

25 **Originality-Significance statement**

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

37

38 **Summary**

39 *Dickeya* is a genus of phytopathogenic enterobacteriales causing soft rot in a variety of plants (e.g.
40 potato, chicory, maize). Among the species affiliated to this genus, *Dickeya aquatica*, described in
41 2014, remained particularly mysterious because it had no known host. Furthermore, while *D.*
42 *aquatica* was proposed to represent a deep-branching species among *Dickeya* genus, its precise
43 phylogenetic position remained elusive.

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

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

56

57 **Introduction**

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

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

81 species isolated from potatoes (Van der Wolf et al., 2014), *Dickeya aquatica* from freshwater rivers
82 (Parkinson et al., 2014), and *Dickeya fangzhongdai* from pear trees displaying symptoms of
83 bleeding canker in China (Tian et al., 2016). Thus, the *Dickeya* genus now comprises eight species
84 with distinctive phenotypic features (Supplementary Table S1). Virulence mechanisms have been
85 extensively studied in *Dickeya dadantii*. During infection, the bacterium enters the host using
86 natural openings and multiplies in intercellular spaces without causing major damage. Then, it
87 suddenly induces the production of aggression factors, such as pectinases, that break down the
88 plant cell wall pectin, causing the macroscopic symptom called soft rot (Reverchon and Nasser,
89 2013). *Dickeya* populations were initially considered to be restricted to tropical and subtropical
90 plant hosts and areas (Perombelon, 1990). This assumption was called into question following the
91 identification of *D. dianthicola* strains from potato plants in Western Europe (Janse and Ruissen,
92 1988) and the first isolation of *D. solani* strains in France, Finland, Poland, the Netherlands, and
93 Israel (Czajkowski et al., 2011; van der Wolf et al., 2014). It was postulated that the pathogen was
94 introduced in Europe via the international trade of potato seeds (van der Wolf et al., 2014). Strains
95 of *D. solani* have also been found in hyacinth, which was interpreted as a possible recent transfer
96 from hyacinth to potato, possibly via contaminated irrigation waters (Sławiak *et al.*, 2009). This
97 highlights the high capacity of adaptation and dissemination of *Dickeya* to new geographic areas
98 and to new hosts.

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

109 species (Zhang et al., 2016). *D. fangzhongdai* branched at the base of the Cluster II in a tree
110 based on seven housekeeping genes (Tian et al., 2016). Finally, according to a tree based on six
111 housekeeping genes, *D. aquatica* emerge just after the divergence of *D. paradisiaca* but before the
112 divergence of Cluster I and Cluster II (Parkinson et al., 2014). Yet, the associated bootstrap values
113 were weak (Parkinson et al., 2014), meaning that the relative order of emergence of *D. paradisiaca*
114 and *D. aquatica* remained to determine. Among *Dickeya*, *D. aquatica* is remarkable because all
115 the three known strains 174/2^T, 181/2, and Dw0440 have been isolated from waterways and, in
116 contrast to other species, have no known vegetal host (Supplementary Table S1) (Parkinson et al.,
117 2014; Ma et al., 2007). These features and its early-branching position within *Dickeya* make *D.*
118 *aquatica* very interesting to study the origin and the early steps of the diversification of *Dickeya*,
119 and in particular, the emergence of their virulence factors and capacity to infect plants.
120 These important questions were the focus of this study. For this purpose, we sequence the
121 complete genome of the *D. aquatica* 174/2 type strain. Based on the phylogenetic analysis of
122 1,341 single copy core protein families, we show that *D. aquatica* represents the second deepest
123 branching lineage within *Dickeya*, reclassify two lineages that likely correspond to new
124 genomospecies (gs.): *Dickeya* gs. poaceaephila and *Dickeya* gs. undicola, and identify a new
125 lineage, tentatively called *Prodigiosinella*, that likely represents the closest relative of *Dickeya*. We
126 also highlight the presence of different virulence factors, suggesting that *D. aquatica* 174/2 could
127 be pathogenic. Stimulated by these findings, we explore the pathogenic potential of *D. aquatica*
128 174/2. Surprisingly, we identify the acidic fruits tomato (pH 4.8) and cucumber (pH 5.1) as potential
129 hosts for *D. aquatica*, reflecting the potential of this aquatic species to flexibly adapt to a new
130 ecological niche provided by acidic fruits with a high water content. Accordingly, we show that this
131 strain displays a specific induction of twitching motility under acidic conditions. Finally, using
132 phylogenomic approaches, we trace back the origin and evolution of key factors associated with
133 virulence and host specificity in *Dickeya*, including *D. aquatica*. From this study we infer the
134 evolution of gene repertoires along the diversification of *Dickeya* and highlight a remarkable
135 general tendency toward proteome reduction in all *Dickeya* species.

136

137 **Results and discussion**

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

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

159

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

161 The massive release of *Dickeya* genomic sequences (7 complete genomes and 41 draft genomes)
162 in public databases (Supplementary Table S3), in addition to the complete genome of *D. aquatica*
163 174/2^T reported in this study, provided an interesting resource to explore the evolutionary history of
164 this genus. We investigated the phylogeny of *Dickeya* using 51 ribosomal proteins (rprots), as

165 these were shown to be well suited to study the systematics of prokaryotes, especially
166 *Proteobacteria* (Yutin et al., 2012; Ramulu et al., 2014), on the one hand, and 1,341 core protein
167 families (core-pf), on the other hand. The maximum likelihood (ML) rprots tree, rooted with
168 *Pectobacterium* (*Pectobacteriaceae*) and *Serratia* (*Yersiniaceae*), recovered the monophyly of
169 *Dickeya* (Bootstrap Value (BV) = 100%, Supplementary Figure S1). Surprisingly, *Serratia* strains,
170 used as outgroup to root the phylogeny of *Dickeya* together with *Pectobacterium*, did not form a
171 monophyletic group, due to the robust clustering of *Serratia* sp. ATCC 39006 with *Dickeya* (BV =
172 99%). A similar grouping was also observed in trees based on the RNA components of the small
173 and large subunits of the ribosome (SSU and LSU rRNA, respectively) (Supplementary Figure S2),
174 suggesting that strain ATCC 39006 represents the sister-lineage of *Dickeya*, and was wrongly
175 affiliated to the *Serratia*, based on its capacity to synthesize prodigiosin, a red pigment secondary
176 metabolite with antimicrobial, anticancer, and immunosuppressant properties, characteristic of
177 *Serratia* species (Thomson et al., 2000). The sister-relationship of ATCC 39006 and *Dickeya* is
178 consistent with the close evolutionary link between the pectinolytic gene clusters (*pel*, *paeY*, *pemA*)
179 shared by these taxa (Duprey et al., 2016b), and with the presence, in *Serratia* sp. ATCC 39006, of
180 an homologue of the Vfm quorum sensing system previously reported as specific to *Dickeya*
181 species (Supplementary Table S4 and Figure S3) (Nasser et al., 2013). The phylogenetic analysis
182 of the genes involved in the biosynthesis of the prodigiosin showed that strain ATCC 39006 genes
183 emerge in a cluster gathering sequences from unrelated taxa: *Serratia* (*Yersiniaceae*), *Hahella*
184 (*Hahellaceae*), and *Streptomyces* (*Actinobacteria*), suggesting that the prodigiosin gene cluster
185 spread among these lineages (including ATCC 39006 strain) through horizontal gene transfer
186 (HGT) (Supplementary Figure S3). Interestingly, strain ATCC 39006, like the early-diverging *D.*
187 *paradisiaca* species, was deprived of the *indABC* gene cluster responsible for the characteristic
188 production of blue-pigmented indigoidine in *Dickeya* spp. (Supplementary Table S4) (Reverchon et
189 al., 2002; Lee and Yu, 2006). Phylogenetic analyses showed that *Dickeya* IndA, IndB, and IndC
190 proteins are closed to sequences from unrelated bacteria (Supplementary Figure S3), suggesting
191 that indigoidine genes underwent HGT and have been secondarily acquired in *Dickeya* after the
192 divergence with *D. paradisiaca*. The evolutionary link between strain ATCC 39006 and *Dickeya* is

193 also consistent with comparison of their proteomes. In fact, strain ATCC 39006 share in average
194 more proteins families with *Dickeya* (2,554/4,221 – 60.5%), compared to *Pectobacterium*
195 (2,387/4,221 - 56.5%) or other *Serratia* species (2,162/4,221 - 51.2%) (Supplementary Table S5).
196 Yet, the large evolutionary distance between ATCC 39006 strain and *Dickeya* (Supplementary
197 Figure S1), its lower GC content (49.2% vs 52.6 - 56.9% in *Dickeya*) (Supplementary Table S3),
198 and the lower number of protein families shared between ATCC 39006 and *Dickeya* (60.5%)
199 compared to between *Dickeya* species (64.7% - 89.6%, Supplementary Table S5), suggested that
200 strain ATCC 39006 does not belong to *Dickeya* and represents rather a close but distinct lineage,
201 possibly a new genus, that we propose to call *Prodigiosinella*, with strain ATCC 39006 being
202 reclassified as *Prodigiosinella confusarubida*. By using leBIBI^{QBPP} phylogenetic SSU rRNA-based
203 positioning tool (Flandrois et al., 2015), we detected two *Erwinia* sp. Strains: MK01 (SSU
204 accession number at the NCBI: AY690711.1) and MK09 (SSU accession number at the NCBI:
205 AY690717.1), isolated from the rhizosphere of *Phragmites communis*, that were very closely
206 related to ATCC 39006 and could belong to *P. confusarubida*.

207 According to these results and to decrease computation time, the core-pf phylogeny was inferred
208 using *P. confusarubida* ATCC 39006 as outgroup. The ML core-pf and rprots trees displayed
209 similar topologies (Figure 2 and Supplementary Figure S1), excepted regarding *D. chrysanthemi*,
210 *D. sp.* NCPPB 569, and *D. solani* (see below). Worth to note, the rprots tree was overall less
211 resolved than the core-pf tree (i.e. the former displayed lower BV than the latter). Both trees
212 supported the monophyly of *D. paradisiaca*, *D. aquatica*, *D. zaeae*, *D. dianthicola*, *D. dadantii*, and
213 *D. solani* (BV ≥ 84%). The monophyly of *D. chrysanthemi* was recovered with core-pf (BV = 100%)
214 but not with rprots. In fact, while the relationships among *D. chrysanthemi* strains were strongly
215 supported in the core-pf tree, they were mostly unresolved (i.e. associated to weak BV) with rprots,
216 meaning that these markers did not contain enough phylogenetic signal to resolved the *D.*
217 *chrysanthemi* phylogeny. A composite group composed of strains annotated as *D. sp.*, *D. solani*,
218 and *D. chrysanthemi*: M074, M005, B16, MK7, S1, NCPPB 3274, and ND14b (formerly
219 misidentified as *Cedecea neteri* and transferred recently to *D. solani*) was present in both trees (BV
220 ≥ 95%). By using leBIBI^{QBPP} tool (Flandrois et al., 2015), we identified strains B16, MK7, and S1 as

221 *D. fangzhongdai*, suggesting that the whole clade could correspond to *D. fangzhongdai* (Alič et al.,
222 2018).

223 Cluster I (*D. zea* and *D. chrysanthemi*) and cluster II (*D. dianthicola*, *D. solani*, *D. dadantii*, *D.*
224 *fangzhongdai* and *D. undicola*) were monophyletic (core-pf: both BV \geq 100% and rprots: both BV \geq
225 80%, respectively). Within cluster II, *D. fangzhongdai* diverged first (BV = 100%), while *D. solani*
226 represented either the sister-group of *D. dadantii* (core-pf: BV = 100%) or *D. dianthicola* (rprots: BV
227 = 98%). The conflicting position of *D. solani* was supported by high BV in both trees, indicating a
228 real inconsistency between the two trees. To go further, AU tests performed on individual core-pf
229 and rprots by using the core-pf and the rprots ML topologies (Supplementary Table S6). These
230 tests showed that 1,063 out of 1,341 (79.3%) core-pf protein families reject the rprots topology,
231 while 853 out of 1,341 (63.6%) core-pf protein families do not reject the core-pf topology. In
232 contrast, 31.4% of the rprots do not reject the core-pf topology and 29.4% do not reject the rprots
233 topology. Thus, a large majority of core-pf protein families favour the core-pf topology, while an
234 equal number of rprots supports either one or the other topology. Accordingly, the rprots topology
235 and in particular the grouping of *D. solani* with *D. dianthicola* could be questioned.

236 Regarding the first divergences within *Dickeya*, both trees pinpointed *D. paradisiaca* and *D.*
237 *aquatica*, as early diverging lineages (core-pf: BV = 100%, rprots: BV = 86%), with *D. paradisiaca*
238 emerging first (core-pf: BV = 100%, rprots: BV = 65%). Finally, the core-pf placed robustly *D. sp.*
239 NCPPB 569, a strain isolated from sugarcane in Australia (Supplementary Table S3), at the base
240 of Cluster I (BV = 100%), while its position was unresolved in the rprots tree (i.e. unsupported
241 grouping with *D. aquatica*, BV = 46%). Irrespective of its position, the evolutionary distances
242 between strain NCPPB 569 and the eight *Dickeya* species were roughly similar, suggesting
243 NCPPB 569 could correspond to an independent lineage within *Dickeya*. This observation was in
244 agreement with a previous work based on the phylogenetic analysis of *recA* (Parkinson et al.
245 2009), and led us to propose that strain NCPPB 569 could represent a distinct genomospecies, we
246 proposed to name *Dickeya* gs. *poaceaephila*. This hypothesis was strengthened by the fact that
247 strain NCPPB 569 could be distinguished from other species according to shared protein families
248 (Figure 2 and Supplementary Figure S4). Among cluster II, *Dickeya* sp. 2B12 deserved attention.

249 In fact, this strain branched as a distant sister-lineage of *D. fangzhongdai* (core-pf: BV = 100% and
250 rprots: BV = 95%). This suggested that this strain could also represent a new genomospecies that
251 we have tentatively called *Dickeya* gs. *undicola*.

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

271

272 ***Virulence-associated phenotypes of D. aquatica***

273 While most *Dickeya* species have known vegetal hosts, all the identified strains of *D. aquatica*
274 (174/2^T, 181/2 and Dw0440) have been isolated from waterways (Supplementary Table S1)
275 (Parkinson et al., 2014). To our knowledge, their pathogenicity has never been demonstrated. The
276 *in silico* survey of the *D. aquatica* 174/2^T genome for transcriptional regulator binding sites revealed

277 the presence of conserved key regulators of virulence (Supplementary Table S7). More precisely,
278 all the regulators (KdgR, PecS, PecT, CRP, Fis, H-NS, GacA-GacS, RsmA-RsmB, MfbR)
279 controlling virulence gene expression in *D. dadantii* (Reverchon et al., 2016) are present in *D.*
280 *aquatica* 174/2^T. The global regulators FNR and CRP displayed the highest number (approximately
281 80) of targets, followed by Fur, CpxR, ArcA, Lrp and KdgR each with more than 30 targets. The
282 activator CRP and the repressor KdgR have a key role in *Dickeya* virulence (Duprey et al., 2016a),
283 as they tightly control the pectin degradation pathway and are involved in coupling central
284 metabolism to pectinase gene expression (Nasser et al., 1997), while Fur acts as a repressor of
285 pectinase genes (Franza et al., 2002). In addition, Fur represses genes involved in the metabolism
286 of iron, a metal that plays an important role as a virulence regulatory signal in *Dickeya* (Franza et
287 al., 2002).

