Open Access

OBM Genetics

Original Research

Diversity in the Bacterial Genus *Dickeya* Grouping Plant Pathogens and Waterways Isolates

Jacques Pédron [*], Frédérique Van Gijsegem [*]

Sorbonne Université, INRA, Institute of Ecology and Environmental Sciences-Paris (iEES-Paris), 75252 Paris cedex, France; E-Mails: jacques.pedron@upmc.fr; vangijse@agroparistech.fr

* **Correspondences:** Jacques Pédron, Frédérique Van Gijsegem; E-Mails: jacques.pedron@upmc.fr; vangijse@agroparistech.fr

Academic Editors: Ben Krause-Kyora, Sergey Pisarenko, Dmitry Kovalev and Alexander Kulichenko

Special Issue: Bacterial Genomes

OBM Genetics
2019, volume 3, issue 4
doi:10.21926/obm.genet.1904098

Received: July 25, 2019
Accepted: November 07, 2019
Published: November 26, 2019

Abstract

***Background*:** Genus *Dickeya* comprises aggressive soft rot plant pathogens with wide geographic distribution and host ranges. Ten *Dickeya* species were characterized. Seven of them (*Dickeya chrysanthemi, D. dadantii, D. dianthicola, D. fangzhongdai, D. solani, D. paradisiaca,* and *D. zeae)* group causative agents of maceration-associated diseases that impact a wide variety of crops or ornamentals as well as isolates from fresh water. The other three species (*D. aquatica, D. lacustris, D. undicola*) were recently isolated only from water sources, so far. Here, we analyzed the *Dickeya* genetic diversity in relation to species affiliation and habitats.

***Methods*:** We compared the genomes of 59 *Dickeya* strains isolated from various hosts and from different environments, determined their relatedness both at the genetic level (ANI) and their pan-genome content and carried out SiLix analysis to explore the occurrence of orthologous or species/strain-specific gene families.

***Results*:** Our study revealed significant conservation of virulence-associated genes in most *Dickeya* species (including "water-specific" ones). We also identified the genome-specific

traits of the various species and highlighted the intra-species diversities. At the species level, we observed a contrasting diversity with some species grouping highly related strains while others were much more diverse or at the limit of subdivision in separate species. The diversity was not related to diversity in habitat or geographical origin neither in the extent of the species accessory genome nor in the number of strain-specific genes.

***Conclusions*:** The genus *Dickeya* pangenome analysis did not highlight strain clustering following host/environment of isolation. The different *Dickeya* species present few specific characteristics and fewer specific gene losses, which are frequently found in specialized human pathogens, reflecting the broad host range of this genus. No "water-specific" genes were identified that would indicate possible spread of *Dickeya* via waterways and implicate irrigation water as a potential threat for economically important crops.

Keywords

Comparative genomics; pectinolytic; enterobacteria; plant pathology; soft rot; blackleg; potato

1. Introduction

Soft rot *Pectobacteriaceae* and *Dickeya* – are *Enterobacterales* responsible for considerable economic losses in several important crops and ornamental plants and were listed as one of the top ten phytopathogenic bacteria [1]. Recent severe outbreaks in potato, one of the primary food crops worldwide, resulted from the action of a cohort of bacteria belonging to different *Pectobacterium* and *Dickeya* species in a complex population dynamics history [2]. These bacteria often exhibit a very broad host range and are able to infect both monocot and dicot plants [2, 3]. Their hallmark is the ability to macerate plant tissues by disintegrating plant cell walls leading to cell lysis and liberation of the cell content. They do so by producing and secreting a battery of plant cell wall degrading enzymes (PCWDEs) that cause maceration of the plant tissue. Their virulence relies however on several other factors that allow them to adapt to environmental changes encountered *in planta* and to face the stresses produced by plant defense responses [4].

The genus *Dickeya* was formed in 2005 by the reclassification of former *Erwinia chrysanthemi* into six species as *Dickeya chrysanthemi*, *D. dadantii*, *D. diffenbachiae*, *D. dianthicola, D. zeae* and *D. paradisiaca* [5]. Subsequently, *D. dieffenbachiae* was reclassified as a subspecies of *D. dadantii* based on DNA-DNA hybridization analysis and MLSA [6]. Each of these species comprises strains isolated from various plant hosts including both dicots and monocots and does not possess real host specificity [5]. Three new *Dickeya* species have also been described: *D. aquatica* isolated from freshwaters in Europe [7], *D. solani* associated with severe outbreaks in the 2000s in potato in Europe and Israel [8, 9] and *D. fangzhongdai* isolated from diseased pear trees but also from several monocots, including orchids, in Asia and Europe and from freshwaters [3, 10, 11]. Comparative analyses of these *Dickeya* proteomes allowed Duprey et al. [12] to infer the complex evolutionary history of the genus *Dickeya*. A recent systematic search for *Dickeya* in waterways led to the characterization of two new *Dickeya* species, *D. lacustris* [13] and *D. undicola* [14] that were isolated exclusively from freshwaters or from the plant rhizosphere at pond edges so far.

Thus, the genus *Dickeya* thus currently comprises ten characterized species that group bacterial isolates from various environments including monocot versus dicot plants, diseased plants versus freshwaters. The purpose of this study was to analyze the inter- and intra-specific diversity in genus *Dickeya* with a special emphasis on the two new species *D. undicola* and *D. lacustris*, isolated from waterways.

2. Materials and Methods

2.1 Genome Collection and Annotation

We downloaded the complete and draft genomes of all *Dickeyas* available as of March 2019 from the NCBI (https://www.ncbi.nlm.nih.gov/), except *D. aquatica* 174/2, whose complete genome was downloaded from the ASAP database (http://asap.ahabs.wisc.edu/asap/home.php). Coding sequences were predicted using the Rapid Annotation Subsystem Technology (RAST) server [15] using default parameters.

2.2 Species Delimitation

Average nucleotide identity (ANI) was computed using the pyani python module [16] with the BLASTp algorithm. In silico DNA-DNA hybridization (DDH) was carried out and analyzed according to Meier-Kolthoff et al. [17], using a dedicated pipeline (http://ggdc.dsmz.de/). Species threshold was set to 96% and 70% for ANI and DDH, respectively. Genome synteny was visualized by a dot plot using the D-Genies web site ([18], http://dgenies.toulouse.inra.fr/).

2.3 Phylogenetic Analysis

In silico multilocus sequence analysis (MLSA) was performed on 1191 concatenated amino acid orthologous sequences (82093 variable sites) of the 59 genomes. *Pectobacterium atrosepticum* 21A was used as an outgroup. The clustering of orthologous sequences into homologous families was conducted using the SiLix software package [19], using an identity threshold of 70% and a minimal overlap of 80%. Orthologous sequences were aligned using the Muscle software [20] and then concatenated. SeaView (version 4.6.5), a multi-platform program for molecular phylogeny, was used to implement specified tools and options [21]. Alignments were curated using Gblocks [22] and phylogenic analysis was performed using the BioNJ algorithm [23] with 100 bootstraps value. Mega7 tool [24] was used to visualize the phylogenetic tree.

For the pangenome clustering, hierarchical clustering was performed for the pan-genome as described by Meric et al. [25]. Briefly, a presence/absence matrix (0/1) for the pangenome was constructed. Manhattan distances were calculated and used for hierarchical clustering to generate the tree. The plotted distance between two genomes revealed the proportion of protein families where their present/absent status differs. Thus, in the pan-genome hierarchical clustering analyses, protein families that were not conserved vary in their presence or absence between genomes. Protein families occurring only in a single genome (singletons) were not included in the analysis. Mega7 tool [24] was used to visualize the phylogenetic tree with the BioNJ algorithm.

2.4 Assessment of Core Genome, Pan Genome, and Species-Specific Genes

Genome-to-genome comparisons were performed by bi-directional protein-protein BLAST sequence comparison of translated open reading frames (ORFs) with a 10^{-5} e-value threshold. Genes were defined using the RAST server [15] applying default parameters and without a minimum length cut-off. Orthologous sequences were clustered into homologous families using the SiLix software package [19], with a constraint of 70% amino acid identity and 80% overlap.

