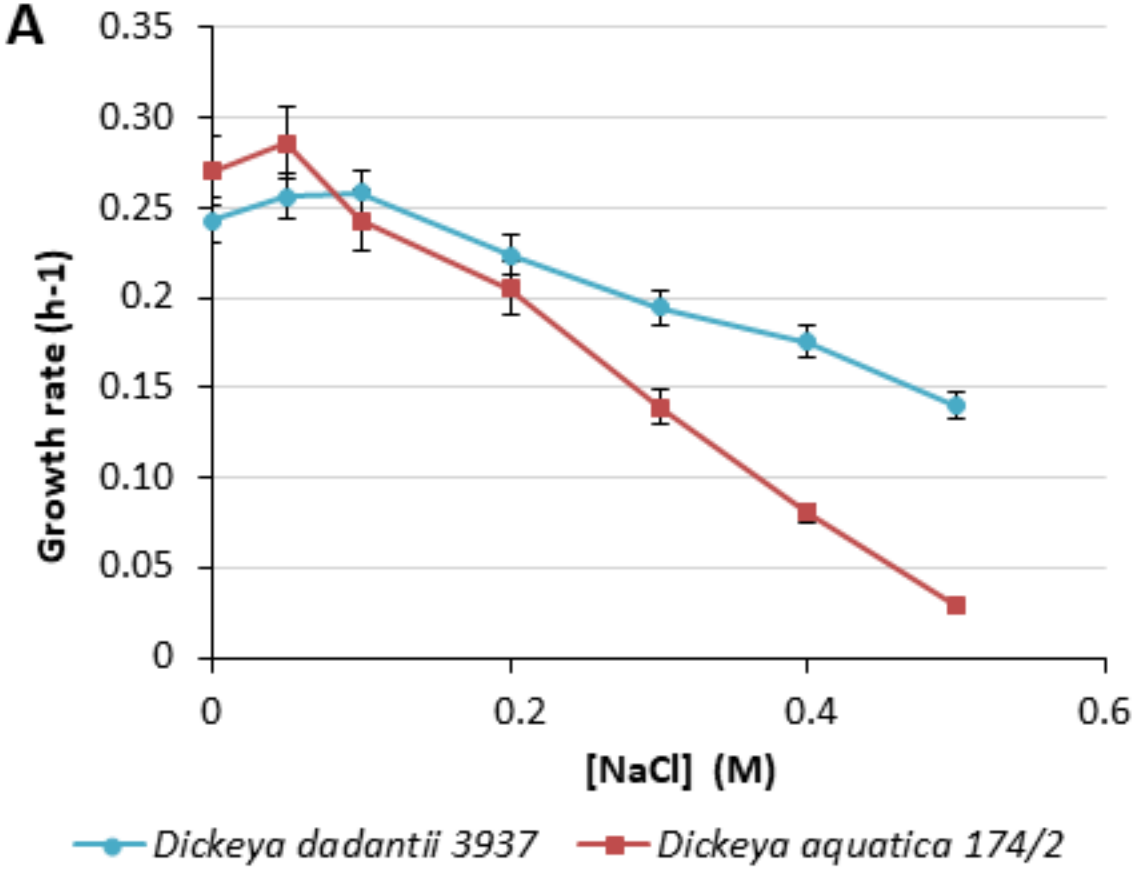
A



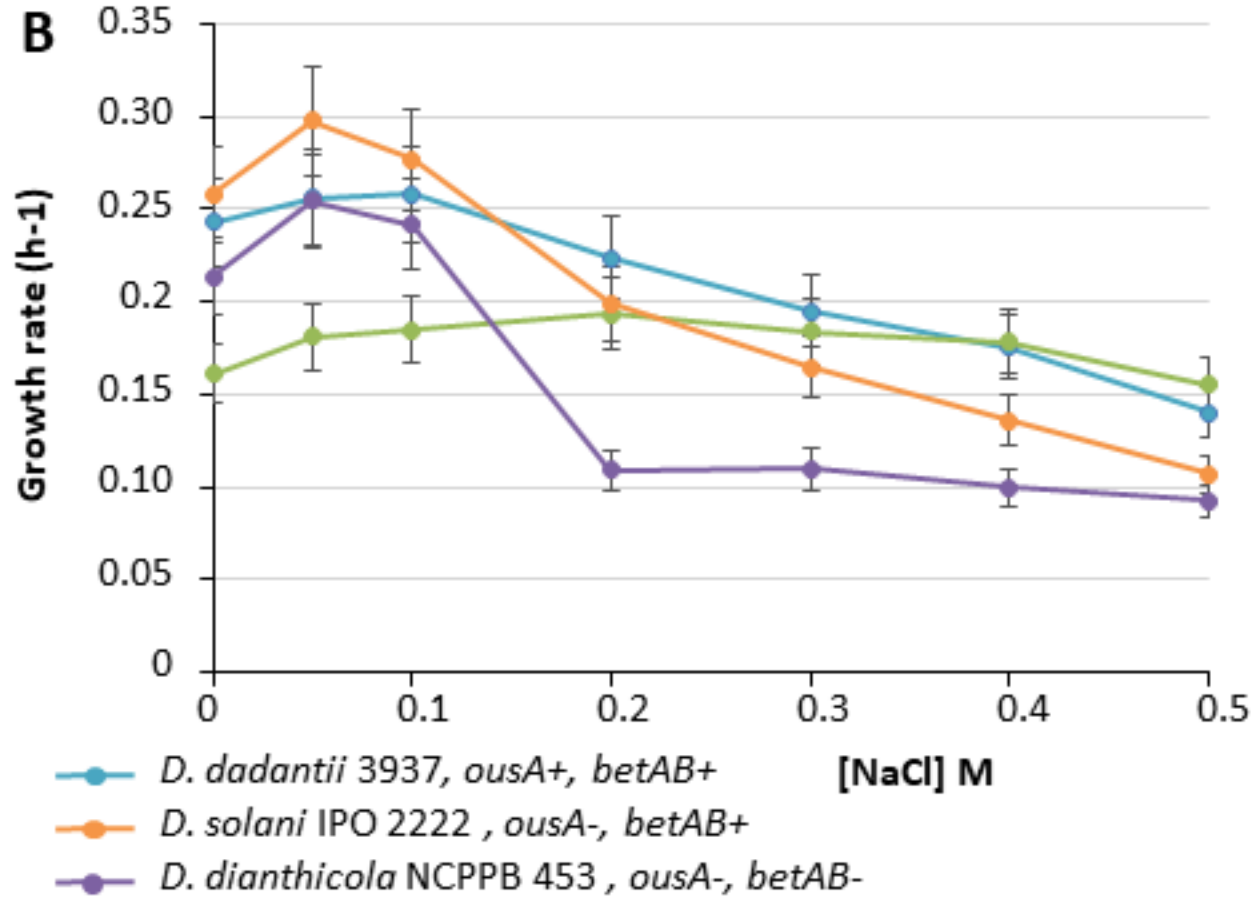
B

Rotten mass (g)	*D. dadantii*	*D. aquatica*
Chicory (leaf)	0.9 ± 0.5	0.02 ± 0.04
Potato (tuber)	5.7 ± 2.4	1.7 ± 0.2
Cucumber (fruit)	0.38 ± 0.49	11.6 ± 10.6