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

304

305 ***Stress resistance phenotypes of D. aquatica***

306 **Osmotic stress**

307 Depending on the plant host, *Dickeya* species encounter various stresses during the infectious
308 process (Reverchon et al., 2016). We therefore assessed stress resistance of *D. aquatica* 174/2^T
309 (Figure 4). This strain proved to be very sensitive to osmotic stress, as it displayed a 50% growth
310 rate reduction on 0.3 M NaCl while *D. dadantii* was only slightly affected (20% growth rate
311 reduction). This effect was even more pronounced on 0.5 M NaCl with a growth rate reduction of
312 90 % for *D. aquatica* and 43% for *D. dadantii* (Figure 4A). This sensitivity to osmotic stress could
313 be linked to the lack of two osmoprotectant biosynthetic pathways in *D. aquatica*: glycine betaine
314 biosynthesis (*betA-betB* gene cluster) and phosphoglycosyl glycerate biosynthesis (*pggS-pggP*
315 gene cluster) (Jiang et al., 2016) (Supplementary Table S4). The *pggSP* gene cluster was present
316 in *P. confusarubida* and all *Dickeya* species except *D. aquatica*. The phylogeny of the two genes
317 was consistent with the phylogeny of species (Supplementary Figure S3), indicating that their
318 absence in *D. aquatica* results from a specific and secondary loss in this taxon. In contrast, the
319 *betABI* gene cluster was absent in *D. paradisiaca*, *D. aquatica*, *D. dianthicola* and Cluster I species
320 (Supplementary Table S4). Their phylogenies suggested that the *betABI* gene cluster could have
321 been present in the ancestor of *P. confusarubida* and *Dickeya*, and secondarily and independently
322 lost in the *Dickeya* lineages mentioned above (Supplementary Figure S3). Finally, the *ousA* gene
323 encoding the major osmoprotectant uptake system in *D. dadantii* is missing in *D. aquatica*, as well
324 as in most other *Dickeya* species, suggesting either a recent acquisition by a few species or
325 multiple losses of an ancestral system (Supplementary Table S4). Interestingly, *D. solani* and *D.*
326 *dianthicola*, both lacking the *ousA* gene, are also more sensitive to osmotic stress than *D. dadantii*
327 (Figure 4B).

328 **Oxidative stress**

329 *D. aquatica* 174/2^T was also very sensitive to oxidative stress, displaying longer lag time than *D.*
330 *dadantii* in presence of 75 μ M H₂O₂ (Figure 4C). This is consistent with the lack of the periplasmic
331 superoxide dismutase SodC, whose gene was likely acquired in *Dickeya* after the divergence of *D.*
332 *paradisiaca* and *D. aquatica* (Supplementary Figure S3), and the lack of the Fe-S cluster assembly

333 SUF system (Supplementary Table S4). Iron-sulfur (Fe-S) clusters are fundamental to numerous
334 biological processes in most organisms, but these protein cofactors can be prone to damage by
335 various oxidants (e.g., O₂, reactive oxygen species, and reactive nitrogen species) (Roche et al.,
336 2013). In addition, the release of free iron in the bacterial cytoplasm amplifies the oxidative stress
337 through the Fenton reaction, producing the highly toxic and reactive hydroxyl radical OH[•]. Most
338 gammaproteobacteria have two Fe-S cluster biogenesis systems SUF and ISC (Roche et al.,
339 2013). For example, *Escherichia coli* cells switch from the ISC to the SUF system under oxidative
340 stress as OxyR, a sensor of oxidative stress, acts as an activator of *suf* operon expression (Lee et
341 al., 2004). Both the SUF and ISC systems were shown to be required for *D. dadantii* growth in the
342 plant host environment that is continually changing in terms of iron availability and redox conditions
343 (Nachin et al., 2001; Expert et al., 2008; Rincon-Enriquez et al., 2008). While both the ISC and
344 SUF systems were likely present in the ancestor of *Dickeya* and *P. confusarubida*, phylogeny and
345 taxonomic distribution suggested that SUF genes were secondarily lost in *D. paradisiaca*, *D.*
346 *aquatica*, *D. gs. poaceaephila*, and *D. zae* (Supplementary Figure S3). The absence of both the
347 SUF system and SodC in *D. aquatica* could be an important factor limiting its host range
348 (Supplementary Figure S3). Finally, the presence of the DNA-binding protein Dps, which has a
349 protective function against a wide range of stresses, in *P. confusarubida* and all *Dickeya* excepted
350 *D. paradisiaca* suggested a specific loss in this latter lineage (Supplementary Figure S3).

351 Regarding other oxidative stress sources, the NorWVR system responsible for nitrite oxide
352 detoxification was absent in the first diverging *Dickeya* species (Supplementary Table S4). The
353 phylogenies of the corresponding proteins suggested secondary losses in *P. confusarubida*, *D.*
354 *paradisiaca*, and *D. aquatica*, as these proteins were present in *Pectobacterium*, a closely related
355 genus of *Dickeya* (Supplementary Figure S3). Finally, the ascorbate degradation pathway encoded
356 by the *ula* genes was present in the *D. paradisiaca*, *D. aquatica*, *D. zae*, and *D. chrysanthemi*
357 species, suggesting that the pathway was also present in the ancestor of *Dickeya* and secondarily
358 lost in *D. gs. poaceaephila* and in the ancestor of Cluster II (Supplementary Figure S3). Ascorbic
359 acid, the major and probably the only antioxidant buffer in the plant apoplast, becomes oxidised
360 during pathogen attack (Pignocchi and Foyer, 2003). Modification of the apoplastic redox state

361 modulates receptor activity and signal transduction to regulate plant defence and growth
362 (Pignocchi and Foyer, 2003). The capacity of some *Dickeya* species to catabolize ascorbate could
363 be a strategy to weaken the plant defence.

364 **Acidic stress**

365 Regarding pH sensitivity, both *D. aquatica* 174/2^T and *D. dadantii* showed optimum growth at pH
366 7.0. *D. aquatica* 174/2^T displayed no significant growth rate reduction down to pH 5.3 while *D.*
367 *dadantii* was slightly affected (20% growth rate reduction at pH 5.3 compared to optimum pH 7.0)
368 (Figure 4D). Below pH 4.9, the growth of *D. aquatica* abruptly diminished compared to that of *D.*
369 *dadantii* indicating that *D. aquatica* growth was much more impacted by low pH (Figure 4D).
370 Bacterial response to acid stress involves evasion of cell damage and adaptation of the enzymatic
371 profile by reducing reactions producing protons and promoting those consuming protons (Bearson
372 et al., 1997). Concerning the protection of proteins against pH damage, *Dickeya* strains, including
373 *D. aquatica* strain 174/2^T, encode the lysine-rich protein Asr acid-inducible periplasmic chaperone,
374 known to protect the proteins by sequestering protons due to its high basic amino acids
375 composition (Seputiene et al., 2003). In *Dickeya*, the corresponding gene is located downstream
376 the *rstAB* gene cluster, known to control the *asr* gene expression (Ogasawara et al., 2007). Yet,
377 due to their atypical amino acid composition these proteins were often wrongly annotated as
378 histone proteins (e.g. in *D. dadantii* NCPPB 898, *D. dianthicola* NCPPB 3534, *D. dianthicola*
379 NCPPB 453, *D. dianthicola* GBBC 2039, and *D. chrysanthemi* ATCC 11663). In terms of metabolic
380 adaptation, a major strategy for bacterial pH homeostasis consists in using decarboxylases to
381 remove cytoplasmic protons (Krulwich et al., 2011). The glutamate-, arginine-, and lysine-inducible
382 decarboxylases classically found in enteric bacteria (Foster, 2004) are absent in *Dickeya* species.
383 However they contain some organic acid decarboxylases such as oxalate and malonate
384 decarboxylases (Supplementary Table S4). Some *Dickeya* species, including *D. dadantii*, have two
385 oxalate decarboxylation pathways: the *frc-oxc* pathway, being involved in the acid tolerance
386 response in *E. coli* (Fontenot et al., 2013) and the OxdD decarboxylase pathway. Most *Dickeya*
387 species harbour at least one of these oxalate decarboxylation pathways, except *D. aquatica*, *D. gs.*
388 *poaceaephila*, *D. chrysanthemi*, and a few *D. zeae* strains, which were deprived of both pathways

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

396 **Twitching motility**

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

417 *Dickeya* pathogenicity is therefore an interesting question for the future studies.

418 **Antimicrobial peptides**

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

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

433

434 ***Distribution of plant cell wall degrading enzymes in Dickeya***

435 The plant cell wall is a complex and dynamic meshwork of polymers (cellulose, hemicellulose,
436 pectin, structural glycoproteins) (Pauly and Keegstra, 2016). Among these polymers, pectin is the
437 most complex and includes both linear regions composed of polygalacturonan and ramified regions
438 (RGI and RGII, respectively). RGI contains a rhamnogalacturonan backbone and various lateral
439 chains such as galactan, arabinan and galacturonan (Caffall and Mohnen, 2009). RGII contains a
440 short galacturonan backbone, carrying four side chains, with a diversity of rare monosaccharides
441 (O'Neill et al., 2004). The carboxylic groups of D-galacturonate residues are methyl-esterified to
442 various degrees (up to 80%) and these residues are, to a lesser extent, acetylated at the C2 and/or
443 C3 positions. Feruloyl esters are a type of modification commonly found in arabinan and galactan
444 chains of ramified regions (Ishii, 1997). The virulence of *Dickeya* is correlated with their ability to

445 synthesize and secrete plant cell wall degrading enzymes, including a full set of pectinases
446 (Hugouvieux-Cotte-Pattat et al., 2014), xylanases and xylosidases (Keen et al., 1996) proteases
447 PrtA, PrtB, PrtC, PrtG (Wandersman et al., 1987), and the cellulase Cel5Z (Py et al., 1991). The
448 presence of these virulence factors varies depending on the species (Matsumoto et al., 2003;
449 Duprey et al., 2016b). To obtain a comprehensive view of this phenomenon, we explored the “Plant
450 cell wall degradosome” of *Dickeya* species including *D. aquatica* (**Figure 6**).

451

452 **The pectinasome**

453 The *Dickeya* pectinasome includes multiple pectate lyases (PelA, PelB, PelC, PelD, PelE, PelL,
454 PelL, PelN, PelW, PelX, PelZ, Pel10), pectin lyases (PnlG, PnlH), polygalacturonases (PehK,
455 PehN, PehV, PehW, PehX), pectin methyl esterases (PemA, PemB), pectin acetyl esterases
456 (PaeX, PaeY), feruloyl esterases (FaeD, FaeT), rhamnogalacturonate lyases (RhiE, RhiF), and
457 one periplasmic endogalactanase (GanA) (Figure 6) (Hugouvieux-Cotte-Pattat et al., 2014). *D. gs.*
458 *poaceaephila* NCPPB 569 was the genospecies with the poorest pectinase content. According to
459 their taxonomic distribution and phylogeny, most of these proteins could be inferred in the ancestor
460 of *Dickeya* and in its sister lineage *P. confusarubida*. Regarding the *pelAED* cluster, while *P.*
461 *confusarubida* has a single gene, the ancestor of *Dickeya* had two copies, *pelA* and an
462 undifferentiated *pelDE* pectate lyase-coding gene (Duprey et al., 2016b), suggesting that a
463 duplication event occurred in the stem of *Dickeya*. This undifferentiated *pelDE* has undergone a
464 duplication event after the emergence of *D. paradisiaca*, giving rise to *pelD* and *pelE*
465 (Supplementary Figure S5) (Duprey et al., 2016b). *D. aquatica* has then lost *pelE*, while *D.*
466 *poaceaephila* has lost both *pelA* and *pelE*. The *pelA* gene was also lost in *D. dianthicola*, while
467 *pelE* was lost in some *D. chrysanthemi* strains (Figure 6, Supplementary Figure S5). The *pelBC*
468 cluster was also likely present in the *Dickeya* ancestor, and conserved in most *Dickeya* species.
469 However, *pelB* was lost in *D. gs. poaceaephila*, and *D. dadantii subsp. dieffenbachiae* NCPPB
470 2976 (Supplementary Figure S3). The pectate lyase Pell was present in *Pectobacterium* and in all
471 *Dickeya*, except *D. paradisiaca*, and in *P. confusarubida*, suggesting that it could be ancestral in
472 *Dickeya*. Similarly, the phylogeny of pectin methyl esterase PemB indicated that the protein was

473 present in *P. confusarubida* and *Pectobacterium*, suggesting an ancestral presence in *Dickeya*.
474 Accordingly, its absence in *D. paradisiaca*, *D. aquatica*, *D. gs. poaceaephila*, and *D. zea* strains
475 likely reflected secondary losses. The phylogeny of this protein also suggested that the *pemB* gene
476 found in *D. chrysanthemi* was acquired by HGT from Cluster II species (Supplementary Figure S3).
477 Regarding the polygalacturonase PehN, a similar scenario could be inferred except that the
478 corresponding gene was also lost in *P. confusarubida*, and that a few *D. zea* strains seemed to
479 have reacquired a PehN from Cluster II species by HGT (Supplementary Figure S3). Regarding
480 the *pehVWX* cluster, the three genes were derived from a single *pehX* gene that was present in the
481 ancestor of *P. confusarubida* and *Dickeya*. A gene duplication event occurred in the ancestor of *D.*
482 *dianthicola*, *D. solani* and *D. dadantii* leading to PehV. A second event led to the divergence of
483 PehW in the ancestor of *D. solani* and *D. dadantii* (Supplementary Figure S5). The
484 polygalacturonase PehK was likely present in the ancestor of all *Dickeya* and secondarily lost in *D.*
485 *gs. poaceaephila*, *D. solani*, and *D. dianthicola* (Supplementary Figure S3). The pectin lyase PnlG
486 was likely present in the ancestor shared by *Dickeya* and *P. confusarubida*, and secondarily lost in
487 *D. aquatica*, *D. gs. poaceaephila*, *D. chrysanthemi*, *D. dianthicola* and in some *D. zea* strains
488 (Supplementary Figure S3). Finally, the rare pectate lyase Pel10 and pectin lyase PnlH displayed
489 patchy taxonomic distributions (Figure 6). Their phylogenies suggested that they spread through
490 HGT in *Dickeya* (Supplementary Figure S3).
491 The saturated and unsaturated digalacturonates resulting from pectin degradation by pectinases
492 are converted into monogalacturonate and 5-keto-4-deoxyuronate by the oligogalacturonate lyase
493 Ogl, which is present in all *Dickeya* species (Figure 6). The phylogenies of the *gan* gene cluster
494 responsible for degradation of galactan chains in pectin-ramified regions and the
495 rhamnogalacturonate lyase RhiE involved in degradation of RGI pectin-ramified regions indicated
496 they were likely present in the ancestor of *P. confusarubida* and *Dickeya*, and secondarily lost in
497 the basal *D. paradisiaca* species, in *D. gs. poaceaephila*, and in a few other strains (Figure 6). The
498 ferulate esterases FaeT and FaeD were absent in *P. confusarubida* and were acquired in the basal
499 *D. paradisiaca* and *D. aquatica* species. The FaeT enzyme was then secondarily lost in *D. gs.*
500 *poaceaephila*, and *D. dadantii subsp dieffenbachiae* (Supplementary Figure S3). Altogether, the

501 variability observed in the pectinasome among *Dickeya* species and even among different strains
502 likely reflects the dynamic evolution of the corresponding gene families involving gene acquisitions
503 and losses. More compellingly, this variability indicates that the pectinasome cannot be used to
504 distinguish among various species.