3. Results and Discussion

3.1 Contrasting Diversities among Dickeya Species

To investigate the diversity of the genus *Dickeya*, we analyzed 59 *Dickeya* genomes from databases, including the 49 genomes analyzed by Duprey et al. [12], two additional genomes of recently described *D. undicola* species [14] and the genome of the new *D. lacustris* species [13]. Our panel comprised 20 fully sequenced genomes and 39 draft genomes consisting of one to nearly 200 scaffolds (Table 1). A significant number of the analyzed strains were isolated from potato (*Solanum tuberosum*) diseased tissues (19 isolates) reflecting the important economic implications of the recent outbreaks that occurred in this crop in Europe and in the USA [2]. The other strains were collected from monocots or other dicots plants, from the rhizosphere of healthy potato plants (2 isolates) or from natural freshwaters (14 isolates). Geographically, the analyzed bacteria originated from all around the world but the distribution was not uniform with half of them isolated from Europe and only one isolated from Comores islands, Africa (Table 1).

As already described [26, 12], a whole-genome MLSA analysis of the 59 studied genomes confirmed that all strains fit in one of the ten characterized *Dickeya* species except *Dickeya* sp. NCPPB569 that was proposed to be a new species [12] (Figure 1).

Based on MLSA analysis, *Dickeya* species could be further organized in different branches, clade I grouping *D. chrysanthemi*, *D. zeae* and *Dickeya* sp. NCPPB569, clade II grouping the new *D. undicola* species with *D. solani*, *D. dadantii*, *D. fangzhongdai* and *D. dianthicola*, a third clade grouping *D. lacustris/D. aquatica* clade, and the *D. paradisiaca* branch (Figure 1).

We analyzed the relatedness between strains of similar species by calculating the two-by-two average nucleotide identities (ANI) for each *Dickeya* genome (Figure 1) and highlighted contrasting situations. Bacteria belonging to *D. aquatica*, *D. dianthicola* and *D. undicola* group shared around 99% or more ANI among them (Figure 1), revealing very high sequence conservation. However, only three genomes of *D. aquatica* and *D. undicola* are available until now and they may not reflect the complete diversity of these two species. The high conservation observed in these three species does not correlate to similar geographical origins or similar hosts. Indeed, if all *D. dianthicola* strains (except one) were isolated from diseased potato, they originate from locations all over the world (Table 1). Similarly, while the three *D. undicola* strains were isolated from freshwaters, they were from either France or Malaysia.

Table 1 General features of the different *Dickeya* strains/genomes analyzed in this study.

Strains	Genome length	# scaffolds	# of CDS	Isolated from	Geographic origin	Year of isolation
D. aquatica 174/2(T)	4.50	complete	4535	river water	UK	
D. aquatica DW_0440	4.34	47	4353	river water	Finland	
D. aquatica CSL RW240	4.39	26	4398	river water	UK	
D. chrysanthemi 1591	4.81	complete	4604	*Zea mays*	USA	1957
D. chrysanthemi L11	4.77	144	4691	lake water	Malaysia	2014
D. chrysanthemi NCPPB3533	4.73	23	4583	*Solanum tuberosum*	USA	
D. chrysanthemi NCPPB402 (T)	4.70	21	4606	*Chrysanthemum morifolium*	USA	1956
D. chrysanthemi NCPPB516	4.62	8	4542	*Parthenium argentatum*	Denmark	1957
D. dadantii 3937	4.92	complete	4271	*Saintpaulia ionantha*	France	1977
D. dadantii DSM18020 (T)	5.00	complete	4732	*Pelargonium capitatum*	Comoros	1960
D. dadantii subsp. *dieffenbachiae* NCPPB2976 (T)	4.82	14	4652	*Dieffenbachia* sp.	USA	1977
D. dadantii NCPPB3537	4.81	1	4508	*Solanum tuberosum*	Peru	1987
D. dianthicola DE440	4.87	55	4792	*Solanum tuberosum*	USA	2016
D. dianthicola GBBC2039	4.80	32	4607	*Solanum tuberosum*	Belgium	
D. dianthicola IPO980	4.84	27	4493	*Solanum tuberosum*	Netherlands	1991
D. dianthicola ME23	4.91	complete	4790	*Solanum tuberosum*	USA	
D. dianthicola NCPPB3534	4.87	1	4663	*Solanum tuberosum*	Netherlands	1987
D. dianthicola NCPPB453 (T)	4.68	1	4477	*Dianthus caryophyllus*	UK	1956

D. dianthicola RNS04.*9*	4.72	complete	4567	*Solanum tuberosum*	France	2004
D. dianthicola SS70	4.80	62	4665	*Solanum tuberosum*	Pakistan	2017
D. dianthicola WV516	4.91	103	4795	*Solanum tuberosum*	USA	2016
D. fangzhongdai B16	4.89	53	4580	*Phalaenopsis orchid*	Slovenia	2010
D. fangzhongdai DMS101947 (T)	5.03	complete	4638	*Pyrus pyrifolia*	China	2009
D. fangzhongdai M005	5.11	138	4788	waterfall	Malaysia	2013
D. fangzhongdai M074	4.95	145	4631	waterfall	Malaysia	2013
D. fangzhongdai MK7	4.93	21	4589	river water	Scotland	
D. fangzhongdai NCPPB3274	4.99	1	4656	*Aglaonema*	St Lucia	1983
D. fangzhongdai ND14b	5.05	complete	4616	waterfall	Malaysia	2013
D. fangzhongdai PA1	4.98	complete	4548	*Phalaenopsis* orchid	China	2011
D. fangzhongdai S1	4.94	51	4652	*Phalaenopsis* orchid	Slovenia	2012
D. lacustris S29 (T)	4,31	118	4377	lake water	France	2017
D. paradisiaca 703	4.68	complete	4526	*Solanum tuberosum*	Australia	
D. paradisiaca NCPPB2511 (T)	4.63	1	4548	*Musa paradisiaca*	Colombia	1970
D. solani Ds0432	4.92	complete	4498	*Solanum tuberosum*	Finland	2004
D. solani F012	4.88	25	4509	*Solanum tuberosum*	Russia	2010
D. solani IFB0099	4.93	complete	4516	*Solanum tuberosum*	Poland	2005
D. solani IFB0223	4.94	complete	4516	potato rhizosphere	Germany	2005
D. solani IFB0221	4.88	38	4518	potato rhizosphere	Germany	2005
D. solani IPO2222 (T)	4.92	complete	4530	*Solanum tuberosum*	Netherlands	2007

D. solani PPO9019	4.96	complete	4641	muscari	Netherlands	2006
D. solani PPO9134	4.87	22	4546	hyacinth	Netherlands	2008
D. solani RNS05.1.2A	4.99	37	4718	*Solanum tuberosum*	France	2005
D. solani RNS08.23.3.1.A	4.92	complete	4536	*Solanum tuberosum*	France	2008
D. undicola 2B12	4.35	77	4178	lake water	Malaysia	2014
D. undicola FVG01	4.61	178	4525	water	France	2017
D. undicola FVG10	4.54	202	4484	water	France	2016
Dickeya sp. NCPPB569	4.22	6	4235	*Saccharum*	Australia	
Dickeya sp. Secpp_1600 *	5.11	complete	4713	radish	China	2016
D. zeae 586	4.82	complete	4515	Philodendron Schott	USA	
D. zeae CSL_RW192	4.70	4	4587	river water	UK	
D. zeae DZ2Q	4.65	26	4456	*Oryza sativa*	Italy	
D. zeae EC1	4.53	complete	4260	*Oryza sativa*	China	1997
D. zeae MK19	4.67	4	4494	river water	UK	
D. zeae MS1	4.75	58	4589	*Musa*	China	2009
D. zeae MS2	4.74	complete	4529	*Musa paradisiaca*	China	2014
D. zeae NCPPB2538 (T)	4.56	7	4380	*Zea mays*	USA	1970
D. zeae NCPPB3531	4.63	2	4354	*Solanum tuberosum*	Australia	
D. zeae NCPPB3532	4.56	1	4390	*Solanum tuberosum*	Australia	
D. zeae ZJU1202	4.59	188	4417	*Oryza sativa*	China	2002

* Our analyses showed that the Secpp_1600 strain belongs to *D. fangzhongdai*.