Tomato (fruit)	8.0 ± 4.6	36 ± 16
Pineapple (fruit)	0.0 ± 0.0	0.0 ± 0.0
Kiwi (fruit)	0.0 ± 0.0	0.0 ± 0.0

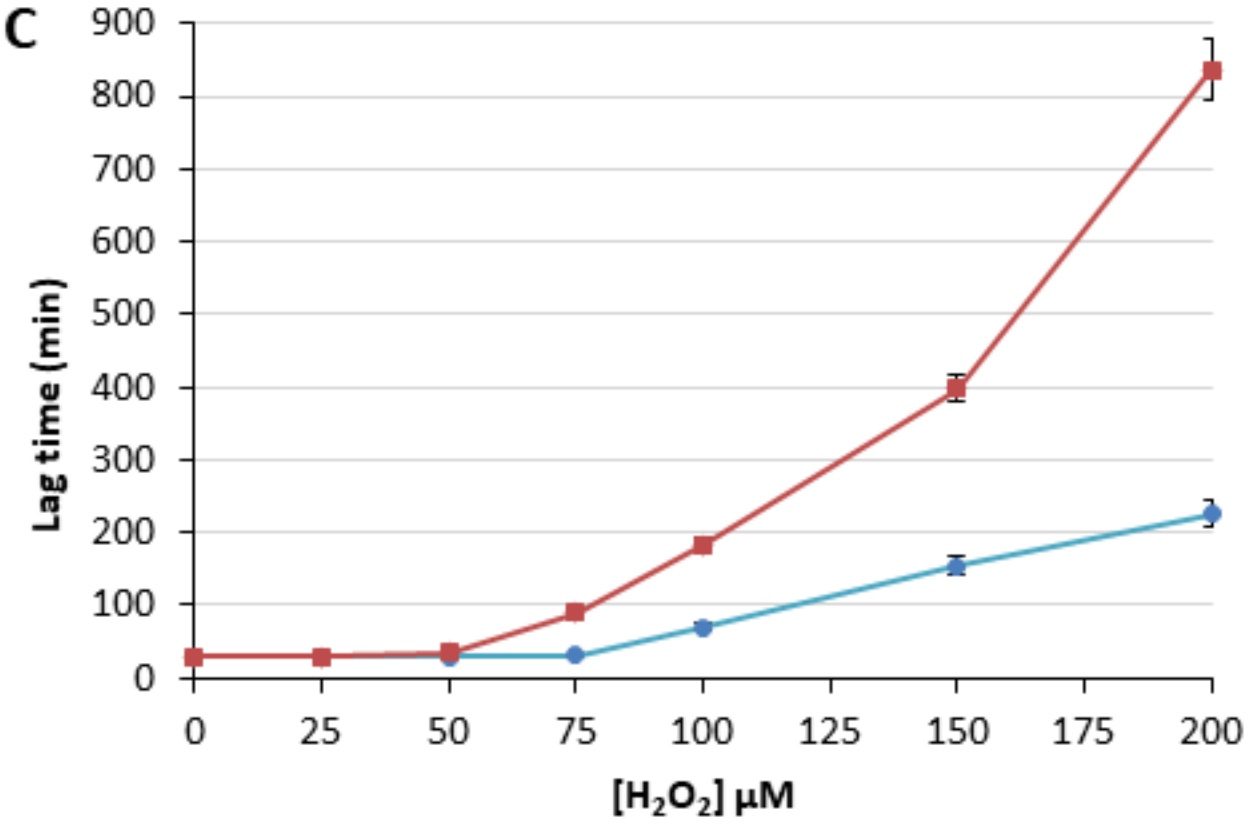
A


Growth rate (h-1)
0.35
0.30
0.25
0.2
0.15
0.10
0.05
0
0
0.2
0.4
0.6
[NaCl] (M)
Dickeya dadantii 3937
Dickeya aquatica 174/2