505

506 **Cellulases and Xylanases**

507 Two cellulases Cel5Z and CelY were likely ancestral in *Dickeya* species except in *D. gs.*
508 *poaceaephila*, which has lost Cel5Z. While Cel5Z was involved in cellulose degradation, CelY
509 belonged to the *bcs* gene cluster responsible for cellulose fiber formation (Prigent-Combaret et al.,
510 2012). Xylanases (XynA, XynB) and xylosidases (XynC, XynD) cleave xylan and xyloglucan, which
511 belong to the hemicelluloses. The β 1,4 xylan is mainly present in plant cell wall of monocots (Pena
512 et al., 2016) and is further decorated, often by acetyl, arabinosyl, and glucuronosyl side-chain
513 substitutions. Notably, the xylan substitution patterns depend on the plant species and are distinct
514 in gymnosperms and angiosperms (Busse-Wicher et al., 2016). Xyloglucan is a β -1,4 glucan that
515 can be substituted with a diverse array of glycosyl and nonglycosyl residues. The type and order of
516 xyloglycan substituents depend on the plant species (Pauly and Keegstra, 2016). Xyloglucan
517 polymers fall into one of two general types. In one type, three out of four backbone glucosyl
518 residues are xylosylated, leading to an XXXG-type xyloglycan, which is predominant in most
519 dicots. Another type of xyloglycan exhibits reduced xylosylation in that only two out of the four or
520 more backbone glucosyl residues are xylosylated, resulting in the XXGGn-type xyloglucan, which
521 is present in early land plants such as liverworts, mosses, lycophytes, and ferns of the order
522 Polypodiales. This type of xyloglucan seems to be absent from the gymnosperms and
523 angiosperms, with the exception of the grasses (Poales) and plants from the order Solanales, such
524 as potato, tobacco and tomato (Pauly and Keegstra, 2016). Thus, it is tempting to hypothesize that
525 the xylanase and xylosidase content of *Dickeya* species could be correlated with their plant host.
526 Interestingly, all *Dickeya* species as well as *P. confusarubida* contained at least one XynA, XynB,
527 XynC, or XynD coding gene (Figure 6). The phylogeny of XynA indicated that the corresponding
528 gene was likely present in the ancestor of *P. confusarubida* and *Dickeya*, and secondarily lost in *D.*

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

548

549 **Proteases**

550 Finally, the proteases PrtA, PrtB, PrtC, and PrtG, resulting from specific duplications that occurred
551 during the diversification of *Dickeya*, and the associated type I protease secretion system PrtDEF
552 were absent in *P. confusarubida* and the basal branching *D. paradisiaca*. Yet, *Pectobacterium*
553 harbour closely related homologues of PrtDEF and a single protease-coding gene closely related
554 to *Dickeya* PrtA, PrtB, PrtC and PrtG. This suggested that the whole system could have been
555 present in the common ancestor they shared with *Dickeya*, and then secondarily lost in *P.*
556 *confusarubida* and *D. paradisiaca* (Supplementary Figure S3). PrtG was specifically lost in *D.*

557 *chrysanthemii*, *D. dianthicola* *D. gs. undicola* and *D. gs. poaceaephila*. This later genomospecies
558 conserved only one protease PrtC whereas *D. dianthicola* strains RNS04-9 and NCPPB453
559 contained a *prtA* pseudogene and retained two proteases PrtC, PrtB (Figure 6).

560

561 **Other factors**

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

572

573 ***Distribution of secretion systems in Dickeya species***

574 For all *Dickeya* species, possession of secretion systems allowing them to actively secrete
575 virulence factors is of crucial importance. Unsurprisingly, different protein secretion systems (T1SS
576 to T6SS) are present in *Dickeya* species (Supplementary Table S4). As previously mentioned, the
577 type I protease secretion system PrtDEF was present in the ancestor of all *Dickeya* species and
578 lost in the basal *D. paradisiaca* species and in *P. confusarubida*. All *Dickeya* species are equipped
579 with the Out specific T2SS responsible for the secretion of most pectinases and the cellulase
580 Cel5Z. The phylogeny of the components of the Out system indicated it was likely present in the
581 ancestor shared by *Pectobacterium*, *P. confusarubida*, and *Dickeya*, and was conserved during the
582 diversification of *Dickeya* (Supplementary Figure S3). A second Stt specific T2SS, allows secretion
583 of the pectin lyase PnIH (Ferrandez and Condemine, 2008). Accordingly, the taxonomic
584 distributions of the pectin lyase PnIH and the Stt T2SS components were similar, being present in

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

608 All *Dickeya* species and *P. confusarubida* were found to possess a two-partner secretion system
609 (T5SS) CdiB-CdiA mediating bacterial intercellular competition. Their phylogenies clearly indicate
610 an ancestral presence in both lineages (Supplementary Figure S3). CdiB is a transport protein that
611 exports and presents CdiA proteins on the cell surface (Willett et al., 2015). The Cdi system is
612 involved in contact-dependent growth inhibition (CDI) by delivering the C-terminal toxin domain of

613 CdiA (CdiA-CT) to target bacteria (Aoki et al., 2010). Some *Dickeya* strains are equipped with two
614 CdiA proteins, for example, *D. dadantii* 3937, which produces two different CdiA-CT toxins: the first
615 one being a tRNase and the second one harbouring DNase activity (Aoki et al., 2010; Ruhe et al.,
616 2013). Each Cdi system also encodes a specific CdiI antitoxin that interacts with the cognate CdiA-
617 CT toxin and prevents auto-inhibition (Willett *et al.*, 2015).

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

632

633 ***Distribution of plant adherence elements in Dickeya species***

634 A chaperone–usher pilus assembly pathway, associated with type I fimbriae (*fimEAICDFGHB*),
635 was present in *D. aquatica* but not in any other *Dickeya* species (Supplementary Table S4).
636 Phylogenetic analyses suggested an acquisition from *Morganellaceae* or *Enterobacteriaceae*,
637 through HGT (Supplementary Figure S3). The adhesin FimH, a two-domain protein at the tip of
638 type I fimbriae is known to recognize mannoside structures and to be responsible for adhesion to
639 both animal epithelial cells and plant surface (Haahtela et al., 1985; Sauer et al., 2000). Therefore,
640 we can hypothesize that type I fimbriae could be involved in adherence of *D. aquatica* to plant

641 surface. Interestingly, in *Xylella fastidiosa*, the Fim system is known to be an antagonist of
642 twitching motility caused by type IV pili (De La Fuente et al., 2007). In *D. aquatica*, the mutually
643 exclusive production of type I fimbriae and type IV pilus could be linked to the specific pH
644 regulation of type IV pilus. When *D. aquatica* penetrates into the intercellular apoplast, which is an
645 acidic compartment, twitching motility would be induced to favour bacteria dissemination in plant
646 tissues, while the type I fimbriae would be no longer required during the colonization. This
647 regulation of twitching would thus contribute to the efficiency of *D. aquatica* for infecting tomatoes,
648 cucumbers, and probably other acidic fruits.

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

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

664

665 ***Distribution of iron assimilation systems in Dickeya species***

666 Iron acquisition by *Dickeya* is required for the systemic progression of maceration symptoms in the
667 plant hosts (Enard et al., 1988, Dellagi et al., 2005, Franza et al., 2005). To chelate iron from the
668 surroundings, most *Dickeya* species synthesize and excrete two siderophores: the

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

697 least four systems were present in each species. This multiplicity underscores the fact that
698 competition for this essential metal is critical for the outcome of the plant-*Dickeya* interaction.

699

700 ***Biosynthesis of secondary metabolites in Dickeya species***

701 In addition to siderophores, some *Dickeya* species produce secondary metabolites such as the
702 phytotoxin zeamine and the antifungal compound oocydin via non-ribosomal peptide synthases
703 (NRPS) and polyketide synthases (PKS) (Zhou et al., 2011a; Matilla et al., 2012). To evaluate the
704 diversity of secondary metabolites produced by the *Dickeya* genus, we screened the eight
705 complete genomes (*D. paradisiaca* Ech703, *D. aquatica* 174/2, *D. zea* EC1, *D. zea* Ech586, *D.*
706 *chrysanthemi* Ech1591, *D. solani* IPO2222, *D. fangzhongdai* N14b, *D. dadantii* 3937) and the three
707 partial genomes (*D. dianthicola* RNS04.9, *D. gs. poaceaephila* Ech569, *D. gs. undicola* 2B12) for
708 gene clusters encoding NRPS or/and PKS. Then, we analysed the evolutionary history of these
709 genes clusters among the 49 *Dickeya* genomes. The *oocBCDEFGJKLMNOPQRSTUVWXYZ* gene
710 cluster coding for oocydin biosynthesis proteins was present in *D. paradisiaca*, *D. zea* strains
711 isolated from rice, *D. chrysanthemi* subspecies *chrysanthemi*, *D. solani*, *D. dianthicola*, and *D.*
712 *fangzhongdai* strain NCPPB 3274 (Supplementary Table S4). This could suggest an ancestral
713 presence in *Dickeya* accompanied by losses in *D. aquatica*, *D. gs. poaceaephila*, *D. dadantii*, *D.*
714 *gs. undicola*, most *D. fangzhongdai* strains, and some Cluster I strains (Supplementary Figure S3).
715 However, the hypothesis of acquisition and spreading through HGT within *Dickeya* could not be
716 excluded. The *zmsABCDEFGHIJKLMNPQRS* gene cluster directing zeamine biosynthesis was
717 restricted to *D. zea* strains isolated from rice, *D. fangzhongdai* and *D. solani* (Supplementary
718 Table S4), suggesting secondary acquisition by HGT (Supplementary Figure S3). In addition,
719 genes involved in coronafic acid biosynthesis, a phytotoxin classically produced by *Pseudomonas*
720 *syringae* (Bender et al., 1999), were present only in *D. gs. poaceaephila* and *D. dadantii*
721 subspecies *dieffenbachiae* (Supplementary Table S4), suggesting specific acquisition via HGT. We
722 detected four additional gene clusters encoding NRPS and PKS, (i) cluster 1 was specific to *D.*
723 *paradisiaca* and *P. confusarubida*, (ii) cluster 2 was specific to *D. paradisiaca*, (iii) cluster 3 was
724 specific to *D. aquatica*, suggesting recent acquisitions by these species, while (iv) cluster 4 was

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

735

736 ***Evolution of Dickeya gene repertoires***

737 The 49 *Dickeya* proteomes used in this study contained in average 4,022 proteins, ranging in size
738 from 3,533 (*D. gs. poaceaephila*) up-to 4,352 (*D. fangzhongdai* NCPPB 3274) proteins,
739 representing a difference of 819 proteins (Supplementary Table S3). Comparison of the 197,073
740 proteins contained in the 49 *Dickeya* proteomes led to the delineation of 11,566 protein families
741 (Figure 7). These protein families correspond to the pan-proteome of *Dickeya*. Among these
742 protein families, 13.9% (1,604) were present at least in one copy in all *Dickeya* proteomes (Figure
743 7) and defined the core-proteome of this genus. Yet, the size of both pan- and core-proteomes
744 could be slightly underestimated because the proteomes of some strains were deduced from draft
745 genomes. Nevertheless, this meant that in average, ~39.9% of the proteins of any *Dickeya*
746 proteome belonged to the core proteome, while a given *Dickeya* proteome encompassed only
747 ~34.8% of the pan-proteome of this genus. Random taxonomic sampling-based rarefaction curves
748 indicated that sequencing more *Dickeya* genomes will probably not change significantly the
749 estimated size of the core-proteome, while it appears that the pan-proteome is far from being fully
750 disclosed (Figure 7B). This highlights the high diversity and plasticity of the gene repertoires in
751 *Dickeya*. This observation coupled to the great diversity of the virulence, stress resistance,
752 metabolism, and secretion systems imply that none of the *Dickeya* strains could be regarded as a

753 representative model for this genus. Core protein families could be punctually lost in a given strain,
754 while some protein families can be present transiently in a few strains. Thus, it is also relevant to
755 consider the persistent (i.e. protein families present in more than 90% of the strains) and volatile
756 proteomes (i.e. protein families present in less than 10% of the strains) (Touchon et al., 2009). In
757 *Dickeya*, most protein families could be classified either as persistent (2,714 protein families,
758 23.5%) or volatile (6,426 protein families, 55.6%) (Figure 7A). Unsurprisingly, the persistent
759 proteome was enriched in proteins with known functions, while volatile proteome encompassed
760 mostly hypothetical proteins, prophage elements, and transposases. It is tempting to consider the
761 genes encoding for the volatile proteome as a reservoir of functional innovations, yet the adaptive
762 potential of these genes remains a matter of debate (Touchon et al., 2009).

763 Among the 11,566 protein families inferred in *Dickeya*, 3,452 corresponded to strain specific
764 families (i.e. being present in a single proteome) (Figure 7A and Supplementary Table S8). The
765 number of strain specific protein families ranged from 336 in *D. aquatica* 174/2^T and 304 in *D. gs.*
766 *Poaceaephila*, down to zero in *Dickeya solani* strains MK16, PPO9134 and RNS0773B
767 (Supplementary Table S8). To determine the origin of these strain specific protein families, we
768 used them as seeds to query with BLASTP (e-value cut-off 10⁻⁴) a local database gathering 3,104
769 complete prokaryotic proteomes, including the 49 *Dickeya* and *P. confusarubida* ATCC 39006
770 proteomes. Results indicated that 1,051 (30.4%) *Dickeya* strain specific protein families displayed
771 best hit in one of the other 48 *Dickeya* strains, meaning that those sequences were wrongly
772 considered as strain specific because they did not satisfy the coverage and identity parameters
773 used to delineate the protein families. This was not surprising because some protein families could
774 be fast-evolving, meaning that applying uniform parameters can fail to delineate correctly these
775 protein families and lead to an overestimation of the strain specific protein families. In contrast,
776 1,625 (47.1%) *Dickeya* strain specific protein families displayed best hits in non-*Dickeya*
777 proteomes, meaning that the corresponding genes were likely acquired by HGT from non-*Dickeya*
778 donors. The taxonomic distribution of the corresponding sequences pinpointed *Proteobacteria*
779 (especially *Enterobacteriales*), and to a less extent *Firmicutes* as major donors (Supplementary
780 Table S8 B-C). Yet, the contribution of these two phyla is likely overestimated due to their

781 overrepresentation in sequence databases compared to other lineages. Accordingly, these results
782 should be interpreted as general trends but additional data would be required to precisely estimate
783 the real contribution of *Firmicutes* and *Proteobacteria*. Finally, 776 (22.5%) *Dickeya* strain specific
784 protein families displayed no significant hits or no hits at all, indicating that the corresponding
785 genes were truly strain specific or corresponded to annotation errors (false positives).

786 At the species level proteome size variation ranged from 435 (*D. aquatica*) to 120 (*D. paradisiaca*)
787 proteins (Supplementary Table S8). The taxonomic distribution of the species-specific protein
788 families displayed overall similar pictures, with most of them being present in all strains of the
789 species (Supplementary Figure S6). This revealed a relative homogeneity of proteomes within
790 species. Interestingly, a few strains diverged from this general trend, such as strain NCPBB 3274
791 within *D. fangzhongdai*, *D. aquatica* 174/2^T, *D. chrysanthemi* NCPPB 402, *D. dadantii* NCPPB
792 2976, and *D. solani* RNS 0512A. This is consistent with some previous studies. For instance, *D.*
793 *dadantii* NCPPB 2976 is part of the *dieffenbachiae* subspecies and has been shown to be clearly
794 different from the other *D. dadantii* subsp *dadantii* strains based on ANI values (Zhang et al.,
795 2016). Among, *D. solani*, strain RNS 0512A was proposed to define a novel *D. solani* sub-group
796 based on the high number and wide distribution of nucleotide variations compared to other *D.*
797 *solani* strains (Khayati et al., 2015).

798 Using COUNT (Csuros, 2010) on the 12,660 protein families built with SILIX and the topology of
799 the core-pf tree, we inferred the ancestral protein repertoires at each node of the *Dickeya* core-pf
800 phylogeny (Figure 8). COUNT provided similar results for different *Dickeya* species, irrespectively
801 of the position of *D. solani*, as either the sister-group of *D. dadantii* or *D. dianthicola*
802 (Supplementary Table S9), as suggested by core-pf and rprots phylogenetic analyses (see above).