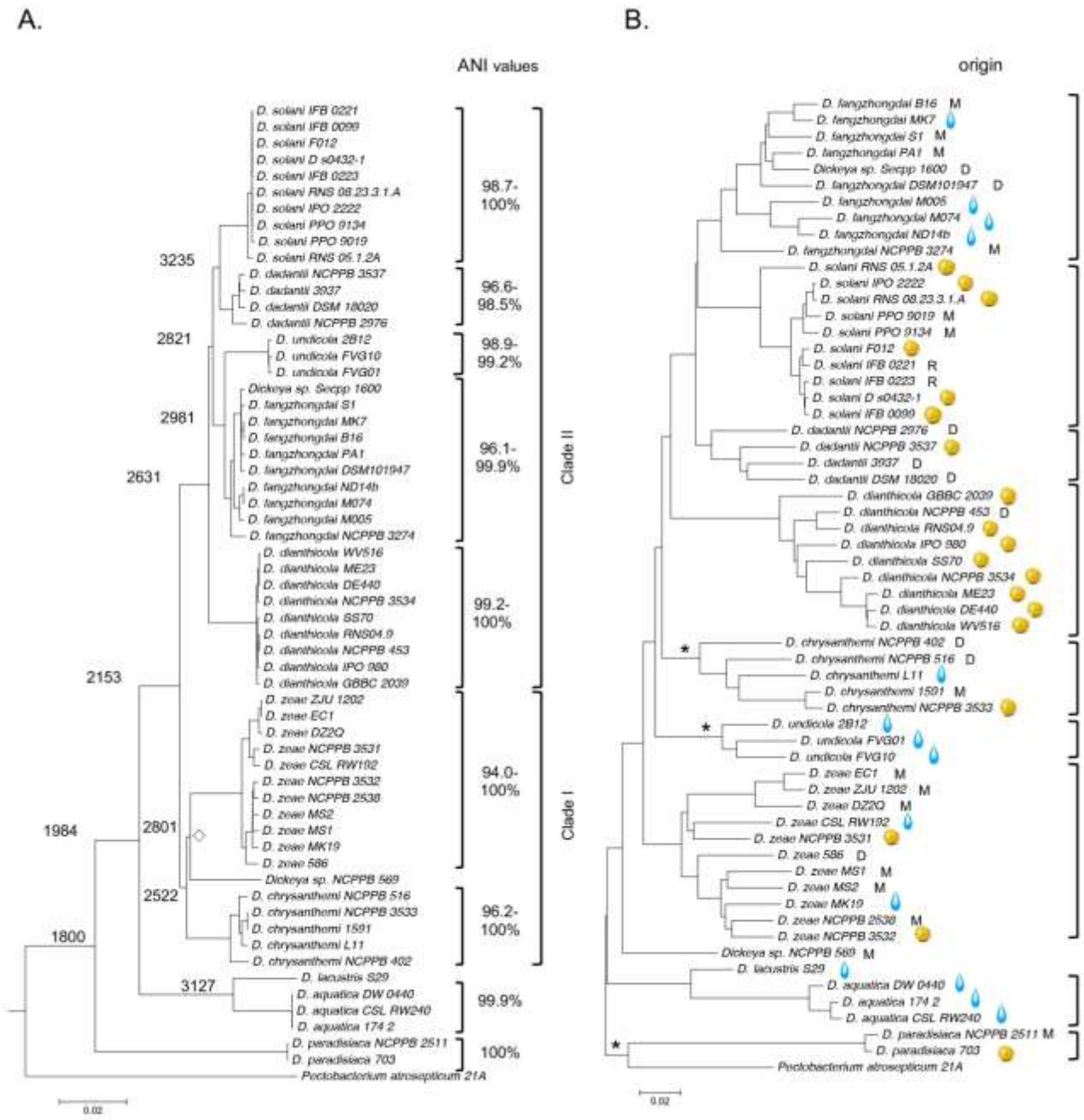

Figure 1 Phylogeny of the *Dickeya* genus. A. Phylogenic tree built up from the concatenated sequences of 1191 homologous protein sequences (82093 variable sites). One hundred bootstrap replicates were performed to assess the statistical support of each node. Only one node (labelled ◇) was below 100% (bootstrap value 49%). *Pectobacterium atrosepticum* 21A was used as an outgroup. The numbers of protein families common to all genomes of the different clades/nodes are written in the left part. ANI values ranges for each species are given. B. Pangenome tree: distance was calculated from a presence/absence matrix of the pangenome (see Materials and Methods). Origins: M: monocot; D: dicot; R: potato rhizosphere; 🥔: potato; 💧: water. The stars indicate differences between the phylogenic and pangenome trees.

Similarly, no link was observed between plant host origin and genome variability. For instance, as already reported [27, 28, 29], most *D. solani* strains exhibiting very high ANI values above 99.9%, were isolated either from potato or from ornamentals, pointing to a clonal origin and possible contamination of potato from nearby fields planted with ornamentals [30]. However, one *D. solani* genome, RNS05.1.2 was found to be more divergent with an ANI value of about 98% with the other *D. solani* genomes [28, 29]. This divergence is not related to a different host or geographical origins because, like other *D. solani* strains of our panel, this strain was isolated from potato in France [28].

Strains grouped in the *D. fangzhongdai*, *D. zeae*, *D. dadantii* and *D. chrysanthemi* species were more diverse and, in some cases, fell below the limits of species definition, now generally recognized as 95%–96% for average identity value (ANI) and 70% for digitally derived DNA-DNA hybridization (dDDH) [31, 32]. While all *D. dadantii* and *D. chrysanthemi* strains harbored more than 96/70% ANI/dDDH values between genomes of the same species, these values were at the limits of species definition for some *D. fangzhongdai* or even fell below these limits for *D. zeae* genomes. In both cases, a further subcladal organization was observed in these species and the distribution of the analyzed genomes could be assigned to three branches, two subclades grouping genomes that share ANI values ranging from 97% to 99%, and one divergent strain (NCPPB3274 in *D. fangzhongdai* and 586 in *D. zeae*) (Figure 1) [33]. The ANI/dDDH values of genomes belonging to the two *D. zeae* subclades were as low as 94/58.2% (Table S1), indicating the separation of these two lineages into two distinct subspecies or even separate species. Further phenotypic characterizations and analysis of more genomes are needed to solve this ambiguity.

3.2 Dickeya Species Share a Large Panel of Genes

In order to further address the relationships between the different *Dickeya* species, we carried out a pan-genome analysis of this genus and compared the protein-coding sequences of the 59 genomes using the SiLix gene family clustering tool. Proteins were classified as homologous to others in a given family if the amino acid identities were above 70% with 80% minimal overlap. Bacterial pan-genomes could be divided into three gene classes as i) the core genome that groups protein families shared by all members of the analyzed genus/species, ii) the specific genome gathering protein families unique to one genome or a genome group, iii) the accessory genome including protein families present in more than one but not all genomes of a group. The genus *Dickeya* core genome comprises 1800 genes representing nearly 40% (37.5% to 43%) of a single genome content (Figure 1). These common genes encompass most of the genes shown to be involved in virulence (for a review, [4]). The hallmark of virulence of the soft rot causing *Dickeya* is the production and secretion of a battery of plant cell wall degrading enzymes, such as pectinases, cellulase, and proteases. The out type 2 secretion system (T2SS) drives the secretion of nine pectate (PelABCDEILNZ), one pectin (Pnl), one or two rhamnogalacturonate (RhiE, RhiEbis) lyases and various enzymes that strip off the pectin decorations (pectin methylesterases PemA and PemB, acetyl esterase PaeY) as well as the cellulase CelZ. As recently described [12], the corresponding encoding genes are largely conserved in the *Dickeya* genus (Table 2). Like all members of clade II, *D. undicola* possesses all PCWDE-encoding genes except the gene encoding the second rhamnogalacturonate lyase. A notable exception is PelA whose gene is truncated in all *D. dianthicola* genomes. Likewise, *D. lacustris* harbors almost the same PCWDE battery as its

closest relative *D. aquatica*, except that it has only one copy of the *pelD/E* genes and possesses the Pel10 encoding gene otherwise present only in *D. fangzhongdai* and some *D. chrysanthemi* strains. *D. paradisiaca* and *Dickeya* sp. 569 were found to be less equipped, and they both lacked PelA, PelI, Rhi, and PemB. In addition, PelZ, and PnlG were absent in *Dickeya* sp. 569 while *D. paradisiaca* harbored only one copy of the PelD/E isoenzymes and lacked PaeY acetyl esterase.