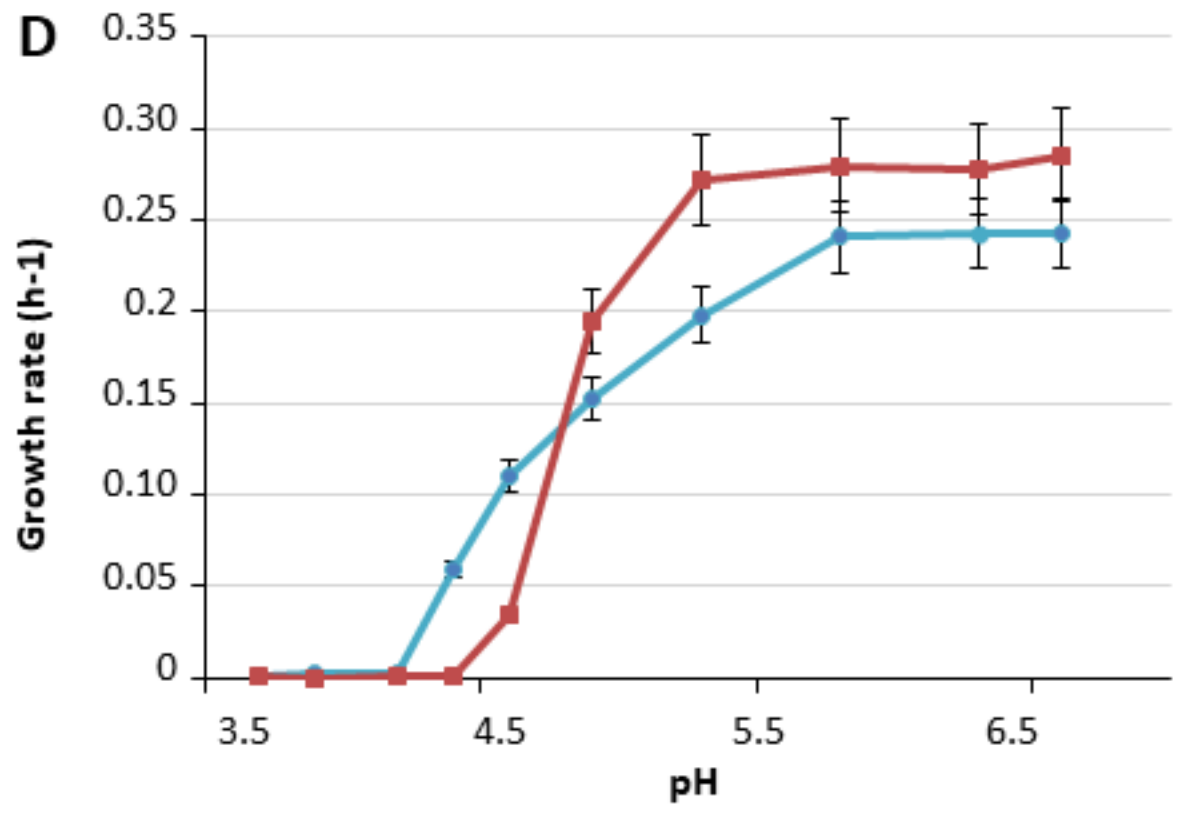
B

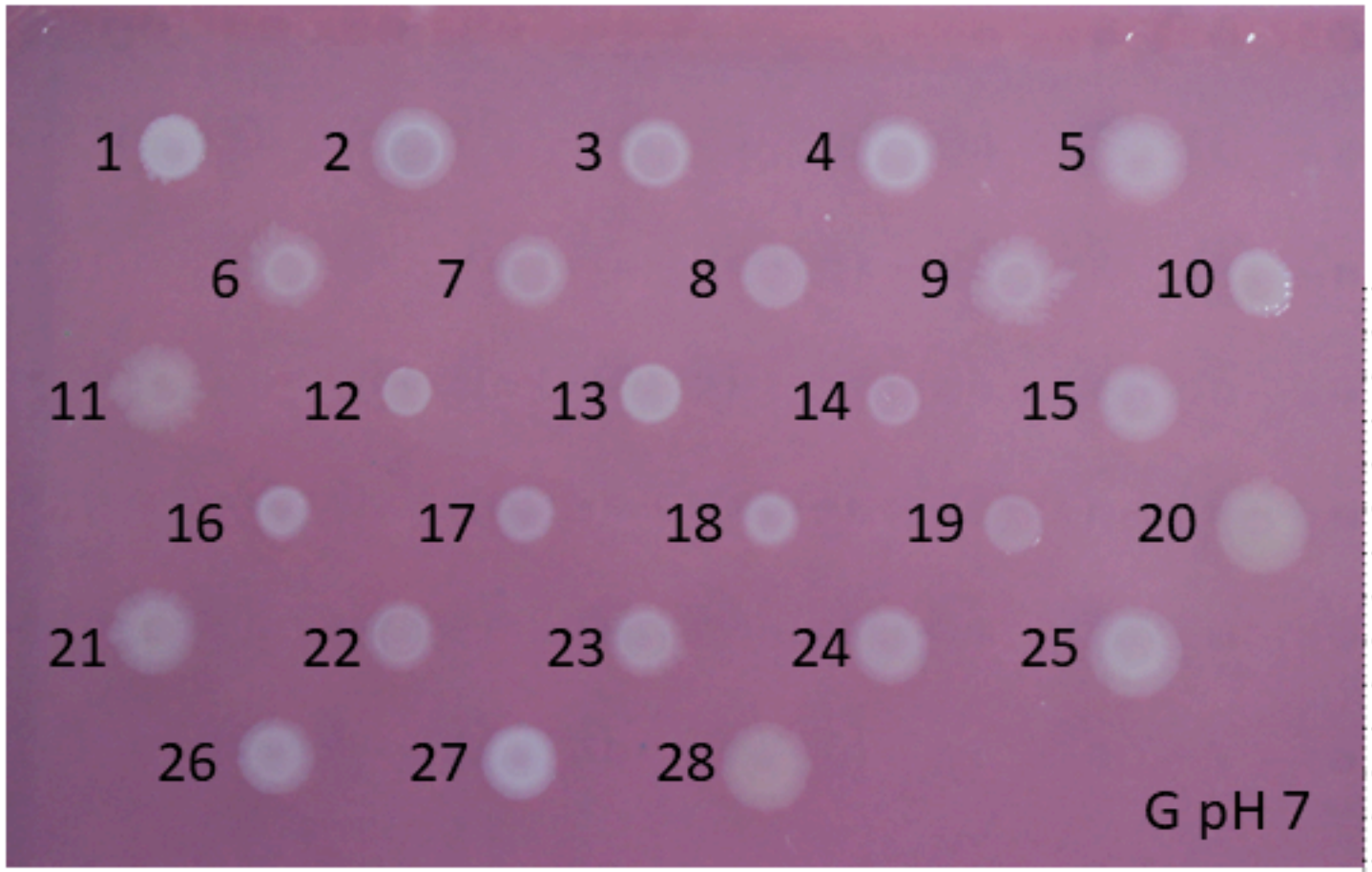

Growth rate (h-1)
0.35
0.30
0.25
0.2
0.15
0.10
0.05
0
0
0.1
0.2
0.3
0.4
0.5
[NaCl] M
D. dadantii 3937, ousA+, betAB+
D. solani IPO 2222 , ousA-, betAB+
D. dianthicola NCPPB 453 , ousA-, betAB-