803 The main difference lies in the numbers of gains and losses at the base of *Dickeya*.

804 We inferred 4,627 protein families in the ancestor of *Dickeya*, while in average 3,921 protein
805 families are contained in present-day *Dickeya* proteomes. This corresponds to a global loss of 18%
806 of the protein families. Interestingly, loss of protein families dominated over gains and affected all
807 *Dickeya* species (Figure 8). Highest protein losses were observed on the stems leading to *D. gs.*
808 *poaceaephila* and *D. aquatica* and to a lesser extent in *D. paradisiaca*, *D. gs. undicola*, and *D.*

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

816

817 *Characterization of the D. aquatica mobilome*

818 Mobile Genetic Elements (MGEs) are the main actors of the HGT and include plasmids, viruses
819 (phages and prophages) and transposons (Jackson et al., 2011). They are often localised within
820 genomic islands on chromosomes. The mobilome of a strain is the repertoire of all the genes
821 associated with MGEs. Using both PHAST and IslandViewer, we detected seven phage elements
822 and ten genomic islands in *D. aquatica* 174/2^T (Figure 1, Supplementary Table S10). Most of the
823 genomic islands contained genes of transposases, integrases or mobile elements that were likely
824 remnants of HGT. They also contain 105 of the 336 ORFAN genes detected in *D. aquatica* 174/2^T
825 strain. Among the seven detected prophages, P2, P6 and P7 were related to transposable Mu-
826 phage. These were also present in some *D. zea* strains isolated from river as well as *D.*
827 *dianthicola* strains (Supplementary Table S10). P3 and P5 were specific to some *D. aquatica*
828 strains, while P1, a defective prophage with only few conserved genes, and P4 were widespread in
829 all *Dickeya* species (Supplementary Table S10). Among the ten genomic islands detected in *D.*
830 *aquatica* 174/2^T, seven (GI1, GI2, GI4, GI5, GI6, GI7, GI9) were mainly composed of mobile
831 elements and small hypothetical proteins, GI2 also contained a type III restriction-modification
832 system, whereas GI4 included a type I restriction-modification system and a toxin-antitoxin system
833 (Supplementary Table S10). Similarly, GI6 contained a toxin-antitoxin system as well as an
834 isolated non-ribosomal peptide synthase, which was also found in *D. gs. poaceaephila* and *D. zea*
835 (Supplementary Table S10). GI7 contained some metabolic proteins, including the previously
836 mentioned cluster 3 encoding NRPS and PKS, as well as transporters, notably a cobalt/nickel ABC

837 transporter (Supplementary Table S10). Excluding the mobile elements, GI1, GI2, GI4, GI5, GI6
838 GI7, and GI9 were specific to *D. aquatica* strains, even if a few genes composing these GIs can be
839 punctually detected in some others strains (Supplementary Table S10). The three other genomic
840 islands (GI3, GI8, GI10) were metabolic islands (Supplementary Table S10). GI3 has been laterally
841 transferred between *Erwinia pyrofolia* and *D. aquatica* (Supplementary Figure S3). GI8 that
842 included proteins related to fatty acid metabolism, was conserved in *D. zeae* strains isolated from
843 rice or originated from China (Supplementary Table S10). GI10 contained proteins related to the
844 complete carbapenem biosynthetic pathway CarABCDE and the associated resistance proteins
845 CarFG (Supplementary Table S10). *P. confusarubida* contained genes coding for CarABCDE as
846 well as CarFG, in agreement with its capacity to synthesize the carbapenem antibiotic (carbapen-
847 2-em-3-carboxylic acid) (Thomson et al., 2000; Coulthurst et al., 2005). This cluster was conserved
848 in *D. undicola*, *D. dadantii* subspecies *dieffenbachiae*, some *D. chrysanthemi* strains and in the *D.*
849 *zeae* CSL-RW192 strain isolated from water (Supplementary Table S10). *D. paradisiaca* only
850 contained the *carFG* resistance genes but was deprived of the biosynthetic genes (Supplementary
851 Table S9). The phylogeny of CarABCDE and CarFG shows that relationships among the *Dickeya*
852 sequences are inconsistent with the species phylogeny, suggesting a secondary acquisition
853 through HGT in this genus (Supplementary Figure S3).

854 Overall, phages and genomic islands from *D. aquatica* are mostly species specific suggesting a
855 wide genomic plasticity in the *Dickeya* genus.

856

857 *Concluding remarks*

858 In this work, we sequenced the complete genome of *D. aquatica* 174/2 and showed that unlike
859 initially supposed, this bacterium, similarly to other *Dickeya* species, is a phytopathogen. Who is
860 the natural host of *D. aquatica* and why is this bacterium found in aquatic environments rather than
861 associated with plants as are its close relatives? Although the data at hand are not sufficient to
862 answer these questions, we speculate that *D. aquatica* can be pathogenic for aquatic plants such
863 as charophytes, which are a family of complex-structured algae living in a variety of wetland and
864 freshwater habitats including those, from which the *D. aquatica* strains have been isolated.

865 Charophytes are thought to be the closest ancestor of land plants. The cells of these algae are
866 surrounded by polysaccharide-based cell walls. However, their cell walls are thin and cannot be
867 distinguished as primary or secondary cell walls. Furthermore, they lack xyloglucans, which are
868 common in most land plants (Sarkar et al., 2009). Intriguingly, *D. aquatica* lacks the xylanases and
869 xylose degradation pathways and we hypothesize that this could reflect an evolutionary adaptation
870 to charophyte hosts.

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

883

884 **Experimental procedures**

885 *Genome sequencing, assembly and annotation*

886 DNA for sequencing of the *D. aquatica* type strain (174/2) genome was extracted from overnight
887 broth culture using Promega bacterial genomic DNA kit. PacBio sequencing to > 350X coverage
888 was performed by Eurofins Genomics (<https://www.eurofinsgenomics.eu/>). Reads were assembled
889 using CANU (Koren et al., 2017). The annotation was performed automatically with RAST (Aziz et
890 al., 2008), then expertly reviewed using MAGE (Vallenet et al., 2013) and literature data. The
891 expert review allowed to assign 2,457 gene names, correct 547 annotations and add 5 CDS
892 missed by RAST. The replication origin (oriC) was predicted by OriFinder

893 (<http://tubic.tju.edu.cn/Ori-Finder>) (Gao and Zhang, 2008). The presence of mobile genetic
894 elements in the *D. aquatica* 174/2 genome was investigated by the following online tools:
895 IslandViewer (<http://pathogenomics.sfu.ca/islandviewer>) (Dhillon et al., 2015) for the GI regions,
896 CRISPRDetect (http://brownlabtools.otago.ac.nz/CRISPRDetect/predict_crispr_array.html) (Biswas
897 et al., 2016) for CRISPR arrays, whereas putative prophage sequences were identified by PHAST
898 and PHASTER analysis (<http://phast.wishartlab.com/>) (Zhou et al., 2011b; Arndt et al., 2016). The
899 genome sequence has been submitted to EMBL database under accession number
900 GCA_900095885 (<http://www.ebi.ac.uk/>).

901

902 *Virulence assays on various hosts*

903 Bacterial cultures were grown in M63G minimal medium (M63 + 0.2% w/v glucose) (Miller 1972)
904 and diluted to a given OD₆₀₀ depending on the host: 0.2 (chicory) or 1 (potato, cucumber, tomato,
905 pineapple and kiwi). For chicory, 5 µL of bacterial suspension were injected into a 2 cm incision at
906 the center of the leaf. For potato, cucumber and tomato, 5, 100 and 200 µL of bacterial suspension
907 were injected into the vegetable, respectively. 200 µL of bacterial suspension were also injected
908 into pineapple and kiwi fruits. Plants were incubated at 30°C with 100% humidity for 18 h (chicory)
909 or 42 h (potato, cucumber and tomato) or 78 h (pineapple and kiwi). The soft rot mass is used to
910 quantify virulence.

911

912 *Stress resistance assays*

913 Bacteria were cultured at 30°C in 96 well plates using M63G (M63 + 0.2% w/v glucose) pH 7.0 as
914 minimal medium. Bacterial growth (OD_{600nm}) was monitored for 48 h using an Infinite® 200 PRO -
915 Tecan instrument. Resistance to osmotic stress was analysed using M63G enriched in 0.05 to 0.5
916 M NaCl. Resistance to oxidative stress was analysed in the same medium by adding H₂O₂
917 concentrations ranging from 25 to 200 µM. The pH effect was analyzed using the same M63G
918 medium buffered with malic acid at different pH ranging from 3.7 to 7.0.

919

920 *Proteome database construction*

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

925

926 *Assembly of Dickeya and Serratia ATCC 39006 protein families*

927 *Dickeya* and *Serratia* ATCC 39006 protein families were assembled with SILIX version 1.2.9 (Miele
928 et al., 2011). More precisely, pairwise comparisons of protein sequences contained in DickeyaDB
929 were performed using the BLASTP program version 2.2.26 with default parameters (Altschul et al.,
930 1997). Proteins in a pair providing HSP (High-scoring Segment Pairs) with identity over 60% and
931 covering at least 80% of the protein lengths were gathered in the same family. This led to the
932 assembly of 12,660 protein families, among which 1,493 were present in at least one copy in all
933 DickeyaDB proteomes. In contrast, considering the 49 *Dickeya* proteomes without taking into
934 account *Serratia* ATCC 39006 led to the assembly of 11,566 protein families, among which 1,604
935 were present at least in one copy in all *Dickeya* and *Serratia* ATCC 39006 proteomes, and 1,420 in
936 exactly one copy.

937

938 *Inference of reference phylogenies of Dickeya*

939 Reference phylogenies of *Dickeya* were inferred using ribosomal proteins (rprots) on the one hand
940 and core protein families (core-pf) on the other hand. The rprots phylogenetic tree was rooted with
941 sequences from *Serratia* ATCC 39006, together with three *Pectobacterium* species
942 (*Pectobacterium carotovorum* PC1, *Pectobacterium atrosepticum* SCRI1043, and *Pectobacterium*
943 *wasabia* WPP163) and five additional *Serratia* species (*Serratia marcescens* FGI94, *Serratia*
944 *fonticola* DSMZ4576, *Serratia liquefaciens* ATCC27592, *Serratia proteamaculans* 568, and
945 *Serratia plymuthica* AS9), while the core-pf phylogenetic tree was rooted with *Serratia* ATCC
946 39006 (see results).

947 Rprots sequences were extracted from the DickeyaDB using the engine of the riboDB database
948 (Jauffrit et al., 2016). Briefly, the riboDB engine allows retrieving rprots sequences through a

949 double approach combining reciprocal best-blast-hits and hidden Markov model (HMM) profiles
950 searches.

951 Starting from the 1,420 core-pf protein families containing exactly one copy in each proteome of
952 the DickeyaDB and 51 rprots, we applied several quality controls. First, within protein families
953 extremely short sequences (<30% of the median length of the family) were discarded. Then, we
954 used FastTreeMP (Price et al., 2010), and PhyloMCOA (de Vienne et al., 2012) to detect and
955 discard outlier sequences, on the basis of nodal and patristic distances. At the end of the quality
956 control process, 1,341 protein families present in more than 35 (70%) out of the 50 considered
957 proteomes and 51 rprots were kept. For each of these protein families, multiple alignments were
958 built using the CLUSTAL-Omega-1.1.0 program (Sievers et al., 2011) and trimmed using
959 GBLOCKS (Castresana, 2000) with parameters set to a minimal trimming. The trimmed multiple
960 alignments corresponding to 1,341 core-pf on the one hand and the 51 rprots on the other hand
961 have been combined using Seaview-4.5.4 (Gouy et al., 2010) to build the core-pf and the rprots
962 supermatrices containing 414,696 and 6,295 amino acid positions, respectively.

963 Maximum likelihood (ML) phylogenies of these supermatrices have been inferred with IQ-TREE-
964 1.5.3 (Nguyen *et al.*, 2015) with a C60 profile mixture of the Le and Gascuel evolutionary model
965 (Le and Gascuel 2008) and a gamma distribution with four site categories (Γ_4) to model the
966 heterogeneity of evolutionary rates across sites, as proposed by the model testing tool
967 (Kalyaanamoorthy *et al.*, 2017) available in IQ-TREE. The robustness of the inferred ML trees was
968 estimated with the non-parametric bootstrap procedure implemented in IQ-TREE-1.5.3 for the
969 rprots supermatrix (100 replicates of the original alignments) and the ultrafast bootstrap approach
970 for the core-pf supermatrix (1,000 replicates).

971

972 *Phylogenetic analysis of single markers*

973 947 SSU rRNA and 125 LSU rRNA complete sequences from *Dickeya*, *Serratia* (including strain
974 ATCC 39006), and *Pectobacterium* available in public databases were retrieved and aligned with
975 MAFFT v7.222. The resulting multiple alignments were trimmed using BMGE-1.1 (default
976 parameters). The phylogeny of the 947 SSU rRNA sequences was inferred using FastTree-2.1.9

977 (Price et al., 2010), with the GTR + gamma + cat 4 model, while the LSU rRNA tree was inferred
978 with IQ-TREE with the TIM3+F+I+Γ4 model according to the BIC criterion, as suggested by the
979 propose model tool available in IQ-TREE.

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

987

988 *Identification of the D. fangzhongdai phylogenetic cluster*

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

996

997 *Ancestral gene content*

998 The program COUNT (Csuros, 2010) was used for gene families evolutionary reconstruction in
999 *Dickeya* species using the topology of the core-pf tree as reference. All the generated 12,660
1000 families were submitted to COUNT, which can perform ancestral genome reconstruction by
1001 posterior probabilities in a phylogenetic birth-and-death model. Rates were optimized using a gain-
1002 loss-duplication model and three discrete gamma categories capturing rate variation across
1003 families, with other parameters set at default and allowing different gain-loss and duplication-loss
1004 rates for different branches. One hundred rounds of optimization were computed. COUNT was run

1005 twice: first with *D. solani* as sister-group of *D. dadantii* and then as sister-group of *D. dianthicola*.

1006

1007 **Acknowledgements**

1008 This work was supported by the Investissement d'Avenir grant (ANR-10-BINF-01-01), by the ANR
1009 Combicontrol grant (ANR-15-CE21-0003-01), by a grant from the FR BioEnviS and using the
1010 computing facilities of the Computing Cluster of the LBBE/PRABI. We would like to thank Georgi
1011 Muskhelishvili for critical reading the manuscript and Frederic Jauffrit as well as Mailys Dumet for
1012 providing the extraction of ribosomal proteins.

1013

1014 **Conflict of interest statement**

1015 The authors declare that no conflicting interests exist.

1016

1017 **References**

- 1018 Alič Š, Naglic T, Llop P, Toplak N, Koren S, Ravnikar M & Dreo T (2015) Draft Genome Sequences of
1019 *Dickeya* sp. Isolates B16 (NIB Z 2098) and S1 (NIB Z 2099) Causing Soft Rot of Phalaenopsis Orchids.
1020 *Genome announcements* **3**.
- 1021 Alič Š, Van Gijsegem F, Pédrón J, Ravnikar M & Dreo T (2018) Diversity within the novel *Dickeya*
1022 *fangzhongdai* sp., isolated from infected orchids, water and pears. *Plant Pathology* **67**, Issue 7, September
1023 2018: 1612-1620.
- 1024 Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997)
1025 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids*
1026 *Res* **25**: 3389-3402.
- 1027 Aoki, S.K., Diner, E.J., de Roodenbeke, C.T., Burgess, B.R., Poole, S.J., Braaten, B.A. et al. (2010) A
1028 widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. *Nature* **468**: 439-
1029 442.
- 1030 Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016) PHASTER: a better,
1031 faster version of the PHAST phage search tool. *Nucleic Acids Res* **44**: W16-21.
- 1032 Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A. et al. (2008) The RAST Server:
1033 rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.
- 1034 Bearson, S., Bearson, B., and Foster, J.W. (1997) Acid stress responses in enterobacteria. *FEMS Microbiol*
1035 *Lett* **147**: 173-180.
- 1036 Bender, C.L., Alarcon-Chaidez, F., and Gross, D.C. (1999) *Pseudomonas syringae* phytotoxins: mode of
1037 action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiol Mol Biol Rev* **63**:
1038 266-292.