Table 2 Conservation of genes encoding pectinolytic enzymes.

	pelA	***peD/E***	***pelB/C***	***pelI***	***pelL/N***	***pelZ***	***rhiE***	***rhiE bis***	***pel10***	***pnlG***	***pemA***	***pemB***	***paeY***
D. dadantii	1	2	2	1	2	1	1	no	no	1	1	1	1
D. solani	1	2	2	1	2	1	1	no	no	1	1	1	1
D. dianthicola	trunc	2	2	1	2	1	1	var	no	var	1	1	1
D. fangzhongdai	1	2	2	1	2	1	1	no	1	1	1	1	1
D. undicola	1	2	2	1	2	1	1	no	no	1	1	1	1
D. chrysanthemi	1	2	2	1	2	1	var	1	var	no	1	no	1
D. zeae	1	2	2	1	2	1	1	1	no	var	1	1	1
Dickeya sp. 569	1	1	2	1	2	1	1	1	1	no	1	no	1
D. lacustris	1	2	2	1	2	1	1	1	no	1	1	no	1
D. aquatica	no	1 div	2	no	2	1	no	no	no	no	1	no	no
D. paradisiaca	no	1	2	1	1	no	no	no	no	1	1	no	1

Trunc: truncated, div: divergent, var: variable.

All *Dickeya* species except *D. paradisiaca* were found to possess a Prt T1SS that allows the secretion of up to four metalloproteases (PrtABCG). The genes encoding the three PrtABC proteases are also largely shared while the *prtG* gene is absent in *D. dianthicola*, *D. undicola*, *D. chrysanthemi* and *Dickeya* sp. NCPPB569 (Table 3).

Four other protein secretion systems (T3SS to T6SS) might be present in gram-negative bacteria where they are involved in interactions with eukaryotic cells or in bacterial intra- and inter-competitions [34]. All four TSS were present in the *D. dadantii* 3937 model strain but T5SS comprised very long proteins truncated in the draft genomes and difficult to tackle. So, we excluded T5SS from our study. The different *Dickeya* species possess variable sets of T3SS, T4SS, and T6SS; *D. paradisiaca* and *Dickeya.* sp. NCPPB569 do not have any; *D. lacustris* and *D. aquatica* possess only T3SS, and for the other species, the presence and number of these TSS are variable (Table 3).

Upon plant colonization, bacteria encounter various stressful environmental conditions like the low pH in apoplast, as well as osmotic and oxidative stresses. The *D. dadantii* 3937 model strain is well equipped to cope with these stresses [35]. Like many other bacteria, it responds to oxidative stress by producing antioxidant enzymes like catalases (KatE, KatG), superoxide dismutases (SodA, SodC) and alkylhydroperoxide reductase (AhpCF) as well as repair systems like the peptide

methionine sulfoxide reductase (MsrA), an enzyme repairing oxidative damage caused to proteins [36], or the Suf and Isc systems, involved in the repair of damaged Fe/S clusters containing proteins [37, 38]. It also produces the ROS scavenging pigment indigoïdine, It can also resist antibacterial peptides produced by plants through the Sap system and detoxify nitric oxide produced by plants during infection with the nitric oxide dioxygenase HmpX [39, 40]. Most of the encoding genes are well conserved in different *Dickeya* species (Table 3). However, the *suf* gene cluster is present only in clade II as well as in *D. chrysanthemi*, HmpX is truncated in *D. undicola* and absent in *Dickeya* sp. 569 and *D. paradisiaca*. The indigoïdine encoding genes are absent in *D. paradisiaca* and *D. lacustris*. (Table 3), although strikingly, *D. lacustris* was more resistant to oxidative stress than *D. dadantii* despite the lack of indigoïdine production [13].

Table 3 Conservation of virulence-related factors.

	AvrL/M	***T1SS prt***	***prtG***	***prtABC***	***T3SS***	***T4SS VirB****	***T4SSb****	***T6SS***	***kat***	***sod***	***ind***	***suf***	***hmpX***
D. dadantii	2	1	1	3	1	var	var	1	2	2	yes	yes	1
D. solani	2	1	1	3	1	1	var	1	2	2	yes	yes	1
D. dianthicola	1	1	no	2 or 3	1	1	no	1	2	2	yes	yes	1
D. fangzhongdai	2	1	1	3	1	var	var	1	2	2	yes	yes	1
D. undicola	1	1	no	3	1	no	var	1	2	2	yes	yes	trunc
D. chrysanthemi	2	1	no	3	1	1	var	1	2	2	yes	yes	1
D. zeae	0 or 1	1	1	3	1	var	no	1	2	2	yes	no	1
Dickeya sp. 569	no	1	no	1	no	no	no	no	1	1	yes	no	no
D. lacustris	2	1	1	3	1	no	no	no	2	2	no	no	1
D. aquatica	2	1	1	1 to 3	1	no	no	no	2	2	yes	no	1
D. paradisiaca	no	no	no	no	no	no	no	no	2	2	no	no	no

*Two types of T4SS might be present in bacteria, either associated to protein secretion encoded by the *virB* operon or associated with plasmid conjugation encoded by a VirD2/VirD4/Trb locus present in an integrative conjugative transposon element (ICE) (T4SSSb). Trunc: truncated, var: variable.

At the species level, core genomes are very large comprising nearly 3000 to 4000 protein families, i.e., around two-thirds of the genetic content in a strain (Figure 2).

This high genetic conservation persists in clades that group different species. Indeed, the *D. solani/D. dadantii* clade, *D. undicola/D. fangzhongdai* clade and *D. lacustris/D. aquatica* clade also share around 3000 protein families (Figure 1). The *D. zeae/D. chrysanthemi/Dickeya* sp. NCPPB569 clade I is less related because it shares only around 2500 protein families, similar to the number of protein families shared by all five clade II *Dickeya* species (Figure 1).

Therefore, even at the genus level, *Dickeya* members share a very large pool of genes as compared to other *Enterobacterales* species. For comparison, using less stringent conditions (50%

identity on 50% of the length of the proteins) the *E. coli* core genome includes only around 1500 orthologous genes [41]. Furthermore, most virulence determinants are shared by most *Dickeya* species with the notable exceptions of *D. paradisiaca* and the *Dickeya* sp. 569 strain.