D. chrysanthemi subsp. chrysanthemi NCPPB 402 , ousA+, betAB-

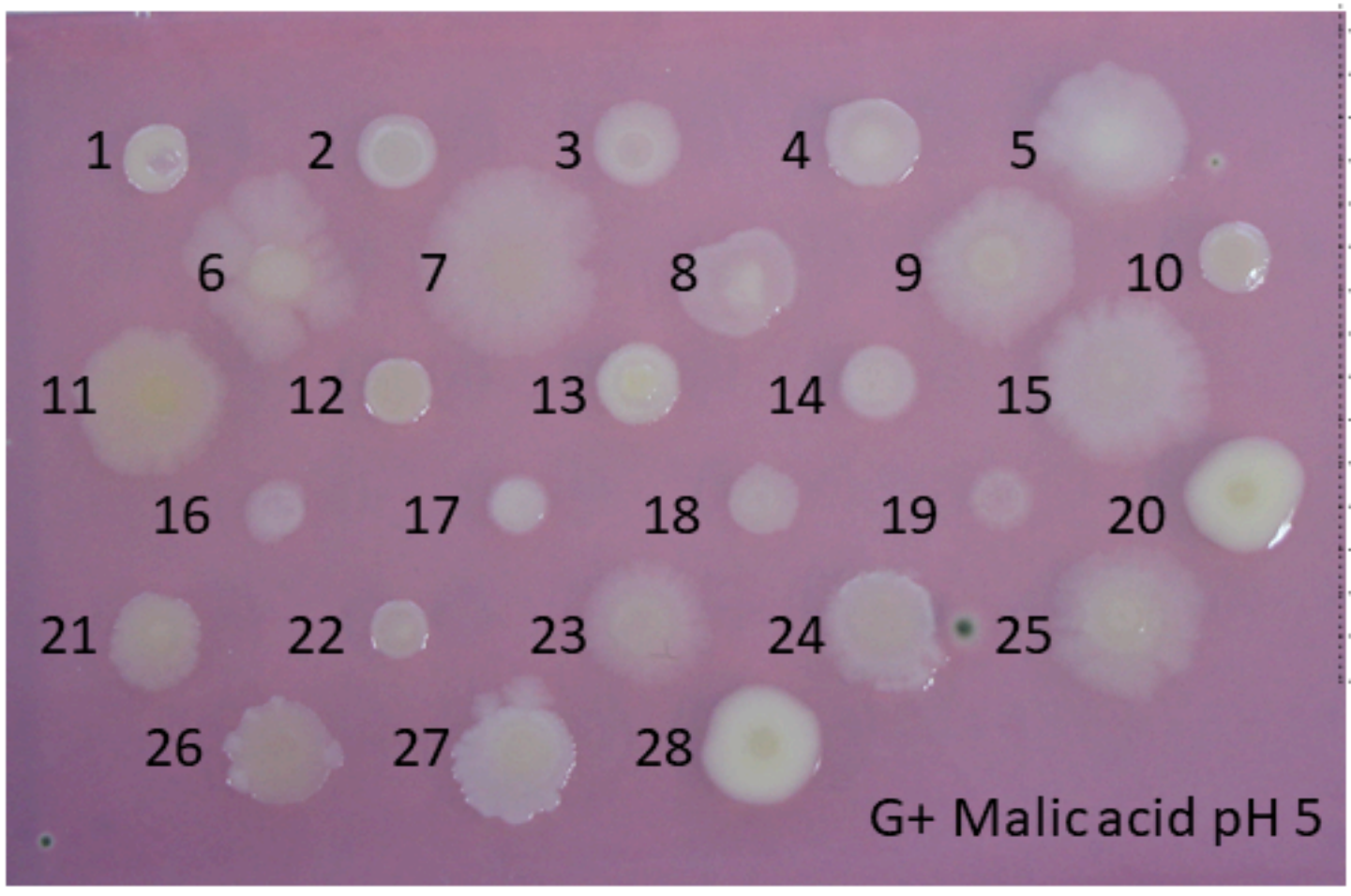
C


Lag time (min)
900
800
700
600
500
400
300
200
100
0
0
25
50
75
100
125
150
175
200
[H2O2] µM
Dickeya dadantii 3937
Dickeya aquatica 174/2