- 1039 Berkey, R., Bendigeri, D., and Xiao, S. (2012) Sphingolipids and plant defense/disease: the "death"
1040 connection and beyond. *Front Plant Sci* **3**: 68.
- 1041 Bertani I, Passos da Silva D, Abbruscato P, Piffanelli P and Venturi V (2013) Draft Genome Sequence of the
1042 Plant Pathogen *Dickeya zea* DZ2Q, Isolated from Rice in Italy. *Genome announcements* **1**.
- 1043 Biswas, A., Staals, R.H., Morales, S.E., Fineran, P.C., and Brown, C.M. (2016) CRISPRDetect: A flexible
1044 algorithm to define CRISPR arrays. *BMC Genomics* **17**: 356.
- 1045 Brady, C.L., Cleenwerck, I., Denman, S., Venter, S.N., Rodriguez-Palenzuela, P., Coutinho, T.A., and De
1046 Vos, P. (2012) Proposal to reclassify *Brenneria quercina* (Hildebrand and Schroth 1967) Hauben et al.
1047 1999 into a new genus, *Lonsdalea* gen. nov., as *Lonsdalea quercina* comb. nov., descriptions of
1048 *Lonsdalea quercina* subsp. *quercina* comb. nov., *Lonsdalea quercina* subsp. *iberica* subsp. nov. and
1049 *Lonsdalea quercina* subsp. *britannica* subsp. nov., emendation of the description of the genus *Brenneria*,
1050 reclassification of *Dickeya dieffenbachiae* as *Dickeya dadantii* subsp. *dieffenbachiae* comb. nov., and
1051 emendation of the description of *Dickeya dadantii*. *Int J Syst Evol Microbiol* **62**: 1592-1602.
- 1052 Brenner, D.J., and Farmer III, J.J. (2005) Family I. *Enterobacteriaceae*. In *Bergey's Manual of Systematic*
1053 *Bacteriology*. D. J. Brenner, N. R. Krieg, J. T. Staley, G. M. Garrity, D. R. Boone, P. Vos et al. (eds). New
1054 York, US: Springer, pp. 587-607.
- 1055 Burdman, S., Bahar, O., Parker, J.K., and De La Fuente, L. (2011) Involvement of Type IV Pili in
1056 Pathogenicity of Plant Pathogenic Bacteria. *Genes (Basel)* **2**: 706-735.
- 1057 Busse-Wicher, M., Li, A., Silveira, R.L., Pereira, C.S., Tryfona, T., Gomes, T.C. et al. (2016) Evolution of
1058 Xylan Substitution Patterns in Gymnosperms and Angiosperms: Implications for Xylan Interaction with
1059 Cellulose. *Plant Physiol* **171**: 2418-2431.
- 1060 Buttner, D. (2016) Behind the lines-actions of bacterial type III effector proteins in plant cells. *FEMS Microbiol*
1061 *Rev.*
- 1062 Caffall, K.H., and Mohnen, D. (2009) The structure, function, and biosynthesis of plant cell wall pectic
1063 polysaccharides. *Carbohydr Res* **344**: 1879-1900.
- 1064 Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic
1065 analysis. *Mol Biol Evol* **17**: 540-552.
- 1066 Chan KG, Kher HL, Chang CY, Yin WF and Tan KH (2015) Analysis of Pectate Lyase Genes in *Dickeya*
1067 *chrysanthemi* Strain L11, Isolated from a Recreational Lake in Malaysia: a Draft Genome Sequence
1068 Perspective. *Genome announcements* **3**.
- 1069 Chandran Darbari, V., and Waksman, G. (2015) Structural Biology of Bacterial Type IV Secretion Systems.
1070 *Annu Rev Biochem* **84**: 603-629.
- 1071 Conway, J.R., Lex, A., and Gehlenborg, N. (2017) UpSetR: an R package for the visualization of intersecting
1072 sets and their properties. *Bioinformatics* **33**: 2938-2940.
- 1073 Coulthurst, S.J., Barnard, A.M., and Salmond, G.P. (2005) Regulation and biosynthesis of carbapenem
1074 antibiotics in bacteria. *Nat Rev Microbiol* **3**: 295-306.
- 1075 Csuros, M. (2010) Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood.
1076 *Bioinformatics* **26**: 1910-1912.
- 1077 Czajkowski, R., Pérombelon, M.C.M., van Veen, J.A., and van der Wolf, J.M. (2011) Control of blackleg and
1078 tuber soft rot of potato caused by *Pectobacterium* and *Dickeya* species: a review. *Plant Pathol* **60**: 999–

- 1079 1013.
- 1080 Daub, J., Eberhardt, R.Y., Tate, J.G., and Burge, S.W. (2015) Rfam: annotating families of non-coding RNA
1081 sequences. *Methods Mol Biol* **1269**: 349-363.
- 1082 de la Cruz, F., Frost, L.S., Meyer, R.J., and Zechner, E.L. (2010) Conjugative DNA metabolism in Gram-
1083 negative bacteria. *FEMS Microbiol Rev* **34**: 18-40.
- 1084 De La Fuente, L., Burr, T.J., and Hoch, H.C. (2007) Mutations in type I and type IV pilus biosynthetic genes
1085 affect twitching motility rates in *Xylella fastidiosa*. *J Bacteriol* **189**: 7507-7510.
- 1086 de Vienne, D.M., Ollier, S., and Aguilera, G. (2012) Phylo-MCOA: a fast and efficient method to detect outlier
1087 genes and species in phylogenomics using multiple co-inertia analysis. *Mol Biol Evol* **29**: 1587-1598.
- 1088 Degraeve, A., Siamer, S., Boureau, T., and Barny, M.A. (2015) The AvrE superfamily: ancestral type III
1089 effectors involved in suppression of pathogen-associated molecular pattern-triggered immunity. *Mol Plant*
1090 *Pathol* **16**: 899-905.
- 1091 Dekker, J., and Frank, K. (2015) *Salmonella*, *Shigella*, and *Yersinia*. *Clin Lab Med* **35**: 225-246.
- 1092 Dellagi, A., Rigault, M., Segond, D., Roux, C., Kraepiel, Y., Cellier, F. et al. (2005) Siderophore-mediated
1093 upregulation of Arabidopsis ferritin expression in response to *Erwinia chrysanthemi* infection. *Plant J* **43**:
1094 262-272.
- 1095 Dhillon, B.K., Laird, M.R., Shay, J.A., Winsor, G.L., Lo, R., Nizam, F. et al. (2015) IslandViewer 3: more
1096 flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res* **43**: W104-
1097 108.
- 1098 Duprey, A., Muskhelishvili, G., Reverchon, S., and Nasser, W. (2016a) Temporal control of *Dickeya dadantii*
1099 main virulence gene expression by growth phase-dependent alteration of regulatory nucleoprotein
1100 complexes. *Biochim Biophys Acta* **1859**: 1470-1480.
- 1101 Duprey, A., Nasser, W., Leonard, S., Brochier-Armanet, C., and Reverchon, S. (2016b) Transcriptional Start
1102 Site Turnover in the Evolution of Bacterial Paralogous Genes: the *peIE-peID* Virulence Genes in *Dickeya*.
1103 *FEBS J.* **283**(22):4192-4207.
- 1104 Enard, C., Dirolez, A., and Expert, D. (1988) Systemic virulence of *Erwinia chrysanthemi* 3937 requires a
1105 functional iron assimilation system. *J Bacteriol* **170**: 2419-2426.
- 1106 Expert, D., Boughammoura, A., and Franza, T. (2008) Siderophore-controlled iron assimilation in the
1107 enterobacterium *Erwinia chrysanthemi*: evidence for the involvement of bacterioferritin and the Suf iron-
1108 sulfur cluster assembly machinery. *J Biol Chem* **283**: 36564-36572.
- 1109 Ferrandez, Y., and Condemine, G. (2008) Novel mechanism of outer membrane targeting of proteins in
1110 Gram-negative bacteria. *Mol Microbiol* **69**: 1349-1357.
- 1111 Fineran PC, Iglesias Cans MC, Ramsay JP, et al. (2013) Draft Genome Sequence of *Serratia* sp. Strain
1112 ATCC 39006, a Model Bacterium for Analysis of the Biosynthesis and Regulation of Prodigiosin, a
1113 Carbapenem, and Gas Vesicles. *Genome announcements* **1**.
- 1114 Fontenot, E.M., Ezelle, K.E., Gabreski, L.N., Giglio, E.R., McAfee, J.M., Mills, A.C. et al. (2013) YfdW and
1115 YfdU are required for oxalate-induced acid tolerance in *Escherichia coli* K-12. *J Bacteriol* **195**: 1446-1455.
- 1116 Foster, J.W. (2004) *Escherichia coli* acid resistance: tales of an amateur acidophile. *Nat Rev Microbiol* **2**:
1117 898-907.
- 1118 Franza, T., and Expert, D. (2013) Role of iron homeostasis in the virulence of phytopathogenic bacteria: an

- 1119 'a la carte' menu. *Mol Plant Pathol* **14**: 429-438.
- 1120 Franza, T., Mahe, B., and Expert, D. (2005) *Erwinia chrysanthemi* requires a second iron transport route
1121 dependent of the siderophore achromobactin for extracellular growth and plant infection. *Mol Microbiol*
1122 **55**: 261-275.
- 1123 Franza, T., Michaud-Soret, I., Piqueret, P., and Expert, D. (2002) Coupling of iron assimilation and
1124 pectinolysis in *Erwinia chrysanthemi* 3937. *Mol Plant Microbe Interact* **15**: 1181-1191.
- 1125 Gao, F., and Zhang, C.T. (2008) Ori-Finder: a web-based system for finding oriCs in unannotated bacterial
1126 genomes. *BMC Bioinformatics* **9**: 79.
- 1127 Garland L, Koskinen P, Rouhiainen L, Laine P, Paulin L, Auvinen P, Holm L and Pirhonen M (2013) Genome
1128 Sequence of *Dickeya solani*, a New soft Rot Pathogen of Potato, Suggests its Emergence May Be
1129 Related to a Novel Combination of Non-Ribosomal Peptide/Polyketide Synthetase Clusters. *Diversity* **5**:
1130 824-842.
- 1131 Glasner, J.D., Yang, C.H., Reverchon, S., Hugouvieux-Cotte-Pattat, N., Condemine, G., Bohin, J.P. et al.
1132 (2011) Genome sequence of the plant-pathogenic bacterium *Dickeya dadantii* 3937. *J Bacteriol* **193**:
1133 2076-2077.
- 1134 Golanowska M, Galardini M, Bazzicalupo M, Hugouvieux-Cotte-Pattat N, Mengoni A, Potrykus M, Slawiak M
1135 and Lojkowska E (2015) Draft Genome Sequence of a Highly Virulent Strain of the Plant Pathogen
1136 *Dickeya solani*, IFB0099. *Genome announcements* **3**.
- 1137 Goto, M. (1979) Bacterial foot rot of rice caused by a strain of *Erwinia chrysanthemi*. *Phytopathology* **69**:
1138 213-216.
- 1139 Gouy, M., Guindon, S., and Gascuel, O. (2010) SeaView version 4: A multiplatform graphical user interface
1140 for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**: 221-224.
- 1141 Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New Algorithms
1142 and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0.
1143 *Systematic Biology*, **59**(3), 307–321.
- 1144 Haahtela, K., Tarkka, E., and Korhonen, T.K. (1985) Type 1 fimbria-mediated adhesion of enteric bacteria to
1145 grass roots. *Appl Environ Microbiol* **49**: 1182-1185.
- 1146 Hauben, L., Moore, E.R., Vauterin, L., Steenackers, M., Mergaert, J., Verdonck, L., and Swings, J. (1998)
1147 Phylogenetic position of phytopathogens within the Enterobacteriaceae. *Syst Appl Microbiol* **21**: 384-397.
- 1148 He, Z., Zhang, H., Gao, S., Lercher, M.J., Chen, W.H., and Hu, S. (2016) Evolvview v2: an online visualization
1149 and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res.* **44** (Web
1150 Server issue): W236-W241.
- 1151 Henrichsen, J. (1972) Bacterial surface translocation: a survey and a classification. *Bacteriol Rev* **36**: 478-
1152 503.
- 1153 Henrichsen, J. (1975) The influence of changes in the environment on twitching motility. *Acta Pathol*
1154 *Microbiol Scand B* **83**: 179-186.
- 1155 Hugouvieux-Cotte-Pattat, N., Condemine, G., and Shevchik, V.E. (2014) Bacterial pectate lyases, structural
1156 and functional diversity. *Environ Microbiol Rep* **6**: 427-440.
- 1157 Ilangovan, A., Connery, S., and Waksman, G. (2015) Structural biology of the Gram-negative bacterial
1158 conjugation systems. *Trends Microbiol* **23**: 301-310.

- 1159 Ishii, T. (1997) Structure and functions of feruloylated polysaccharides. *Plant Sci* **127**: 111-127.
- 1160 Jackson, R.W., Vinatzer, B., Arnold, D.L., Dorus, S., and Murillo, J. (2011) The influence of the accessory
1161 genome on bacterial pathogen evolution. *Mob Genet Elements* **1**: 55-65.
- 1162 Jacobs, J.M., Milling, A., Mitra, R.M., Hogan, C.S., Ailloud, F., Prior, P., and Allen, C. (2013) *Ralstonia*
1163 *solanacearum* requires PopS, an ancient AvrE-family effector, for virulence and To overcome salicylic
1164 acid-mediated defenses during tomato pathogenesis. *MBio* **4**: e00875-00813.
- 1165 Janse, J.D., and Ruissen, M.A. (1988) Characterization and classification of *Erwinia chrysanthemi* strains
1166 from several hosts in The Netherlands. *Phytopathology* **78**: 800-808.
- 1167 Jauffrit, F., Penel, S., Delmotte, S., Rey, C., de Vienne, D.M., Gouy, M. et al. (2016) RiboDB Database: A
1168 Comprehensive Resource for Prokaryotic Systematics. *Mol Biol Evol* **33**: 2170-2172.
- 1169 Jiang, X., Zghidi-Abouzid, O., Oger-Desfeux, C., Hommais, F., Greliche, N., Muskhelishvili, G. et al. (2016)
1170 Global transcriptional response of *Dickeya dadantii* to environmental stimuli relevant to the plant infection.
1171 *Environ Microbiol.* **18** (11):3651-3672.
- 1172 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermini, L.S. (2017) ModelFinder:
1173 fast model selection for accurate phylogenetic estimates. *Nature methods* **14**: 587-589.
- 1174 Keen, N.T., Boyd, C., and Henrissat, B. (1996) Cloning and characterization of a xylanase gene from corn
1175 strains of *Erwinia chrysanthemi*. *Mol Plant Microbe Interact* **9**: 651-657.
- 1176 Khayi, S., Blin, P., Chong, T.M., Chan, K.G., and Faure, D. (2016) Complete genome anatomy of the
1177 emerging potato pathogen *Dickeya solani* type strain IPO 2222T. *Stand Genomic Sci* **11**: 87.
- 1178 Khayi, S., Blin P, Pédrón J, Chong TM, Chan KG, Moumni M, Hélias V, Van Gijsegem F, Faure D (2015)
1179 Population genomics reveals additive and replacing horizontal gene transfers in the emerging pathogen
1180 *Dickeya solani* *BMC Genomics* **16**:788
- 1181 Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017) Canu: scalable
1182 and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**:
1183 722-736.
- 1184 Krulwich, T.A., Sachs, G., and Padan, E. (2011) Molecular aspects of bacterial pH sensing and homeostasis.
1185 *Nat Rev Microbiol.* **9**(5):330-43.
- 1186 Le, S. Q., and Gascuel, O. (2008) An improved general amino acid replacement matrix. *Molecular Biology*
1187 *and Evolution*, **25**(7), 1307–1320.
- 1188
- 1189 Lee, J.H., Yeo, W.S. and Roe, J.H. (2004) Induction of the *sufA* operon encoding Fe-S assembly proteins by
1190 superoxide generators and hydrogen peroxide: involvement of OxyR, IHF and an unidentified oxidant-
1191 responsive factor. *Mol Microbiol* **51**: 1745-1755.
- 1192 Lee, Y.A., and Yu, C.P. (2006) A differential medium for the isolation and rapid identification of a plant soft
1193 rot pathogen, *Erwinia chrysanthemi*. *J Microbiol Methods* **64**: 200-206.
- 1194 Letunic, I. and Bork, P. (2016) Interactive Tree Of Life (iTOL) v3: an online tool for the display and annotation
1195 of phylogenetic and other trees. *Nucleic Acids Res* **44** (W1): W242-245.
- 1196 Li, B., Shi, Y., Ibrahim, M., Liu, H., Shan, C., Wang, Y., Kube, M., Xie, G.L. and Sun, G. (2012) Genome
1197 sequence of the rice pathogen *Dickeya zeae* strain ZJU1202. *J Bacteriol* **194**: 4452-4453.
- 1198 Ma, B., Hibbing, M.E., Kim, H.S., Reedy, R.M., Yedidia, I., Breuer, J. et al. (2007) Host range and molecular