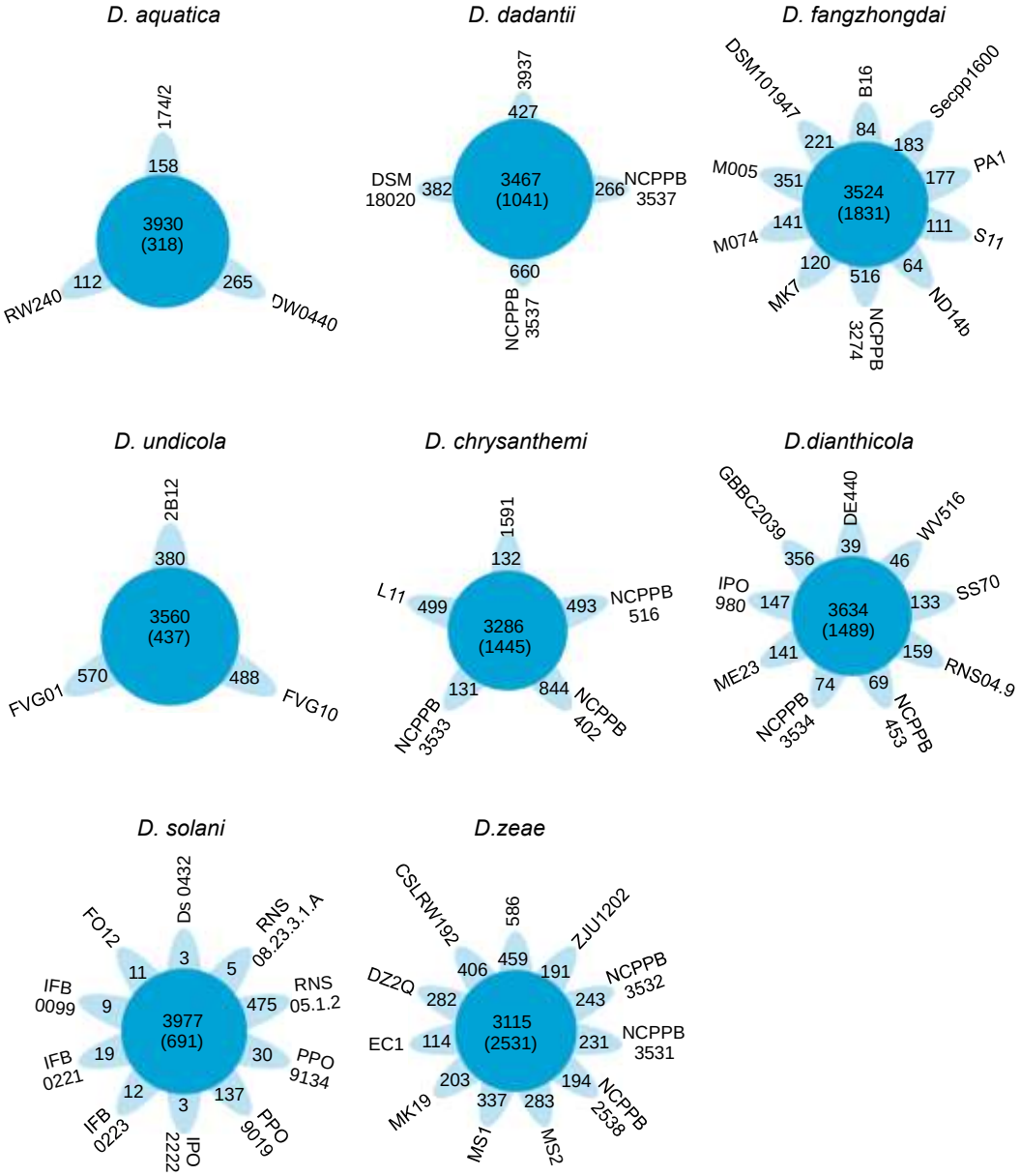

Figure 2 Conserved and strain-specific gene content in the different *Dickeya* species. The protein families present in the core, accessory (in brackets) and strain-specific (in petals) are depicted.

3.3 Accessory Genome and Strain-Specific Genes

The sizes of the accessory genome of species, which groups protein families present in more than one but not all genomes of a given species, ranged from about 400 to more than 2100 protein families (Figure 2). To analyze them further, we performed a pan-genome clustering analysis that builds a hierarchical clustering based on the proportion of the presence/absence status of each gene family in each genome (Figure 1B). The allocation in species determined by the whole

genome MLSA analysis is conserved reflecting the high number of protein families present in the core genome of each species. Strikingly, however, the relatedness between *Dickeya* species at the pan-genome level is quite different.

For instance, *D. chrysanthemi* did not cluster with *D. zeae* anymore but was now closer to clade II species, *D. undicola* fell outside clade II and *D. paradisiaca* was closer to *P. atrosepticum* 21A than to any *Dickeya* species (Figure 1B). At the species level, the distribution of *D. zeae* strains into two branches as per the MLSA and ANI analyses was conserved in the pan-genome distribution, further substantiating the separation into two subspecies/species. No correlation was observed between this pan-genome clustering and habitat or geographical origin (Figure 1B and Table 1) except in *D. undicola* (Figure 1B). Indeed, in *D. undicola*, the two strains isolated from France were closer to each other than to the 2B12 strain in the pan-genome clustering, while in the MLSA core genome analysis, *D. undicola* FVG10 was closer to 2B12 than to the other strain from France. Although the current analysis was based on only three genomes, it points to a possible closeness of the accessory genomes in relation to their geographical origin.

We also determined the number of protein families specific to each analyzed genome, and absent from all other *Dickeya* genomes considered in this study (Figure 2). We found that *D. solani* strains harbored very few strain-specific protein families (except 05.1.1.2), reflecting the already reported clonal origin of these strains [27, 28, 29, 30]. Three *D. dianthicola* strains (DE440, WV516, and ME23) also possess only a few dozen strain-specific genes coupled with very high ANI values (> 99.9%) among them. Interestingly, these three strains were isolated from potato in the USA, pointing to the same clonal origin as reported for the recent blackleg outbreak devastating potato crops in this country [2]. For the other strains, the number of strain-specific genes ranged from a few dozen to several hundreds. Further, we did not observe any link between the proximity of the core genomes revealed by ANI values in a given species and the extent of the strain-specific genome. For example, the members of the *D. dianthicola* and *D. undicola* species exhibited ANI values above 99% within a given species but harbored up to several hundred strain-specific genes, in the same range as *D. dadantii* members that shared only 96.7% ANI among them.

3.4 Features Specific to Each Dickeya Species

To tackle the functions that might be specific to each *Dickeya* species, we identified the species-specific genes present in all members of one species and missing in all other *Dickeya* genomes analyzed in this study (corresponding proteins harboring less than 70% identity). The number of species-specific genes ranged from 20 to 560 gene families (Table 4), although most of these species-specific genes encode hypothetical proteins or proteins with undefined function.

This may be due to the fact that the RAST CDS definition program considers very short proteins (less than 30 aa) for which no annotations are available. Expression profiling analyses of *D. solani* and *D. dianthicola* strains, however, revealed that most of the small CDS are expressed and may be differentially expressed in macerated tubers versus their growth in rich medium [42]. Annotation of the species-specific genes is listed in Table S2. Out of the five *D. undicola* genes with a known or predicted function, two genes encode regulatory proteins, and interestingly, a few (1 to 7) species-specific regulatory genes are present in all *Dickeya* species except *D. dadantii*. Several species-specific genes clustered in genomic regions, grouping at least five genes. *D. aquatica* harbors seven such genomic regions that group 5 to 33 genes. These regions mainly contain genes encoding

hypothetical proteins. Further, GR1 encodes an ABC-transport system related to nickel transporters and GR2 harbors genes related to secondary metabolism (Table S2). Most *D. chrysanthemi* species-specific genes with a known function are grouped in a genomic region of 26 genes that contain the Flp/Tad operon and its associated two-component regulatory system (Table S2). The genes in this operon encode the synthesis of a type IVb pilus shown to be involved in biofilm formation and virulence in several bacteria including *Pectobacteria* [43].

Table 4 *Dickeya* species and *D. aquatica/D. lacustris* clade specificities.

	Species-specific gene families	**Of which hypothetical**	**Species specifically absent genes**	**Of which hypothetical**
D. solani	122	101	0	
D. dadantii	22	22	0	
D. dianthicola	164	114	10	0
D.fangzhongdai	20	12	0	
D. undicola	129	120	3	0
D. chrysanthemi	77	54	2	0
D. zeae	48	30	0	
D. aquatica	560	481	9	3
D. lacustris	ND		8	0
D. paradisiaca	ND		143	32
	Clade-specific gene families	**Of which hypothetical**	**Clade specifically absent genes**	**Of which hypothetical**
D. aquatica/ D. lacustris clade	239	124	54	5

ND: not determined because too few genomes were available.