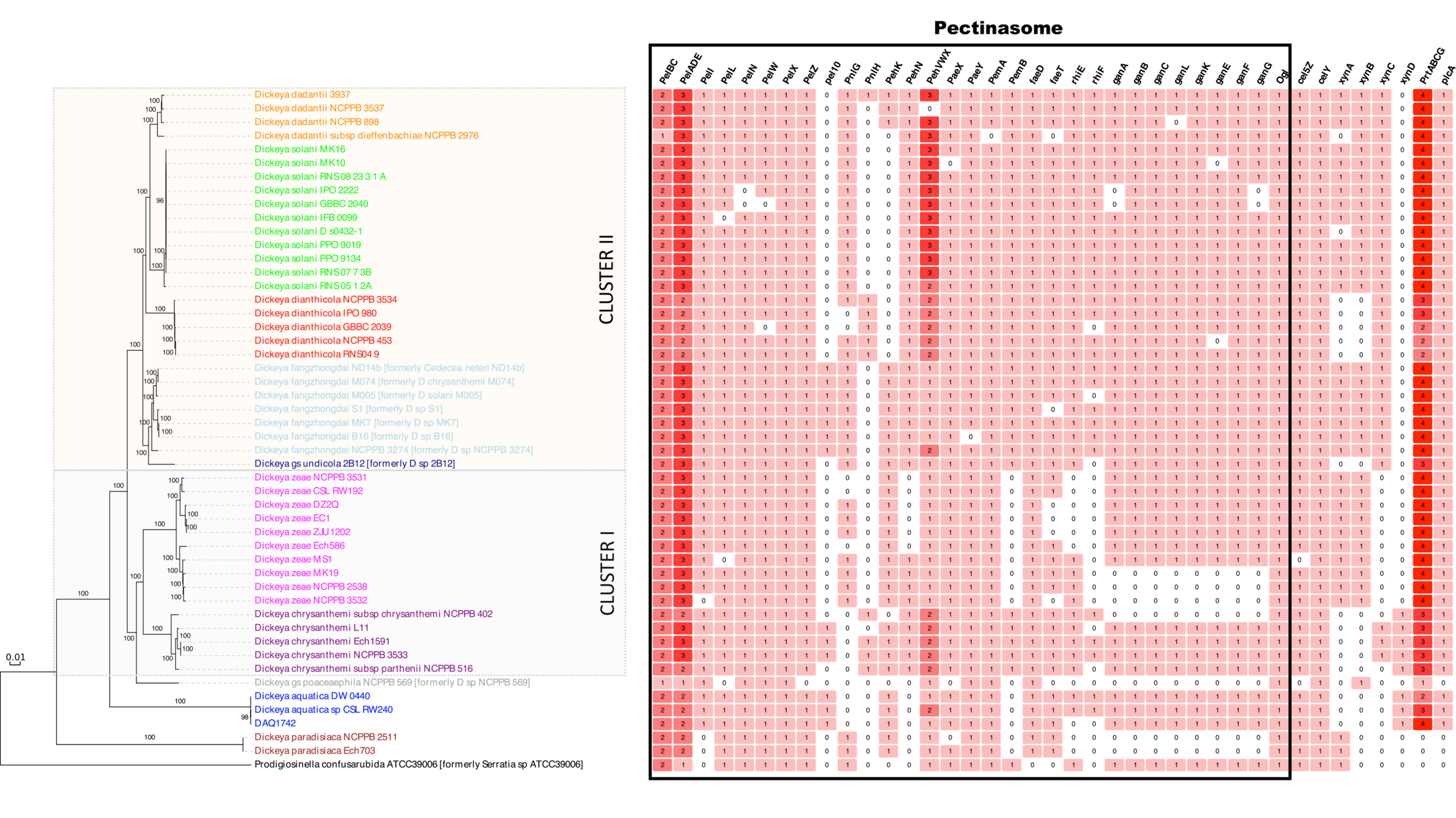
D

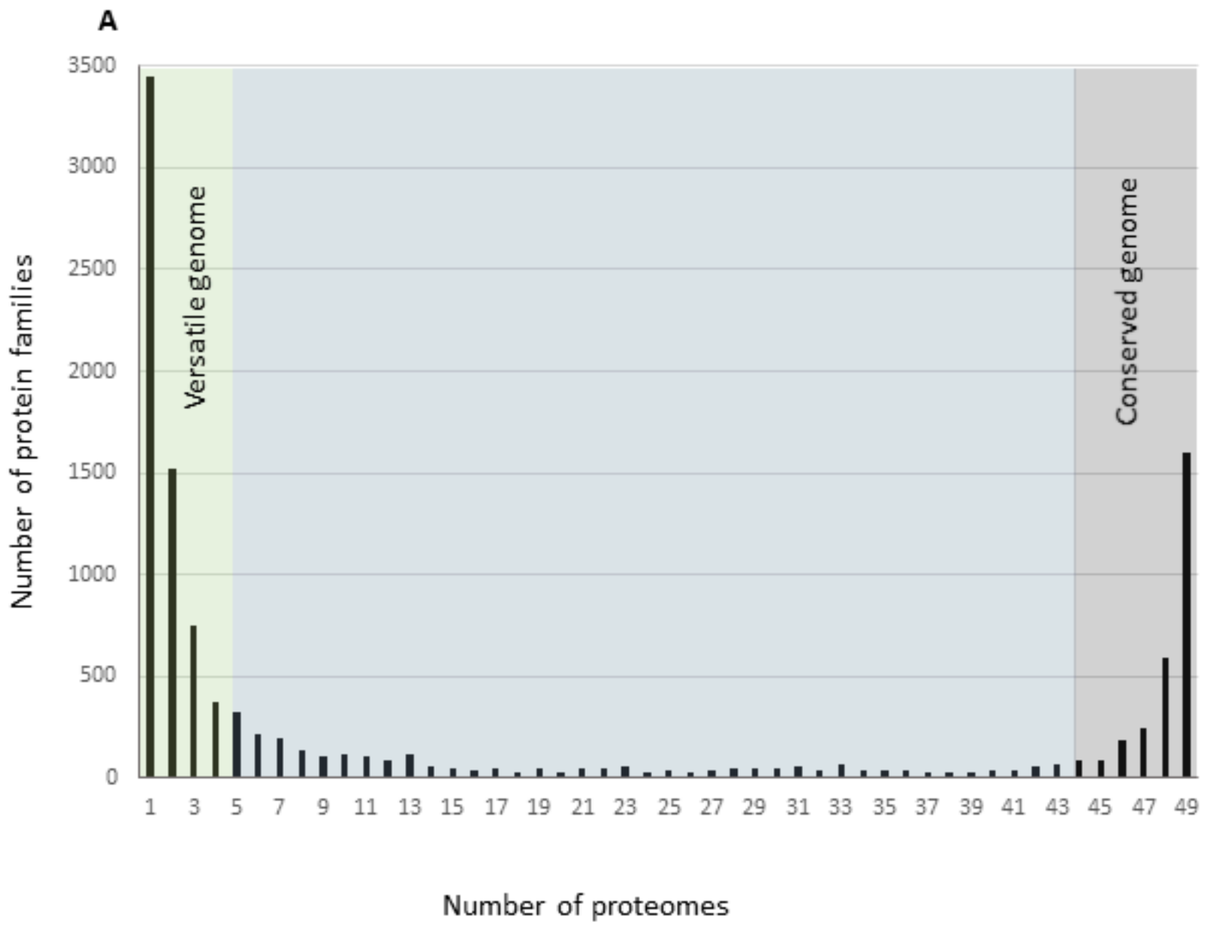

Growth rate (h-1)
0.35
0.30
0.25
0.2
0.15
0.10
0.05
0
3.5
4.5
5.5
6.5
pH
Dickeya dadantii 3937
Dickeya aquatica 174/2