- 1199 phylogenies of the soft rot enterobacterial genera *Pectobacterium* and *Dickeya*. *Phytopathology* **97**: 1150-
1200 1163.
- 1201 Maderbocus, R., Fields, B.L., Hamilton, K., Luo, S., Tran, T.H., Dietrich, L.E.P. and Tong, L. (2017) Crystal
1202 structure of a *Pseudomonas* malonate decarboxylase holoenzyme hetero-tetramer. *Nature*
1203 *communications* **8**: 160.
- 1204 Maier, B., and Wong, G.C. (2015) How Bacteria Use Type IV Pili Machinery on Surfaces. *Trends Microbiol*
1205 **23**: 775-788.
- 1206 Matilla, M.A., Stockmann, H., Leeper, F.J., and Salmond, G.P. (2012) Bacterial biosynthetic gene clusters
1207 encoding the anti-cancer haterumalide class of molecules: biogenesis of the broad spectrum antifungal
1208 and anti-oomycete compound, oocydin A. *J Biol Chem* **287**: 39125-39138.
- 1209 Matsumoto, H., Jitareerat, P., Baba, Y., and Tsuyumu, S. (2003) Comparative study of regulatory
1210 mechanisms for pectinase production by *Erwinia carotovora* subsp. *carotovora* and *Erwinia chrysanthemi*.
1211 *Mol Plant Microbe Interact* **16**: 226-237.
- 1212 Miele, V., Penel, S., and Duret, L. (2011) Ultra-fast sequence clustering from similarity networks with SiLiX.
1213 *BMC Bioinformatics* **12**: 116.
- 1214 Miller JH (1972) *Experiments in Molecular Genetics*. New York.
- 1215 Munzinger, M., Budzikiewicz, H., Expert, D., Enard, C., and Meyer, J.M. (2000) Achromobactin, a new citrate
1216 siderophore of *Erwinia chrysanthemi*. *Z Naturforsch [C]* **55**: 328-332.
- 1217 Nachin, L., El Hassouni, M., Loiseau, L., Expert, D., and Barras, F. (2001) SoxR-dependent response to
1218 oxidative stress and virulence of *Erwinia chrysanthemi*: the key role of SufC, an orphan ABC ATPase. *Mol*
1219 *Microbiol* **39**: 960-972.
- 1220 Nasser, W., Dorel, C., Wawrzyniak, J., Van Gijsegem, F., Groleau, M.C., Deziel, E., and Reverchon, S.
1221 (2013) Vfm a new quorum sensing system controls the virulence of *Dickeya dadantii*. *Environ Microbiol*
1222 **15**: 865-880.
- 1223 Nasser, W., Robert-Baudouy, J., and Reverchon, S. (1997) Antagonistic effect of CRP and KdgR in the
1224 transcription control of the *Erwinia chrysanthemi* pectinolysis genes. *Mol Microbiol* **26**: 1071-1082.
- 1225 Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015) IQ-TREE: a fast and effective
1226 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**: 268-274.
- 1227 Nykyri, J., Mattinen, L., Niemi, O., Adhikari, S., Koiv, V., Somervuo, P. et al. (2013) Role and regulation of
1228 the Flp/Tad pilus in the virulence of *Pectobacterium atrosepticum* SCRI1043 and *Pectobacterium*
1229 *wasabiae* SCC3193. *PLoS One* **8**: e73718.
- 1230 Ogasawara, H., Hasegawa, A., Kanda, E., Miki, T., Yamamoto, K., and Ishihama, A. (2007) Genomic SELEX
1231 search for target promoters under the control of the PhoQP-RstBA signal relay cascade. *J Bacteriol.*
1232 **189**(13):4791-9.
- 1233 O'Neill, M.A., Ishii, T., Albersheim, P., and Darvill, A.G. (2004) Rhamnogalacturonan II: structure and
1234 function of a borate cross-linked cell wall pectic polysaccharide. *Annu Rev Plant Biol* **55**: 109-139.
- 1235 Parkinson, N., DeVos, P., Pirhonen, M., and Elphinstone, J. (2014) *Dickeya aquatica* sp. nov., isolated from
1236 waterways. *Int J Syst Evol Microbiol* **64**: 2264-2266.

- 1237 Parkinson, N., Stead, D., Bew, J., Heeney, J., Tsrer Lahkim, L. and Elphinstone, J. (2009) *Dickeya* species
1238 relatedness and clade structure determined by comparison of *recA* sequences. *Int J Syst Evol Microbiol*
1239 **59**: 2388-2393.
- 1240 Pauly, M., and Keegstra, K. (2016) Biosynthesis of the Plant Cell Wall Matrix Polysaccharide Xyloglucan.
1241 *Annu Rev Plant Biol* **67**: 235-259.
- 1242 Pedron, J., Mondy, S., des Essarts Y.R., Van Gijsegem, F., and Faure, D. (2014) Genomic and metabolic
1243 comparison with *Dickeya dadantii* 3937 reveals the emerging *Dickeya solani* potato pathogen to display
1244 distinctive metabolic activities and T5SS/T6SS-related toxin repertoire. *BMC Genomics* **15**: 283.
- 1245 Pena, M.J., Kulkarni, A.R., Backe, J., Boyd, M., O'Neill, M.A., and York, W.S. (2016) Structural diversity of
1246 xylans in the cell walls of monocots. *Planta* **244**: 589-606.
- 1247 Perez-Mendoza, D., Coulthurst, S.J., Humphris, S., Campbell, E., Welch, M., Toth, I.K., and Salmond, G.P.
1248 (2011) A multi-repeat adhesin of the phytopathogen, *Pectobacterium atrosepticum*, is secreted by a Type
1249 I pathway and is subject to complex regulation involving a non-canonical diguanylate cyclase. *Mol*
1250 *Microbiol* **82**: 719-733.
- 1251 Perombelon, M.C. (1990) *Ecology and pathology of soft rot Erwinias: an overview*. Budapest: Akademiai
1252 Kiado.
- 1253 Persmark, M., Expert, D., and Neilands, J.B. (1989) Isolation, characterization, and synthesis of
1254 chrysoactin, a compound with siderophore activity from *Erwinia chrysanthemi*. *J Biol Chem* **264**: 3187-
1255 3193.
- 1256 Pignocchi, C., and Foyer, C.H. (2003) Apoplastic ascorbate metabolism and its role in the regulation of cell
1257 signalling. *Curr Opin Plant Biol* **6**: 379-389.
- 1258 Poole, S.J., Diner, E.J., Aoki, S.K., Braaten, B.A., t'Kint de Roodenbeke, C., Low, D.A., and Hayes, C.S.
1259 (2011) Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition
1260 (CDI) and rearrangement hotspot (Rhs) systems. *PLoS genetics* **7**: e1002217.
- 1261 Potrykus, M., Golanowska, M., Sledz, W., Zoledowska, S., Motyka, A., Kolodziejska, A. Butrymowicz, J., and
1262 Lojkowska, E. (2016) Biodiversity of *Dickeya* spp. isolated from potato plants and water sources in
1263 temperate climate. *Plant Disease* **100** (2): 408-417
- 1264 Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2--approximately maximum-likelihood trees for
1265 large alignments. *PLoS One* **5**: e9490.
- 1266 Prigent-Combaret, C., Zghidi-Abouzid, O., Effantin, G., Lejeune, P., Reverchon, S., and Nasser, W. (2012)
1267 The nucleoid-associated protein Fis directly modulates the synthesis of cellulose, an essential component
1268 of pellicle-biofilms in the phytopathogenic bacterium *Dickeya dadantii*. *Mol Microbiol* **86**: 172-186.
- 1269 Pritchard, L., Humphris, S., Baeyen, S., Maes, M., Van Vaerenbergh, J., Elphinstone, J., Saddler, G., and
1270 Toth, I. (2013a) Draft Genome Sequences of Four *Dickeya dianthicola* and Four *Dickeya solani* Strains.
1271 *Genome announcements* **1**.
- 1272 Pritchard, L., Humphris, S., Saddler, G.S., Elphinstone, J.G., Pirhonen, M., and Toth, I.K. (2013b) Draft
1273 genome sequences of 17 isolates of the plant pathogenic bacterium *Dickeya*. *Genome announcements* **1**.
- 1274 Py, B., Bortoli-German, I., Haiech, J., Chippaux, M., and Barras, F. (1991) Cellulase EGZ of *Erwinia*
1275 *chrysanthemi*: structural organization and importance of His98 and Glu133 residues for catalysis. *Protein*
1276 *Eng* **4**: 325-333.

- 1277 Ramulu, H.G., Groussin, M., Talla, E., Planel, R., Daubin, V., and Brochier-Armanet, C. (2014) Ribosomal
1278 proteins: toward a next generation standard for prokaryotic systematics? *Mol Phylogenet Evol* **75**: 103-
1279 117.
- 1280 Raoul des Essarts, Y., Mondy, S., Helias, V. and Faure, D. (2015) Genome Sequence of the Potato Plant
1281 Pathogen *Dickeya dianthicola* Strain RNS04.9. *Genome announcements* **3**.
- 1282 Reverchon, S., and Nasser, W. (2013) *Dickeya* ecology, environment sensing and regulation of virulence
1283 program. *Environ Microbiol Rep* **5**: 622-636.
- 1284 Reverchon, S., Muskhelishvili, G., and Nasser, W. (2016) Virulence Program of a Bacterial Plant Pathogen:
1285 The *Dickeya* Model. *Prog Mol Biol Transl Sci* **142**: 51-92.
- 1286 Reverchon, S., Rouanet, C., Expert, D., and Nasser, W. (2002) Characterization of indigoidine biosynthetic
1287 genes in *Erwinia chrysanthemi* and role of this blue pigment in pathogenicity. *J Bacteriol* **184**: 654-665.
- 1288 Rincon-Enriquez, G., Crete, P., Barras, F., and Py, B. (2008) Biogenesis of Fe/S proteins and pathogenicity:
1289 IscR plays a key role in allowing *Erwinia chrysanthemi* to adapt to hostile conditions. *Mol Microbiol* **67**:
1290 1257-1273.
- 1291 Roche, B., Aussel, L., Ezraty, B., Mandin, P., Py, B., and Barras, F. (2013) Iron/sulfur proteins biogenesis in
1292 prokaryotes: formation, regulation and diversity. *Biochim Biophys Acta* **1827**: 455-469.
- 1293 Ruhe, Z.C., Low, D.A., and Hayes, C.S. (2013) Bacterial contact-dependent growth inhibition. *Trends*
1294 *Microbiol* **21**: 230-237.
- 1295 Ryu, C.M. (2015) Against friend and foe: type 6 effectors in plant-associated bacteria. *J Microbiol* **53**: 201-
1296 208.
- 1297 Samson, R., Legendre, J.B., Christen, R., Fischer-Le Saux, M., Achouak, W., and Gardan, L. (2005)
1298 Transfer of *Pectobacterium chrysanthemi* (Burkholder et al. 1953) Brenner et al. 1973 and *Brenneria*
1299 *paradisiaca* to the genus *Dickeya* gen. nov. as *Dickeya chrysanthemi* comb. nov. and *Dickeya*
1300 *paradisiaca* comb. nov. and delineation of four novel species, *Dickeya dadantii* sp. nov., *Dickeya*
1301 *dianthicola* sp. nov., *Dickeya dieffenbachiae* sp. nov. and *Dickeya zeeae* sp. nov. *Int J Syst Evol Microbiol*
1302 **55**: 1415-1427.
- 1303 Sandy, M., and Butler, A. (2011) Chrysobactin siderophores produced by *Dickeya chrysanthemi* EC16. *J Nat*
1304 *Prod* **74**: 1207-1212.
- 1305 Sarkar, P., Bosneaga, E., and Auer, M. (2009) Plant cell walls throughout evolution: towards a molecular
1306 understanding of their design principles. *J Exp Bot* **60**: 3615-3635.
- 1307 Sauer, F.G., Barnhart, M., Choudhury, D., Knight, S.D., Waksman, G., and Hultgren, S.J. (2000) Chaperone-
1308 assisted pilus assembly and bacterial attachment. *Curr Opin Struct Biol* **10**: 548-556.
- 1309 Seeman, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30** (14) : 2068-2069
- 1310 Seputiene, V., Motiejunas, D., Suziedelis, K., Tomenius, H., Normark, S., Melefors, O., and Suziedeliene, E.
1311 (2003) Molecular characterization of the acid-inducible *asr* gene of *Escherichia coli* and its role in acid stress
1312 response. *J Bacteriol* **185**: 2475-2484.
- 1313 Siamer, S., Guillas, I., Shimobayashi, M., Kunz, C., Hall, M.N., and Barny, M.A. (2014) Expression of the
1314 bacterial type III effector DspA/E in *Saccharomyces cerevisiae* down-regulates the sphingolipid
1315 biosynthetic pathway leading to growth arrest. *J Biol Chem* **289**: 18466-18477.
- 1316 Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W. et al. (2011) Fast, scalable generation of

- 1317 high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.
- 1318 Sławiak, M., Beckhoven, J., Speksnijder, A., Czajkowski, R., Grabe, G. and van der Wolf, J.M. (2009)
- 1319 Biochemical and genetical analysis reveal a new clade of biovar 3 *Dickeya* spp. strains isolated from
- 1320 potato in Europe. *European Journal of Plant Pathology* **125**: 245-261.
- 1321 Tan, K.H., Sheng, K.Y., Chang, C.Y., Yin, W.F., and Chan, K.G. (2015) Draft Genome Sequence of a
- 1322 Quorum-Sensing Bacterium, *Dickeya* sp. Strain 2B12, Isolated from a Freshwater Lake. *Genome*
- 1323 *announcements* **3**.
- 1324 Thomas, C.M., and Nielsen, K.M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between
- 1325 bacteria. *Nat Rev Microbiol* **3**: 711-721.
- 1326 Thomson, N.R., Crow, M.A., McGowan, S.J., Cox, A., and Salmond, G.P. (2000) Biosynthesis of
- 1327 carbapenem antibiotic and prodigiosin pigment in *Serratia* is under quorum sensing control. *Mol Microbiol*
- 1328 **36**: 539-556.
- 1329 Tian, Y., Zhao, Y., Yuan, X., Yi, J., Fan, J., Xu, Z. et al. (2016) *Dickeya fangzhongdai* sp. nov., a plant-
- 1330 pathogenic bacterium isolated from pear trees (*Pyrus pyrifolia*). *Int J Syst Evol Microbiol* **66**: 2831-2835.
- 1331 Toth, I., van der Wolf, J.M., Saddler, G., Lojkowska, E., Hélias, V., Pirhonen, M. et al. (2011) *Dickeya*
- 1332 species: an emerging problem for potato production in Europe. *Plant Pathology* **60**: 385-399.
- 1333 Touchon M, Hoede C, Tenailon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C,
- 1334 Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard M, Dossat C, Karoui ME,
- 1335 Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguéne C, Lescat M, Mangenot S, Martinez-
- 1336 Jehanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf CS, Schneider D, Turret J,
- 1337 Vacherie B, Vallenet D, Médigue C, Rocha EP, Denamur E. (2009) Organised genome dynamics in the
- 1338 *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* **5**: e1000344.
- 1339 Urbanikova, L., Vrsanska, M., Morkeberg Krogh, K.B., Hoff, T., and Biely, P. (2011) Structural basis for
- 1340 substrate recognition by *Erwinia chrysanthemi* GH30 glucuronoxylanase. *FEBS J* **278**: 2105-2116.
- 1341 Vallenet, D., Belda, E., Calteau, A., Cruveiller, S., Engelen, S., Lajus, A. et al. (2013) MicroScope--an
- 1342 integrated microbial resource for the curation and comparative analysis of genomic and metabolic data.
- 1343 *Nucleic Acids Res* **41**: D636-647.
- 1344 van der Wolf, J.M., Nijhuis, E.H., Kowalewska, M.J., Saddler, G.S., Parkinson, N., Elphinstone, J.G. et al.
- 1345 (2014) *Dickeya solani* sp. nov., a pectinolytic plant-pathogenic bacterium isolated from potato (*Solanum*
- 1346 *tuberosum*). *IJSEM* **64**: 768-774.
- 1347 Varani, A.M., Siguier, P., Gourbeyre, E., Charneau, V., and Chandler, M. (2011) ISsaga is an ensemble of
- 1348 web-based methods for high throughput identification and semi-automatic annotation of insertion
- 1349 sequences in prokaryotic genomes. *Genome Biol* **12**: R30.
- 1350 Wandersman, C., Delepelaire, P., Letoffe, S., and Schwartz, M. (1987) Characterization of *Erwinia*
- 1351 *chrysanthemi* extracellular proteases: cloning and expression of the protease genes in *Escherichia coli*. *J*
- 1352 *Bacteriol* **169**: 5046-5053.
- 1353 Wang, H.C., Minh, B.Q., Susko, E., and Roger, A.J. (2017) Modeling Site Heterogeneity with Posterior Mean
- 1354 Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Syst Biol*.
- 1355 Wheeler, T.J., and Eddy, S.R. (2013) nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**:
- 1356 2487-2489.