In animal pathogens, diversity in virulence at the species level is often related to the absence instead of the presence of genes, suggesting that high virulence is related to specialization as it was described for *Yersinia pestis* [44]. We thus examined if species specificity may reside in the absence of specific protein families that are present in all other *Dickeya* species. Very few such families are present in *Dickeya* species, less than ten in all, except in *D. paradisiaca* that lacks 124 such genes (Table S2). As already described above, those families absent from *D. paradisiaca* include several proteins associated with virulence like those encoding the Prt type 1 secretion system of proteases, the pectin acetylesterase PaeY, a catalase, part of the achromobactine siderophore biosynthesis pathway as well as several genes related to metabolism and regulation (Table S2). Strikingly, *D. paradisiaca* strains possess all genes encoding proteins involved in flagellum biosynthesis, although, these genes are more related to orthologous genes in *Lonsdalea* and *Brenneria* than to their *Dickeya* counterparts.

3.5 The D. aquatica/ D. lacustris Clade

The *D. aquatica*/*D. lacustris* clade is the only clade grouping strains isolated only from water, hence, we analyzed the genomic specificities of this clade. Members of this clade share 239 genes that are absent in all other *Dickeya* species (or exhibit less than 70% identity) (Table 4). Besides genes encoding hypothetical proteins or proteins with undefined function (124 genes), these clade-specific genes could be divided mainly into categories related to metabolism (45 genes), transport (14 genes), regulation (6 genes), and resistance to stress (9 genes) (Table S3). Several of these genes grouped into nine genomic regions (GR) defined as regions clustering at least five genes (Table 5). Five of these genomic regions include genes involved in interesting pathways/features.

Table 5 Genomic regions present only in the *D. aquatica/D. lacustris* clade.

Genomic regions	CSL RW240 coordinates	Predicted function
GR1 (12 genes)	298401-310142	*aga* operon: N-acetylglucosamine transport and metabolism
GR2 (6 genes)	454821-459568	CPS/LPS
GR3 (5 genes)	716104-725325	Cell envelope
GR4 (18 genes)	779773-797946	metabolism
GR5 (8 genes)	881328-892104	*fim* operon (type I fimbriae)
GR6 (15 genes)	996952-1010038	Tellurium resistance + metabolism
GR7 (8 genes)	2539184-2553382	hypothetical
GR8 (7 genes)	2801783-2810461	hypothetical

GR1 encodes genes involved in the transport and catabolism of N-acetyl-glucosamine, the monomer of chitin present in the cuticles and shells of insects and crustaceans and in fungal cell walls. The genes of GR2 and GR3 are predicted to be involved in capsule biosynthesis and LPS export indicating specificities in the bacterial cell envelope. GR5 genes are similar to the *fim* genes involved in the biosynthesis of type 1 fimbriae that confer adhesion to a variety of eukaryotic cells. In uropathogenic *E. coli*, these fimbriae play a critical role both in the colonization of the lower urinary tract and in the formation of biofilms by intracellular bacteria after the invasion of epithelial cells [45]. GR6 encompasses genes homologous to the Ter stress response gene cluster involved in tellurium resistance and that are activated during pathogenesis in *Yersinia* [46].

Likewise, 54 genes absent in the strains of the *D. aquatica/D. lacustris* clade were present in all other *Dickeya* species (Tables 4 and Table S3). Among them, only five were annotated as hypothetical proteins and several of these genes encode proteins involved in transport systems or metabolism. In particular, *D. aquatica/D. lacustris* lacks the gene cluster encoding proteins involved in the metabolism of xylose available in plant xyloglucans, and thus, these bacteria are not able to use xylose as a carbon source [12, 13].

For comparison, we performed a similar analysis with the members of the *D. zeae*/*D. chrysanthemi*/*Dickeya* sp. 569 clade I. This clade shares only nine genes that are absent in all other *Dickeya* species and conversely, no genes are absent in its members while present in all other *Dickeya*.

Therefore, although *D. aquatica* and *D. lacustris* were isolated from water until now, they are well equipped for invading plants as they possess most of the virulence-related genes (see 3.2) and, contrary to *D. paradisiaca*, none of the genes absent in these species are known to be involved in interactions with plants, except the genes involved in xyloglucan catabolism. Accordingly, *D. aquatica* was shown to be capable of macerating acidic fruits like tomato or cucumber [12].

3.6 Does D. paradisiaca belong to the Dickeya Genus?

Contrary to the definition of bacterial species, the delimitation of genus adjuncts is still a matter of debate. One of the proposed limits is an average nucleotide identity above 80% [32]. In the case of *D. paradisiaca*, the ANI values ranged from 78 to 80% with the other *Dickeya* species (at or below this limit), *D. aquatica/D. lacustris* being the more distant clade. Furthermore, *D. paradisiaca* lacks several enzymes involved in pectinolysis as well as virulence determinants (see 3.2). It was however considered as the earlier branching lineage in the *Dickeya* genus in evolution studies that analyzed the emergence of the different virulence-related genes [47, 12]. Finally, in our pan-genome analysis (Figure 1B), *D. paradisiaca* clusters with *P. atrosepticum*, outside the *Dickeya* genus.

To further analyze the relatedness of *D. paradisiaca* to the *Dickeya* genus, we performed synteny analyses taking advantage of the availability of complete genomes in our panel (Figure 3).

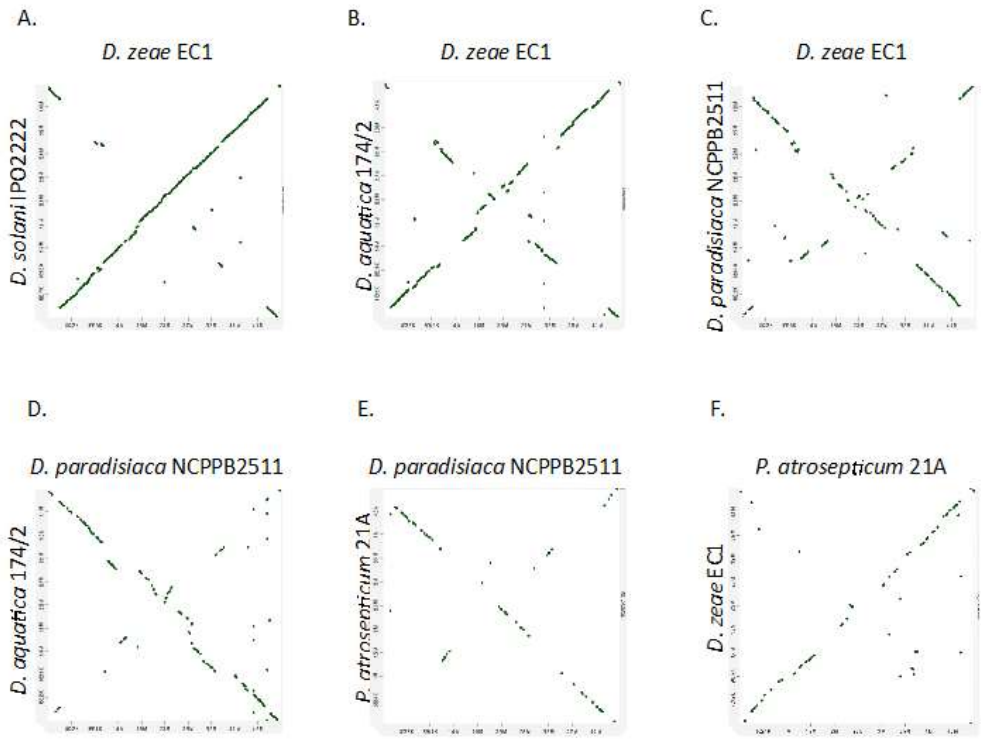

Figure 3 Synteny between *D. paradisiaca* and the other *Dickeya* clades.