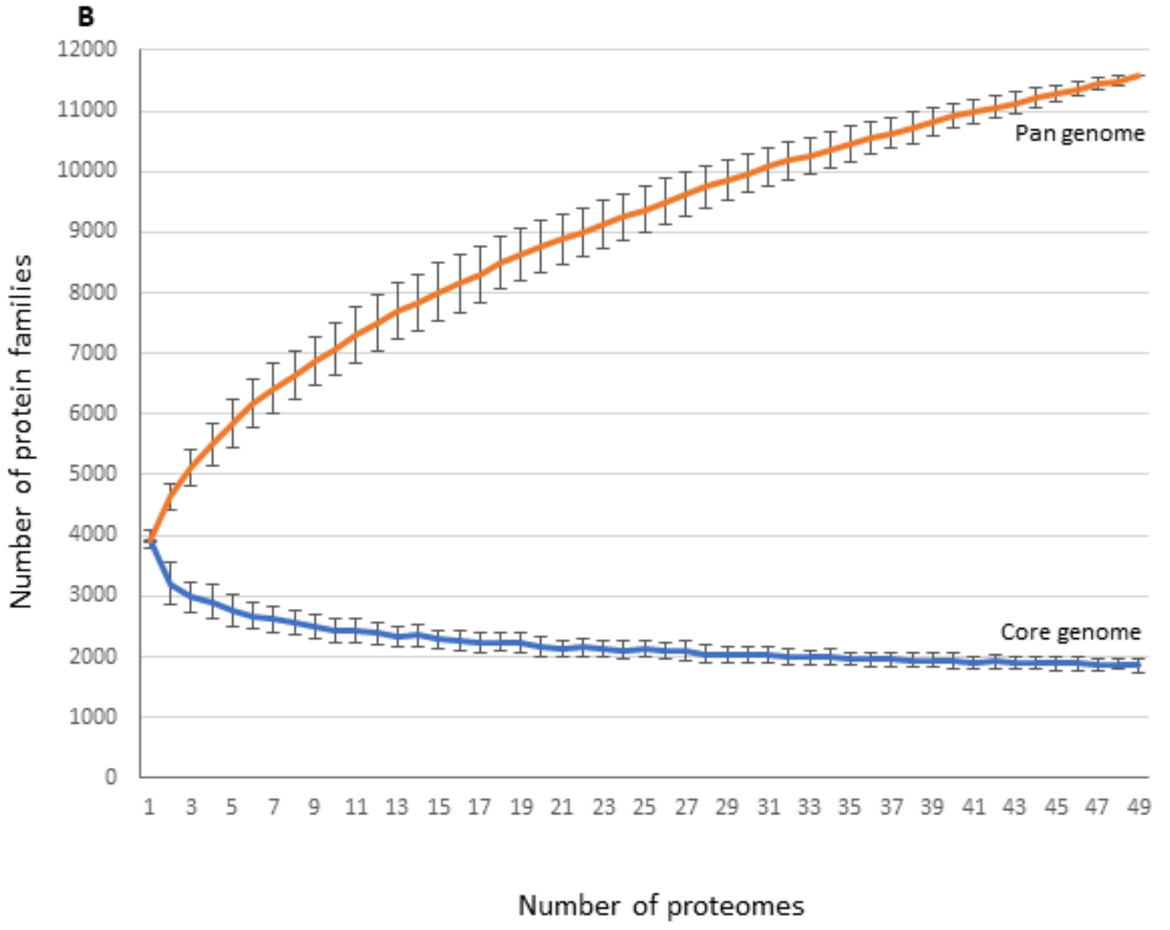


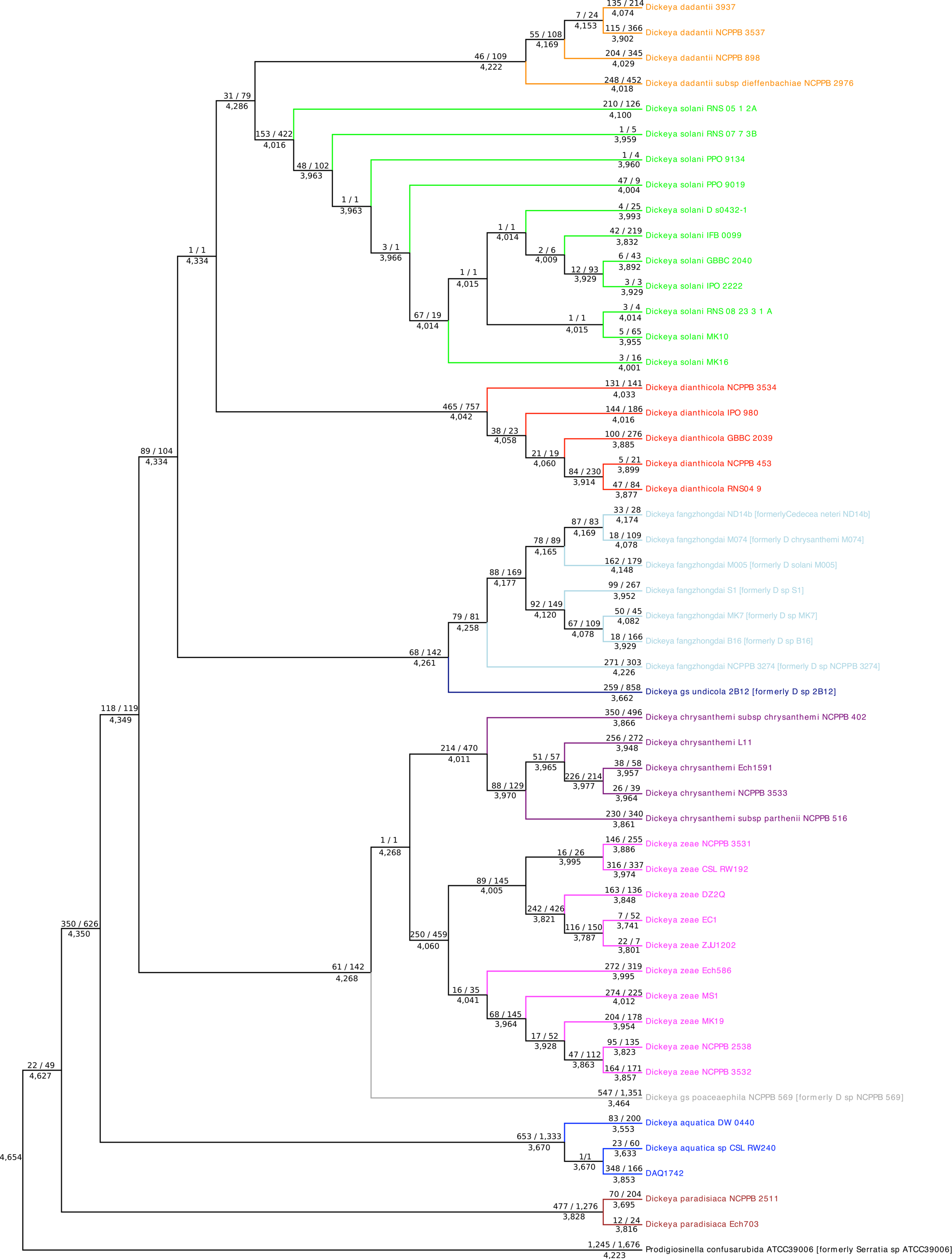


Colony	Strain	Twitching
1	D. dadantii 3937	-
2	D. dadantii 3937	-
3	D. dadantii NCPPB 3537	-
4	D. dadantii NCPPB 898	-
5	D. aquatica 174/2	+
6	D. fangzhongdai NCPPB 3274	+
7	D. fangzhongdai B16	+
8	D. fangzhongdai NCPPB 3274	+
9	D. undicola 2B12	+
10	D. poaceaphila NCPPB 569	-
11	D. zeae NCPPB 3532	+
12	D. zeae NCPPB 2547	-
13	D. zeae NCPPB 2538	-
14	D. chrysanthemi NCPPB 402	-
15	D. chrysanthemi NCPPB 3533	+
16	D. dianthicola RNS04.9	-
17	D. dianthicola CFBP 1888	-
18	D. dianthicola CFBP 2982	-
19	D. dianthicola NCPPB 453	-
20	D. paradisiaca NCPPB 2511	-
21	D. solani IPO2222	-
22	D. solani PP09019	-
23	D. solani RNS07.7.3B	+
24	D. solani GBBC2040	-
25	D. solani Ds0432-1	+
26	D. solani RNS05.1.2A	-
27	D. dadantii subsp. dieffenbachiae CFBP3694	-
28	D. dadantii subsp. dieffenbachiae NCPPB 2976	-


Pectinasome
PelBC
PelADE
PelI
PelL
PelN
PelW
PelX
PelZ
pel10
PnlG
PnlH
PehK
PehN
PehVWX
PaeX
PaeY
PemA
PemB
faeD
faeT
rhlE
rhlF
ganA
ganB
ganC
ganL
ganK
ganE
ganF
ganG
OgL
celSZ
celY
xynA
xynB
xynC
xynD
PrtABCG
plcA
CLUSTER II
CLUSTER I
Dickeya dadantii 3937
Dickeya dadantii NCPPB 3537
Dickeya dadantii NCPPB 898
Dickeya dadantii subsp dieffenbachiae NCPPB 2976
Dickeya solani MK16
Dickeya solani MK10
Dickeya solani RNS08 23 3 1 A
Dickeya solani IPO 2222