- 1357 Willett, J.L., Ruhe, Z.C., Goulding, C.W., Low, D.A., and Hayes, C.S. (2015) Contact-Dependent Growth
1358 Inhibition (CDI) and CdiB/CdiA Two-Partner Secretion Proteins. *J Mol Biol* **427**: 3754-3765.
- 1359 Wolanski, M., Donczew, R., Zawilak-Pawlik, A., and Zakrzewska-Czerwinska, J. (2014) oriC-encoded
1360 instructions for the initiation of bacterial chromosome replication. *Front Microbiol* **5**: 735.
- 1361 Yutin, N., Puigbo, P., Koonin, E.V., and Wolf, Y.I. (2012) Phylogenomics of prokaryotic ribosomal proteins.
1362 *PLoS One* **7**: e36972.
- 1363 Zhang, J.X., Lin, B.R., Shen, H.F., and Pu, X.M. (2013) Genome Sequence of the Banana Pathogen *Dickeya*
1364 *zeae* Strain MS1, Which Causes Bacterial Soft Rot. *Genome announcements* **1**.
- 1365 Zhang, Y., Fan, Q., and Loria, R. (2016) A re-evaluation of the taxonomy of phytopathogenic genera *Dickeya*
1366 and *Pectobacterium* using whole-genome sequencing data. *Syst Appl Microbiol.* **39**(4):252-259.
- 1367 Zhou, J., Zhang, H., Wu, J., Liu, Q., Xi, P., Lee, J. et al. (2011a) A novel multidomain polyketide synthase is
1368 essential for zeamine production and the virulence of *Dickeya zeae*. *Mol Plant Microbe Interact* **24**: 1156-
1369 1164.
- 1370 Zhou, J., Cheng, Y., Lv, M., Liao, L., Chen, Y., Gu, Y. et al. (2015) The complete genome sequence of
1371 *Dickeya zeae* EC1 reveals substantial divergence from other *Dickeya* strains and species. *BMC*
1372 *Genomics* **16**: 571.
- 1373 Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., and Wishart, D.S. (2011b) PHAST: a fast phage search tool.
1374 *Nucleic Acids Res* **39**: W347-352.

1375 **Figure legends**

1376

1377 **Figure 1. Genomic organisation of *Dickeya aquatica* 174/2^T chromosome**

1378 The chromosome is represented as a wheel and the origin of replication (OriC) and terminus (Ter)
1379 are indicated. The circles from outside to inside represent protein-coding sequences (CDS) on the
1380 forward strand, CDS on the reverse strand, distribution of ribosomal RNA operons (four on the right
1381 replichore and three on the left replichore), distribution of tRNA genes, and then distribution of
1382 ncRNA. The blue areas correspond to phage elements detected using PHAST. The thin black lines
1383 correspond to four CRISPR arrays and the red areas represent the ten genomic islands predicted
1384 using IslandViewer. The next circle (black) indicates the GC content and the central circle
1385 (green/purple) shows the GC-skew. The window size of the GC content and GC-skew is 100
1386 nucleotides. Figure 1 was drawn using Gview <https://server.gview.ca/>.

1387

1388 **Figure 2. Phylogeny and proteome comparison of *Dickeya***

1389 Maximal likelihood phylogeny (left) of 1,341 *Dickeya* single copy protein families (50 sequences,
1390 414,696 amino acid positions). The tree was computed with IQ-TREE with the LG+C60 model and
1391 rooted using *Serratia* ATCC 39006. Numbers associated to branches correspond to ultrafast
1392 bootstrap values. The scale bar corresponds to evolutionary distance (i.e. the average number of
1393 the substitutions inferred per site). The table (right) corresponds to the S_{AB} association coefficient
1394 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
1395 protein families present in strain A, N_B is the number of protein families present in strain B and N_{AB}
1396 is the number of protein families shared by strain A and strain B. This coefficient ranged from 0
1397 when both strains do not share any gene family to 100 when all the families present in strain A
1398 were also present in strain B. The figure was generated using Evolview (He et al., 2016).

1399

1400 **Figure 3. Virulence of *D. aquatica* on potato, chicory, cucumber and tomato.** Bacterial
1401 cultures were grown in M63G minimal medium (M63 + 0.2% w/v glucose) and diluted to a given
1402 OD₆₀₀ depending on the host: 0.2 (chicory) or 1 (potato, cucumber and tomato). For chicory, 5 µL

1403 of bacterial suspension were injected into a 2 cm incision at the center of the leaf. For potato,
1404 cucumber and tomato, 5, 100 or 200 μ L of bacterial suspension were injected into the vegetable,
1405 respectively. Plants were incubated at 30°C with 100% humidity for 18 h (chicory) or 42 h (potato,
1406 cucumber and tomato). **A)** Picture of representative specimens of infected plants after incubation.
1407 Note that the rotten area was removed for potato. **B)** Quantification of the soft rot mass. Data is
1408 represented as mean \pm SD of 6 replicates. No soft rot symptoms were detected after 78 h for
1409 pineapple and kiwi fruits infected with 200 μ L of bacterial suspension at OD₆₀₀ 1.

1410

1411 **Figure 4. *D. aquatica* stress resistance.** Bacteria were cultured at 30°C in 96 well plates using
1412 M63G (M63 + 0.2% w/v glucose) pH 7.0 as minimal medium. Bacterial growth (OD_{600nm}) was
1413 monitored for 48 h using an Infinite® 200 PRO - Tecan instrument. A and B) Resistance to
1414 osmotic stress was analysed using M63G enriched in 0.05 to 0.5 M NaCl (abscissa) and growth
1415 rates (ordinate) were determined. C) Resistance to oxidative stress was analysed in the same
1416 medium by adding H₂O₂ concentrations ranging from 25 to 200 μ M (abscissa). The lag time
1417 (ordinate) is represented instead of the growth rate because after the degradation of H₂O₂ by
1418 bacterial catalases, the growth rates are similar. D) The pH effect on growth rate (ordinate) was
1419 analyzed using the same M63G medium buffered with malic acid at different pH ranging from 3.7
1420 to 7.0 (abscissa).

1421

1422 **Figure 5. Twitching motility induced under acidic condition in *Dickeya* species.** Colony
1423 morphologies of various *Dickeya* strains grown on M63G pH 7.0 and M63G buffered with malic
1424 acid at pH 5 in agar plates. The strains corresponding to colony numbers and their twitching
1425 phenotypes are indicated on the right.

1426

1427 **Figure 6. Distribution of plant cell wall degrading enzymes in *Dickeya*.** The type of cell wall
1428 degrading enzyme and the number of homologues of each enzyme are indicated for each of the
1429 strains. The phylogeny on the left corresponds the core-pf ML tree. The figure was generated using

1430 Evolview (He et al., 2016).

1431

1432 **Figure 7: The core-, pan-, versatile-, and persistent-proteomes of *Dickeya* genus**

1433 The delineation of core-, pan-, versatile-, and persistent-proteomes was based on the taxonomic
1434 distribution of the 11,566 protein families identified by analyzing the 49 proteomes of *Dickeya*.
1435 The core-proteome is defined by the protein families present at least in one copy in all 49 *Dickeya*
1436 proteomes, the pan-proteome is defined by the 11,566 protein families, while the versatile- and the
1437 persistent-proteomes are defined as the protein families present in less than 10% and in more than
1438 90% of the 49 *Dickeya* proteomes, respectively.

1439 **(A) Distribution 11,566 protein families across the 49 *Dickeya* proteomes.**

1440 The 3,452 protein families present in a single *Dickeya* proteome (i.e. strain specific families) are on
1441 the left of the x-axis, while the 1,604 protein families defining the core-proteome are on the right of
1442 the x-axis.

1443 **(B) Estimation of the *Dickeya* core and pan-genomes**

1444 The graph shows the estimated sizes of the core- and pan-proteomes of *Dickeya* according to the
1445 number of considered strains. The curves were computed by calculating the core- and the pan-
1446 proteomes for an increasing number of strains randomly selected among the 49 *Dickeya* strains
1447 (100 replicates at each point). When all the 49 *Dickeya* strains were considered, the core- and the
1448 pan-proteomes encompassed 1,604 and 11,566 protein families.

1449

1450 **Figure 8. Evolution of protein family repertoires along the *Dickeya* phylogeny**

1451 The number of protein family repertoires, gains, and losses inferred by COUNT are mapped on the
1452 topology of the core-pf reference phylogeny of *Dickeya*. Values mapped above the branches
1453 correspond to gains / losses, while values below branches correspond to the number of protein
1454 families inferred. Considering the alternative placement of *D. solani* as sister-group of *D.*
1455 *dianthicola* provided very similar results (see Supplementary Table S9).

1456

1457 **Supplementary materials**

1458

1459 **Supplementary Figure S1: Maximum Likelihood phylogenetic tree of *Dickeya* strains based**
1460 **on the Fr-protein supermatrix gathering 51 ribosomal proteins** (LG+C60, 58 sequences, 6,295
1461 amino acid positions).

1462 Numbers at branch correspond to bootstrap values (100 replicates of the original data set). The
1463 scale bar indicates the average number of substitutions per site. The figure was generated using
1464 Evolvview (He et al., 2016).

1465

1466 **Supplementary Figure S2: Phylogenetic position of “*Serratia* sp. ATCC 39006” based on**
1467 **LSU and SSU rDNA trees**

1468 ML trees of the *Dickeya* genus inferred with a collection of (A) 947 SSU rDNA and (B) 125 LSU
1469 rDNA sequences retrieved from public databases. The trees were rooted using *Pectobacterium*
1470 and *Serratia* sequences. The scale bars represent the estimated average number of substitution
1471 per site. Numbers at nodes represent ultrafast bootstrap values (A) and bootstrap values (B).
1472 *Pectobacterium* sequences are in pink, *Serratia* sequences in purple, and *Dickeya* sequences in
1473 black. Worth to note, LSU and SSU rDNA sequences from *Serratia* sp. 39006 robustly branch with
1474 *Dickeya* sequences and not within the *Serratia* genus. The trees were drawn using the iTOL
1475 webserver (Letunic and Bork, 2016).

1476

1477 **Supplementary Figure S3 - Single gene phylogenies**

1478 Maximum Likelihood phylogenetic trees of 912 proteins of interest. Numbers associated with each
1479 branch correspond to ultrafast bootstrap values (1000 replicates of the original data set). The scale
1480 bars indicate the average number of substitutions per site. For each tree, the name and the
1481 annotation of the seed is provided in red.

1482

1483 **Supplementary Figure S4 – Non-metric multi-dimensional scaling (NMDS) plot of *Dickeya***
1484 **genomes according to their gene content.** A Bray distance similarity matrix was calculated
1485 based on the presence/absence profiles of 8,115 gene families (present in at least 2 genomes) for

1486 the 49 analyzed genomes and used to generate NMDS coordinates for each strain. The shorter
1487 distance linking two genomes indicates higher similarities between these genomes. Genomes from
1488 different species are indicated in different colors.

1489

1490 **Supplementary Figure S5: Phylogenies and genetic organisation of *peIAED* and *pehVWX***
1491 **clusters.**

1492 Maximum likelihood trees of *peIAED* and *pehVWX* clusters together with genetic organisation of
1493 these clusters are shown. Numbers associated with each branch correspond to ultrafast bootstrap
1494 values (1000 replicates of the original data set). The scale bar indicates the average number of
1495 substitutions per site. For each tree, the name and the annotation of the seed is provided in red.

1496

1497 **Supplementary Figure S6: The number of shared protein families between and within**
1498 ***Dickeya* species.** Figures were generated using the package UpSetR (Conway et al. 2017).

1499

1500 **Supplementary Table S1: Phenotypic differentiation of species within the genus *Dickeya*.**

1501 Table showing the phenotypic features of *Dickeya* species based on Samson et al. (2005),
1502 Parkinson et al. (2014), van der Wolf et al. (2014) and Tian et al. (2016).

1503

1504 **Supplementary Table S2: Genomic features of the *Dickeya* strains with completely**
1505 **sequenced genomes**

1506

1507 **Supplementary Table S3: List of the 49 *Dickeya* genomes used in this study indicating the**
1508 **plant host or habitat and the geographical origin of each strain**

1509

1510 **Supplementary Table S4: Distribution of genes with a role in pathogenicity or in adaptation**
1511 **to plant niche in *Dickeya* species**

1512 Genes of *Dickeya* species were considered as present if identity of the encoded protein was higher
1513 than 60% of full-length amino acids sequence. If local alignments were too short with regard to the

1514 length of similar sequences, we performed a nucleotide BLAST on full-length DNA sequences with
1515 similar thresholds (10^{-5} e-value, 80% identity of full length sequence). This allowed us to eliminate
1516 false genes. Note that because some genomes represent only drafts, false negatives may occur.

1517

1518 **Supplementary Table S5: Protein family distribution within proteomes from *Dickeya*,**
1519 ***Pectobacterium* and *Serratia*.**

1520 Considered strains were *Dickeya aquatica* 174/2, *Dickeya dadantii* 3937, *Dickeya paradisiaca*
1521 NCPPB 2511, *Dickeya solani* IPO 2222, *Dickeya zeae* EC1, *Pectobacterium atrosepticum*
1522 SCRI1043, *Pectobacterium carotovorum* subsp *carotovorum* PC1, *Pectobacterium wasabiae*
1523 WPP163, *Serratia fonticola*, *Serratia liquefaciens* ATCC27592, *Serratia marcescens* FGI94,
1524 *Serratia plymuthica* AS9, *Serratia proteamaculans* 568 and *Serratia* sp ATCC39006. Families were
1525 built using SILIX version 1.2.9 with 60% of sequence identity and 80% of sequence coverage.

1526

1527 **Supplementary Table S6: Results of the AU tests performed on core-pf and rprots**
1528 **topologies using the 1,341 core-pf and the 51 rprots.**

1529

1530 **Supplementary Table S7: Distribution of Transcription factor-binding sites in *D. aquatica***
1531 **174/2 genome.** Binding sites for 56 transcriptional regulators were predicted in *D. aquatica* using
1532 nhmmer (Wheeler and Eddy, 2013).