Dickeya solani GBBC 2040
Dickeya solani IFB 0099
Dickeya solani D s0432-1
Dickeya solani PPO 9019
Dickeya solani PPO 9134
Dickeya solani RNS07 7 3B
Dickeya solani RNS05 1 2A
Dickeya dianthicola NCPPB 3534
Dickeya dianthicola IPO 980
Dickeya dianthicola GBBC 2039
Dickeya dianthicola NCPPB 453
Dickeya dianthicola RNS04 9
Dickeya fangzhongdai ND14b [formerly Cedecea neteri ND14b]
Dickeya fangzhongdai M074 [formerly D chrysanthemi M074]
Dickeya fangzhongdai M005 [formerly D solani M005]
Dickeya fangzhongdai S1 [formerly D sp S1]
Dickeya fangzhongdai MK7 [formerly D sp MK7]
Dickeya fangzhongdai B16 [formerly D sp B16]
Dickeya fangzhongdai NCPPB 3274 [formerly D sp NCPPB 3274]
Dickeya gs undicola 2B12 [formerly D sp 2B12]
Dickeya zeae NCPPB 3531
Dickeya zeae CSL RW192
Dickeya zeae DZ2Q
Dickeya zeae EC1
Dickeya zeae ZJU1202
Dickeya zeae Ech586
Dickeya zeae MS1
Dickeya zeae MK19
Dickeya zeae NCPPB 2538
Dickeya zeae NCPPB 3532
Dickeya chrysanthemi subsp chrysanthemi NCPPB 402
Dickeya chrysanthemi L11
Dickeya chrysanthemi Ech1591
Dickeya chrysanthemi NCPPB 3533
Dickeya chrysanthemi subsp parthenii NCPPB 516
Dickeya gs poaceaephila NCPPB 569 [formerly D sp NCPPB 569]
Dickeya aquatica DW 0440
Dickeya aquatica sp CSL RW240
DAQ1742
Dickeya paradisiaca NCPPB 2511
Dickeya paradisiaca Ech703
Prodigiosinella confusarubida ATCC39006 [formerly Serratia sp ATCC39006]
0.01


A
Number of protein families
Number of proteomes
Versatile genome
Conserved genome
0
500
1000
1500
2000
2500
3000
3500
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49


B
Number of protein families
Number of proteomes
Pan genome
Core genome
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
12000
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49


135 / 214
4,074
Dickeya dadantii 3937
7 / 24
4,153
115 / 366
3,902
Dickeya dadantii NCPPB 3537
55 / 108
4,169
204 / 345
4,029
Dickeya dadantii NCPPB 898
46 / 109
4,222
248 / 452
4,018
Dickeya dadantii subsp dieffenbachiae NCPPB 2976
31 / 79
4,286
210 / 126
4,100
Dickeya solani RNS 05 1 2A
153 / 422
4,016
1 / 5
3,959
Dickeya solani RNS 07 7 3B
48 / 102
3,963
1 / 4
3,960
Dickeya solani PPO 9134
1 / 1
3,963
47 / 9
4,004
Dickeya solani PPO 9019
3 / 1
3,966
4 / 25
3,993
Dickeya solani D s0432-1
1 / 1
4,014
42 / 219
3,832
Dickeya solani IFB 0099
2 / 6
4,009
6 / 43
3,892
Dickeya solani GBBC 2040
12 / 93
3,929
3 / 3
3,929
Dickeya solani IPO 2222
1 / 1
4,334
1 / 1
4,015
67 / 19
4,014
3 / 4
4,014
Dickeya solani RNS 08 23 3 1 A
1 / 1
4,015
5 / 65
3,955
Dickeya solani MK10
3 / 16
4,001
Dickeya solani MK16
131 / 141
4,033
Dickeya dianthicola NCPPB 3534
465 / 757
4,042
144 / 186
4,016
Dickeya dianthicola IPO 980
38 / 23
4,058
100 / 276
3,885
Dickeya dianthicola GBBC 2039
89 / 104
4,334
21 / 19
4,060
5 / 21
3,899
Dickeya dianthicola NCPPB 453
84 / 230
3,914
47 / 84
3,877
Dickeya dianthicola RNS04 9
33 / 28
4,174
Dickeya fangzhongdai ND14b [formerlyCedecea neteri ND14b]
87 / 83
4,169
78 / 89
4,165
18 / 109
4,078
Dickeya fangzhongdai M074 [formerly D chrysanthemi M074]
162 / 179
4,148
Dickeya fangzhongdai M005 [formerly D solani M005]
88 / 169
4,177
99 / 267
3,952
Dickeya fangzhongdai S1 [formerly D sp S1]
92 / 149
4,120
50 / 45
4,082
Dickeya fangzhongdai MK7 [formerly D sp MK7]
79 / 81
4,258
67 / 109
4,078
18 / 166
3,929
Dickeya fangzhongdai B16 [formerly D sp B16]
68 / 142
4,261
271 / 303
4,226
Dickeya fangzhongdai NCPPB 3274 [formerly D sp NCPPB 3274]
259 / 858
3,662
Dickeya gs undicola 2B12 [formerly D sp 2B12]
118 / 119
4,349
350 / 496
3,866
Dickeya chrysanthemi subsp chrysanthemi NCPPB 402
214 / 470
4,011
256 / 272
3,948
Dickeya chrysanthemi L11
51 / 57
3,965
38 / 58
3,957
Dickeya chrysanthemi Ech1591
226 / 214
3,977
88 / 129
3,970
26 / 39
3,964
Dickeya chrysanthemi NCPPB 3533
230 / 340
3,861
Dickeya chrysanthemi subsp parthenii NCPPB 516
1 / 1
4,268
146 / 255
3,886
Dickeya zeae NCPPB 3531
16 / 26
3,995
316 / 337
3,974
Dickeya zeae CSL RW192
89 / 145
4,005
163 / 136
3,848
Dickeya zeae DZ2Q
350 / 626
4,350
242 / 426
3,821
7 / 52
3,741
Dickeya zeae EC1
116 / 150
3,787
250 / 459
4,060
22 / 7
3,801
Dickeya zeae ZJU1202
61 / 142
4,268
272 / 319
3,995
Dickeya zeae Ech586
16 / 35
4,041
274 / 225
4,012
Dickeya zeae MS1
68 / 145
3,964
204 / 178
3,954
Dickeya zeae MK19
17 / 52
3,928
95 / 135
3,823
Dickeya zeae NCPPB 2538
47 / 112
3,863
164 / 171
3,857
Dickeya zeae NCPPB 3532
22 / 49
4,627
547 / 1,351
3,464
Dickeya gs poaceaephila NCPPB 569 [formerly D sp NCPPB 569]
83 / 200
3,553
Dickeya aquatica DW 0440
653 / 1,333
3,670
23 / 60
3,633
Dickeya aquatica sp CSL RW240
4,654
1/1
3,670
348 / 166
3,853
DAQ1742
70 / 204
3,695
Dickeya paradisiaca NCPPB 2511
477 / 1,276
3,828
12 / 24
3,816
Dickeya paradisiaca Ech703
1,245 / 1,676
4,223
Prodigiosinella confusarubida ATCC39006 [formerly Serratia sp ATCC39006]