1533

1534 **Supplementary Table S8: Protein family distribution within the proteomes of the 49 *Dickeya***
1535 **strains.**

1536 Families were computed with SILIX version 1.2.9 using 60% of sequence identity and 80% of
1537 sequence coverage.

1538

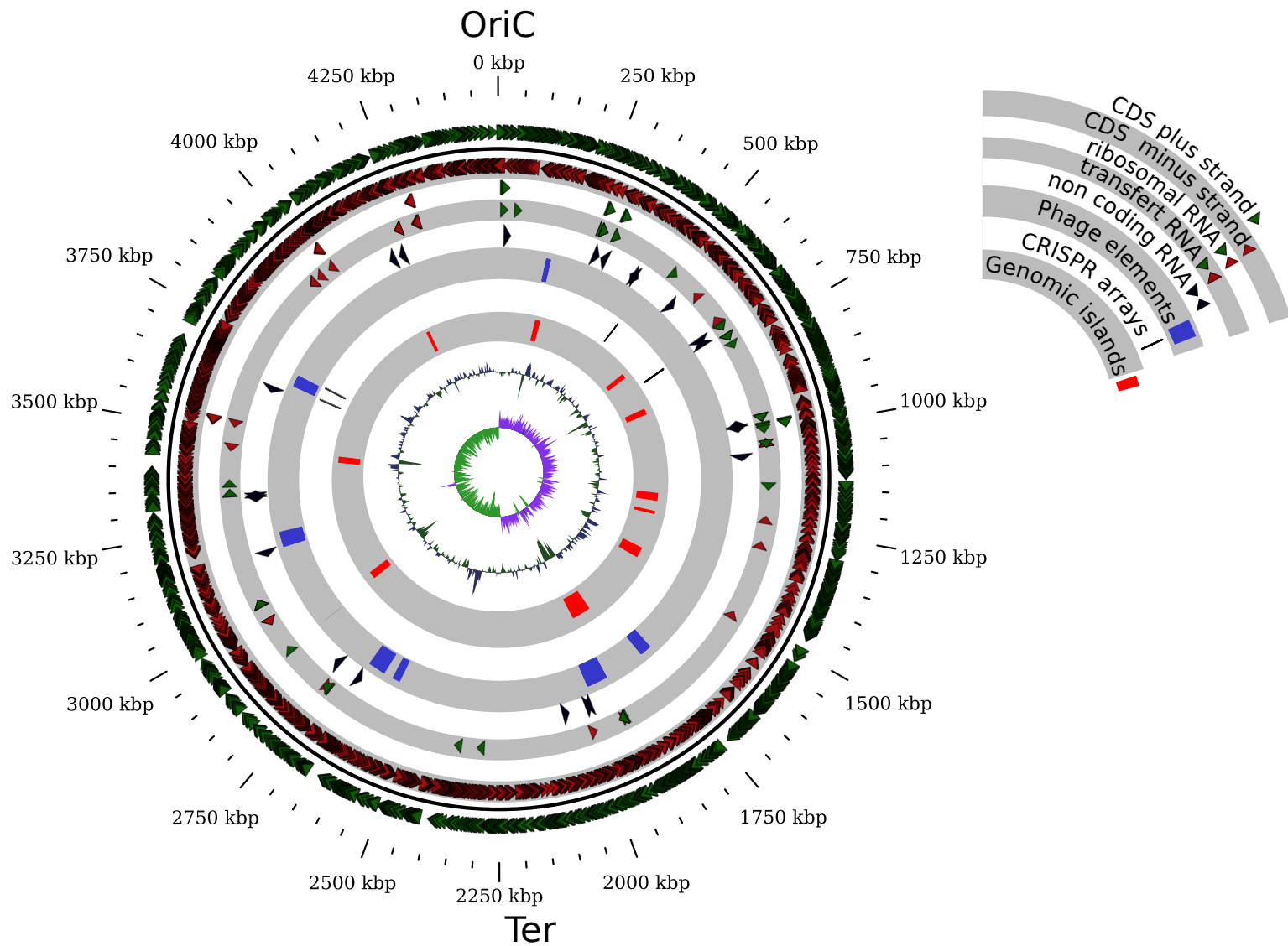
1539 **Supplementary Table S9: Comparison of inferences of ancestral gene repertoires in *Dickeya***
1540 **considering the two possible alternative positions for *D. solani***

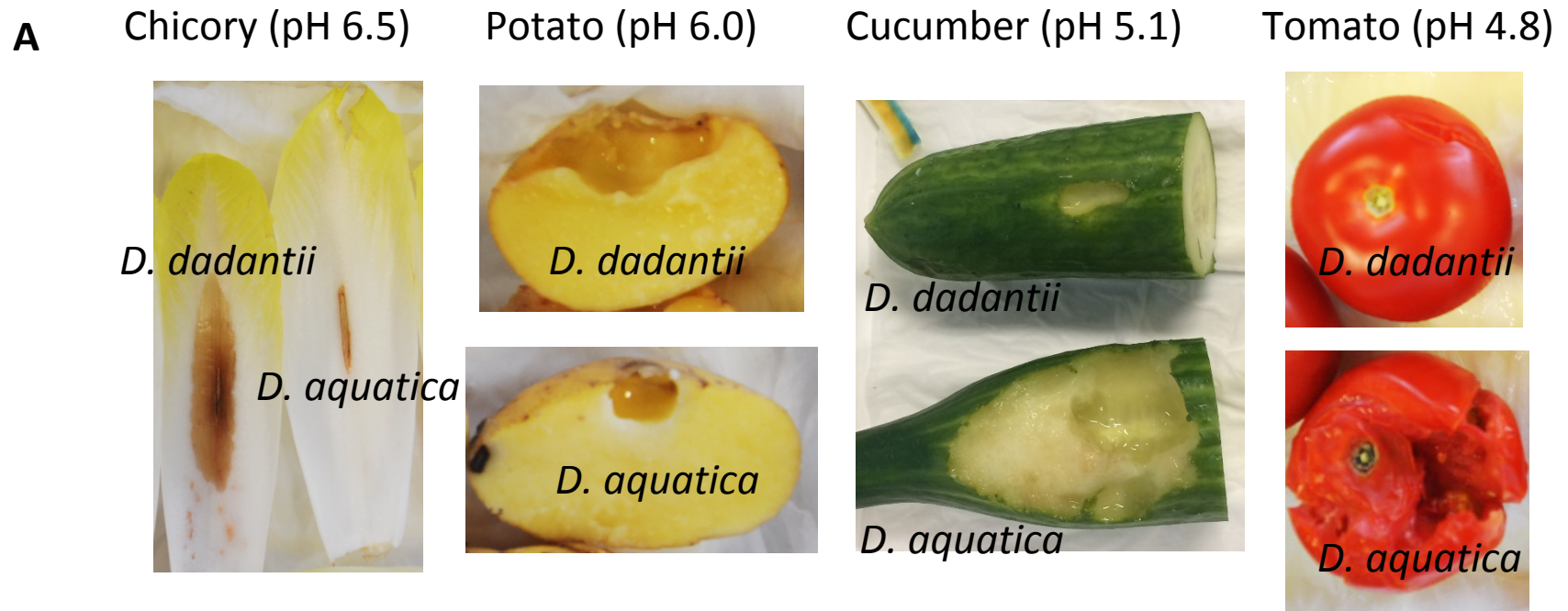
1541

1542 **Supplementary Table S10: Characterization of the mobilome of *D. aquatica* 174/2 and its**
1543 **conservation in other *Dickeya* species**

1544 Genes of *Dickeya* species were considered as present if identity of the encoded protein was higher
1545 than 60% of full-length amino acids sequence. If local alignments were too short with regard to the
1546 length of similar sequences, we performed a nucleotide BLAST on full-length DNA sequences with
1547 similar thresholds (10^{-5} e-value, 80% identity of full length sequence). This allowed us to eliminate
1548 false genes. Note that because some genomes represent drafts, false negatives may occur.

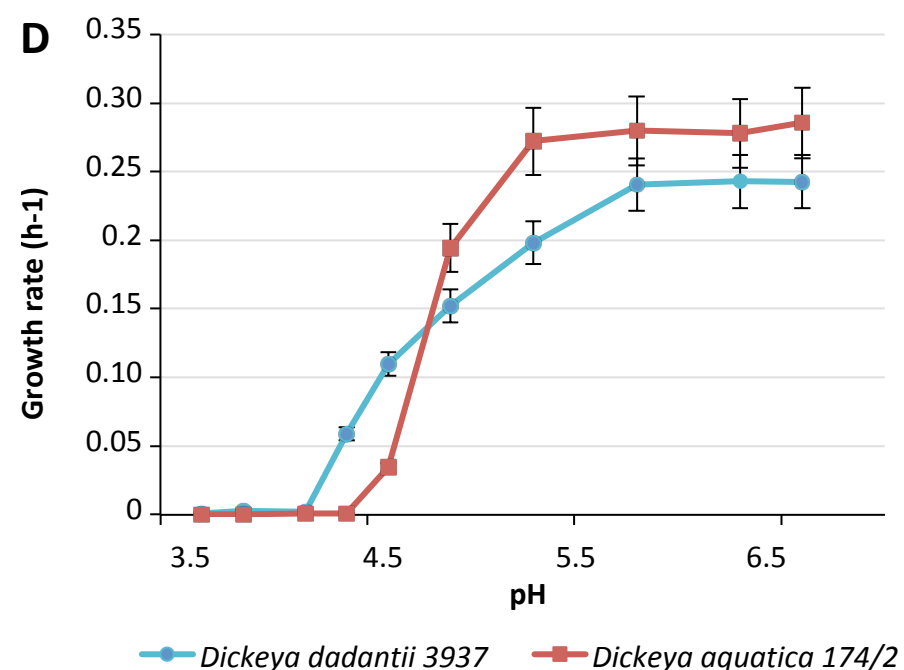
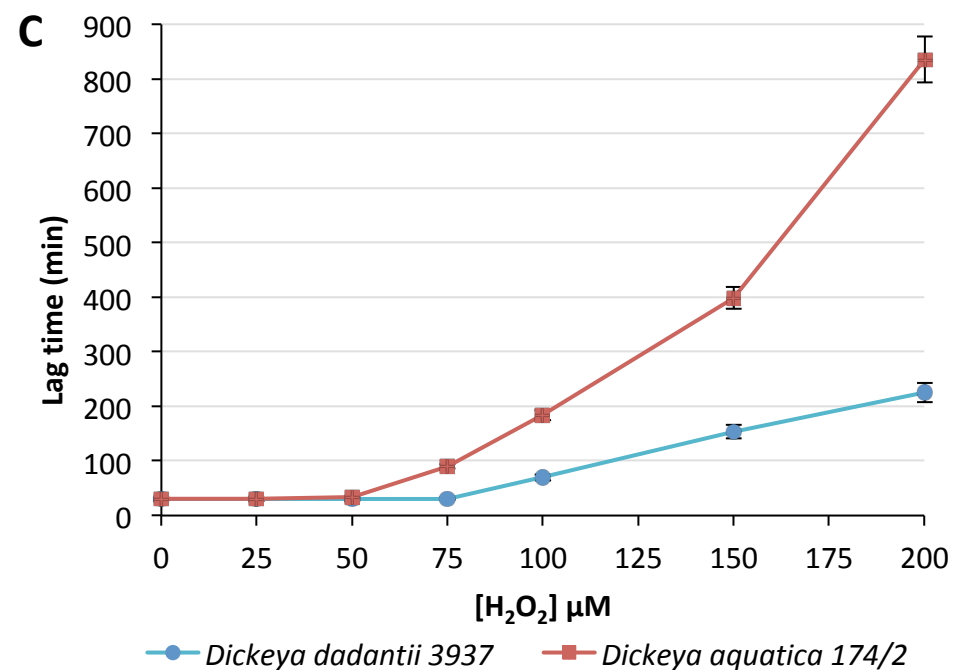
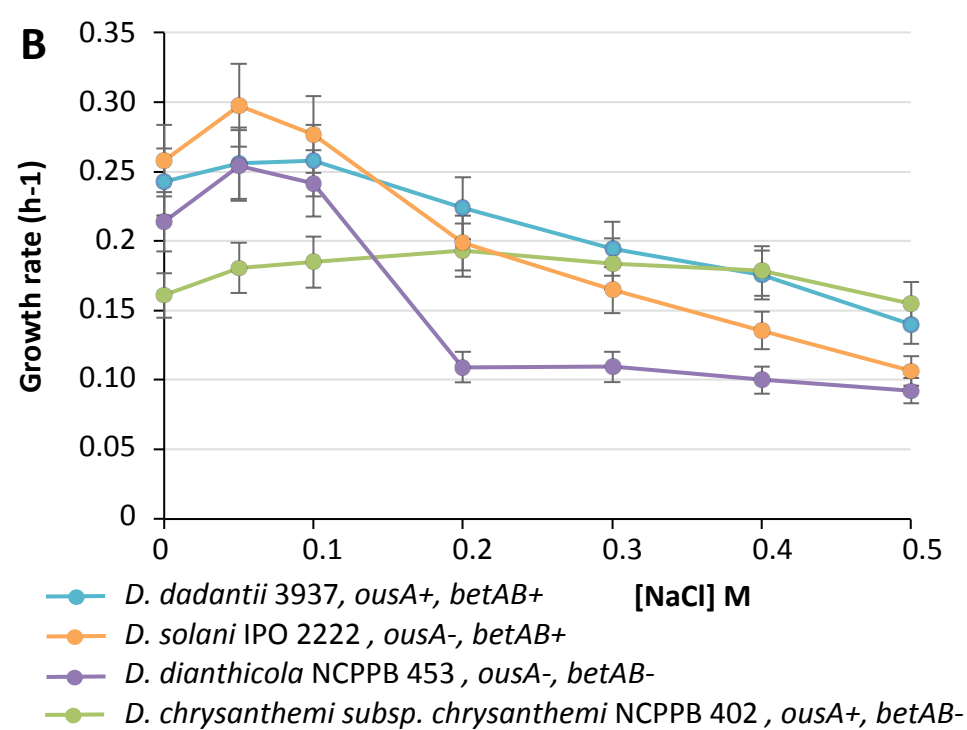
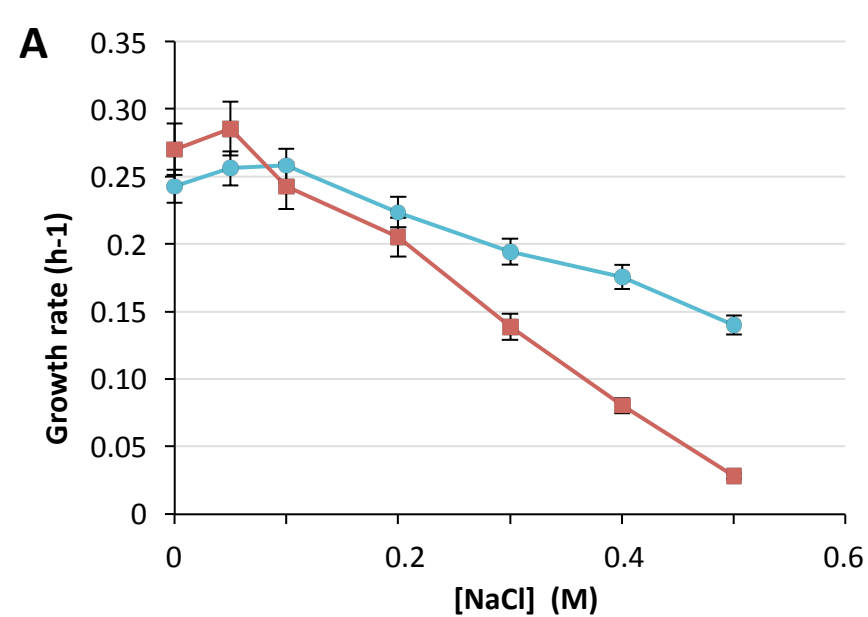
Dickeya aquatica 174/2 genome





B

Rotten mass (g)	<i>D. dadantii</i>	<i>D. aquatica</i>
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





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