The synteny of clade II members of *Dickeya* species is very high as already described using the MAUVE program [27, 42]. This high synteny was conserved between the clade I and clade II clusters that show collinearity with only a few inversions of genomic regions, as exemplified by the *D. zeae* EC1/*D. solani* IPO2222 synteny (Figure 3A). In *D. aquatica* 174/2, although the inversed regions were larger compared to these clusters the syntenic regions were still dense (Figure 3B). In contrast, the syntenic regions between *D. paradisiaca* NCPPB2511 and members of clades I and II were more scattered (Figure 3C). Comparison between *D. paradisiaca* and *D. aquatica* gave similar results (Figure 3D). For comparison, we also analyzed the synteny between *Dickeya* and the P*ectobacterium* related genus exemplified by a *P. atrosepticum* strain (Figure 3E and Figure 3F). The homologous regions were even more scattered despite a quite similar overall organization of *D. zeae* and *P. atrosepticum*. This analysis revealed a gradient of synteny between the members of the genus *Dickeya* with a very high closeness between the genomes of species belonging to clade I and clade II while less homologous regions were observed in *D. paradisiaca*, and an intermediate synteny was revealed in *D. aquatica*. Although *D. aquatica* was proposed as the second deepest branching lineage in the genus *Dickeya*, it was not more syntenic to *D. paradisiaca* than to the other *Dickeya* species. Nevertheless, our analyses only comprise two *D. paradisiaca* genomes (the only ones available) and further phenotypic, as well as genomic analyses of a larger number of strains, should be carried out before proposing a possible categorization of this bacterial group into different genera.

4. Conclusions

One hallmark of the genus *Dickeya* is the high genetic closeness among its members. Indeed, despite using high stringency parameters (70% identity with minimal overlap of 80%), all genomes analyzed shared around as much as 1800 protein families and this was also true for *D. paradisiaca* that is at the limit of affiliation to the genus *Dickeya*. At the species level, very few gene specificities exist and even fewer specific gene losses frequently found in specialized human pathogens. This, accompanied by high conservation of the characterized virulence-related genes in most species, may account for the broad host range of the different *Dickeya* species that can infect both monocot and dicot plants. Nevertheless, such a high relatedness in genetic content does not necessarily imply similarity in strategies to adapt to the plant hosts. In fact, transcriptomic studies during maceration of potato tubers revealed differences in gene expression even in very close *Dickeya* species like *D. solani* and *D. dianthicola* [42]. Likewise, pangenome analysis does not highlight strain clustering according to host or environment of their origin, further substantiating the lack of specialization to a given environment. Among the 239 protein families present only in the quite distant *D. aquatica/D. lacustris* clade that groups the strains isolated only from water, we did not find any genetic trait that might contribute to habitat specificity. Instead, the members of this clade shared two gene clusters that play a role in the interactions of animal pathogens with their hosts and the other one facilitates interaction with insects and fungi. No protein family was shared by the three *Dickeya* species found only in water so far, corroborating the absence of "water-specific" genes already reported in *D. fangzhongdai* water isolates [33]. None of the protein families shown to play a role in interactions with plants were found missing in the *D. aquatica/D. lacustris* clade. Hence, their spread via waterways, specifically irrigation water might be a potential threat for economically important crops.

Earlier studies, by compiling hundreds of strains previously identified as *Erwinia chrysanthemi*, defined seven *Dickeya* species [3, 26, 48]. In addition, the strain NCPPB569 (isolated from sugar cane plantations in Australia) is significantly distinct to belong to an additional new species [3, 26, 48]. Recent studies in aquatic environments allowed the identification of three new *Dickeya* species that were found only in water so far. It would be interesting to extend the search for *Dickeya* isolates to new environments, for instance, insects that are recognized as hosts for some *Dickeya* [49] to explore if this will further broaden the diversity of genus *Dickeya*.

Acknowledgments

We thank Nicole Hugouvieux-Cotte-Pattat for sharing the *D. lacustris* genomic sequence before publication and for fruitful discussions.

Additional Materials

The following additional materials are uploaded at the page of this paper.

1. Table S1: Relatedness of the *Dickeya* species. *in silico* DNA-DNA hybridization (DDH, upper triangle) and Average Nucleotide Identity (ANI, lower triangle) values are presented. The specific threshold value is 70% for DDH and 96% for ANI. Strains fulfilling these thresholds are highlighted in tangerine.
2. Table S2: *Dickeya s*pecies-specific and species specifically absent gene families.
3. Table S2: Gene families specific to and specifically absent from the *D. aquatica/D. lacustris* clade.

Author Contributions

JP performed the bioinformatics analyses. Both authors analysed the data and wrote the manuscript.

Funding

This research was supported by the COMBICONTROL ((ANR-17-CE32-0004-04) and SPREE (ANR-17-CE32-0004-04) projects financed by the French National Agency for Research.

Competing Interests

The authors have declared that no competing interests exist.

References

1. Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, et al. Top 10 plant pathogenic bacteria in molecular plant pathology. Mol Plant Pathol. 2012; 13: 614-629.
2. Charkowski AO. The changing face of bacterial soft-rot diseases. Annu Rev Phytopathol. 2018; 56: 269-288.
3. Suharjo R, Sawada H, Takikawa Y. Phylogenetic study of Japanese *Dickeya* spp. and development of new rapid identification methods using PCR-RFLP. J Gen Plant Pathol. 2014; 80: 237-254.

4. Charkowski A, Blanco C, Condemine G, Expert D, Franza T, Hayes C, et al. The role of secretion systems and small molecules in soft-rot *Enterobacteriaceae* pathogenicity. Annu Rev Phytopathol. 2012; 50: 425-449.
5. Samson R, Legendre JB, Christen R, Fischer-Le Saux M, Achouak W, et al. Transfer of *Pectobacterium chrysanthemi* (Burkholder et al 1953) Brenner et al 1973 and *Brenneria paradisiaca* to the genus *Dickeya gen*. nov. as *Dickeya chrysanthemi* comb. nov. and *Dickeya paradisiaca* comb. nov. and delineation of four novel species. Int J Syst Evol Microbiol. 2005; 55: 1415-1427.
6. Brady CL, Cleenwerck I, Denman S, Venter SN, Rodrıguez-Palenzuela P, Coutinho TA, et al. Proposal to reclassify *Brenneria quercina* (Hildebrand & Schroth 1967) Hauben et al. 1999 into a novel genus, *Lonsdalea gen.* nov., as *Lonsdalea quercina* comb. nov., descriptions of *Lonsdalea quercina* subsp. *quercina* comb. nov., *Lonsdalea quercina* subsp. *Iberica* subsp. nov., and *Lonsdalea quercina* subsp. *britannica* subsp. nov., emendation of the description of the genus *Brenneria*, reclassification of *Dickeya dieffenbachiae* as *Dickeya* A new clade of *Dickeya dadantii* subsp. *dieffenbachiae* comb. nov., and emendation of the description of *Dickeya dadantii*. Int J Syst Evol Microbiol. 2012; 62: 1592-1602.
7. Parkinson N, DeVos P, Pirhonen M, Elphinstone J. *Dickeya aquatica* sp. nov., isolated from waterways. Int J Syst Evol Microbiol. 2014; 64: 2264-2266.
8. Toth IK, van der Wolf JM, Saddler G, Lojkowska E, Helias V, Pirhonen M, et al. *Dickeya* species: An emerging problem for potato production in Europe. Plant Pathol. 2011; 60: 385-399.
9. van der Wolf JM, Nijhuis EH, Kowalewska MJ, Saddler GS, Parkinson N, Elphinstone JG, et al. *Dickeya solani* sp. nov., a pectinolytic plant-pathogenic bacterium isolated from potato (*Solanum tuberosum*). Int J Syst Evol Microbiol. 2014; 64: 768-774
10. Tian Y, Zhao Y, Yuan X, Yi J, Fan J, Xu Z, et al. *Dickeya fangzhongdai* sp. nov., a plant pathogenic bacterium isolated from pear trees (*Pyrus pyrifolia*). Int J Syst Evol Microbiol. 2016; 66: 2831-2835.
11. Alič Š, Van Gijsegem F, Pédron J, Ravnikar M, Dreo T. Diversity within the novel *Dickeya fangzhongdai* sp., isolated from infected orchids, water and pears. Plant Pathol. 2018; 67: 1612-1620.
12. Duprey A, Taib N, Leonard S, Garin T, Flandrois, JP Nasser W, et al. The phytopathogenic nature of *Dickeya aquatica* 174/2 and the dynamic early evolution of *Dickeya* pathogenicity. Env Microbiol. 2019; 21: 2809-2835.
13. Hugouvieux-Cotte-Pattat N, Jacot-des-Combes C, Briolay J. *Dickeya lacustris* sp. nov., a pectinolytic bacterium isolated from lakes in the French region of La Dombes. Int J Syst Evol Microbiol. 2019; 69: 721-726.
14. Oulghazi S, Pédron J, Cigna J, Lau YY, Moumni M, Van Gijsegem F, et al. *Dickeya undicola* sp. nov., a novel species for pectinolytic isolates from surface waters in Europe and Asia. Int J Syst Evol Microbiol. 2019; 69: doi: 10.1099/ijsem.0.003497
15. Aziz RK, Bartels D, Best AA, Dejongh M, Disz T, Edwards RA, et al. The RAST server: Rapid annotations using subsystems technology. BMC Genomics .2008; 9: 75.
16. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Tothc IK. Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. Anal Methods. 2016; 8: 12-24

17. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 2013; 14: 60.
18. Cabanettes F, Klopp C. D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. PeerJ. 2018; 6: e4958.
19. Miele V, Penel S, Duret L. Ultra-fast sequence clustering from similarity networks with SiLiX. BMC Bioinformatics. 2011; 12: 116.
20. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32: 1792-1797.
21. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylo- genetic tree building. Mol Biol Evol, 2010; 27: 221-224.
22. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000; 17: 540-552.
23. Gascuel O. BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol. 1997; 14: 685-695.
24. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016; 33: 1870-1874
25. Méric G, Yahara K, Mageiros L, Pascoe B, Maiden MCJ, Jolley KA, et al. A reference pan-genome approach to comparative bacterial genomics: Identification of novel epidemiological markers in pathogenic *Campylobacter*. PLoS One. 2014; 9: e92798.
26. Zhang Y, Fan Q, Loria R. A re-evaluation of the taxonomy of phytopathogenic genera *Dickeya* and *Pectobacterium* using whole-genome sequencing data. Syst Appl Microbiol. 2016; 39: 252-259.
27. Pédron J, Mondy S, Raoul des Essarts Y, Van Gijsegem F, Faure D. Genomic and metabolic comparison with *Dickeya dadantii* 3937 reveals the emerging *Dickeya solani* potato pathogen to display distinctive metabolic activities and T5SS/T6SS-related toxin repertoire. BMC Genomics. 2014; 15: 1-13
28. Khayi S, Blin P, Pédron J, Chong TM, Chan KG, Moumni M, et al. Population genomics reveals additive and replacing horizontal gene transfers in the emerging pathogen *Dickeya solani*. BMC Genomics. 2015; 16: 1-13.
29. Golanowska M, Potrykus M, Motyka-Pomagruk A, Kabza M, Bacci G, Galardini M, et al. Comparison of highly and weakly virulent *Dickeya solani* strains, with a view on the pangenome and panregulon of this species. Front Microbiol. 2018; 9: 1940.
30. Sławiak M, Beckhoven JRCM, Speksnijder AGCL, Czajkowski R, Grabe G, Wolf JM. Biochemical and genetical analysis reveal a new clade of biovar 3 *Dickeya* spp. strains isolated from potato in Europe. Eur J Plant Pathol. 2009; 125.
31. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Am Sci USA. 2009; 106: 19126-19131
32. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. Proc Natl Am Sci USA. 2005; 102: 2567-2572.
33. Alic Š, Pedron J, Dreo T, Van Gijsegem F. Genomic characterisation of the new *Dickeya fangzhongdai* species regrouping plant pathogens and environmental isolates. BMC Genomics. 2019; 20: 34.
34. Tseng TT, Tyler BM, Setubal JC. Protein secretion systems in bacterial-host associations, and their

description in the gene ontology. BMC Microbiol. 2009; 9: S2.
35. Reverchon S, Nasser W. *Dickeya* ecology, environment sensing and regulation of virulence programme. Environ Microbiol Rep. 2013; 5: 622-636.
36. El Hassouni M, Chambost JP, Expert D, Van Gijsegem F, Barras F. Identification of peptide methionine sulfoxide reductase as a virulence determinant of the plant pathogen *Erwinia chrysanthemi*. Proc Natl Acad Sci. 1999; 96: 887-892
37. Nachin L, El Hassouni M, Loiseau L, Expert D, Barras F. SoxR-dependent response to oxidative stress and virulence of *Erwinia chrysanthemi*: The key role of SufC, an orphan ABC ATPase. Mol Microbiol. 2001; 39: 960-972.
38. Rincon-Enriquez G, Crete P, Barras F, Py B. Biogenesis of Fe/S proteins and pathogenicity: IscR plays a key role in allowing *Erwinia chrysanthemi* to adapt to hostile conditions. Mol Microbiol. 2008; 67: 1257-1273.
39. López-Solanilla E, Llama-Palacios A, Collmer A, García-Olmedo F, Rodríguez-Palenzuela P. Relative effects on virulence of mutations in the sap, pel, and hrp loci of *Erwinia chrysanthemi*. Mol Plant Microbe Interact. 2001; 14: 386-393.
40. Boccara M, Mills CE, Zeier J, Anzi C, Lamb C, Poole RK, et al. Flavohaemoglobin HmpX from Erwinia chrysanthemi confers nitrosative stress tolerance and affects the plant hypersensitive reaction by intercepting nitric oxide produced by the host. Plant J. 2005; 43: 226-237.
41. Lukjancenko O, Wassenaar TM, Ussery DW. Comparison of 61 sequenced *Escherichia coli* genomes. Microb Ecol. 2010; 60: 708-720.
42. des Essarts YR, Pédron J, Blin P, Van Dijk E, Faure D, Van Gijsegem F. Common and distinctive adaptive traits expressed in *Dickeya dianthicola* and *Dickeya solani* pathogens when exploiting potato plant host. Envi Microbiol. 2019; 21: 1004-1018.
43. Nykyri J, Mattinen L, Niemi O, Adhikari S, Koiv V, Somervuo P, et al. Role and regulation of the Flp/Tad pilus in the virulence of *Pectobacterium atrosepticum* SCRI1043 and *Pectobacterium wasabiae* SCC3193. PLoS One. 2013; 8: e73718.
44. Merhej V, Georgiades K, Raoult D. Postgenomic analysis of bacterial pathogens repertoire reveals genome reduction rather than virulence factors. Brief Funct Genomics. 2013; 12: 291-304.
45. Wright KJ, Seed PC, Hultgren SJ. Development of intracellular bacterial communities of uropathogenic *Escherichia coli* depends on type 1 pili. Cell Microbiol. 2007; 9: 2230-2241.
46. Ponnusamy D, Hartson SD, Clinkenbeard KD. Intracellular *Yersinia pestis* expresses general stress response and tellurite resistance proteins in mouse macrophages. Vet Microbiol. 2011; 150: 146-151.
47. Duprey A, Nasser W, Leonard S, Brochier-Armanet C, Reverchon S. Transcriptional start site turnover in the evolution of bacterial paralogous genes: The *pelE-pelD* virulence genes in *Dickeya*. FEBS J. 2016; 283: 4192-4207.
48. Parkinson N, Stead D, Bew J, Heeney J, Tsror L, Elphinstone J. *Dickeya* species relatedness and clade structure determined by comparison of *recA* sequences. Int J Syst Evol Microbiol. 2009; 59: 2388-2393.
49. Rossmann S, Dees MW, Perminow J, Meadow R, Brurberg MB. Soft rot Enterobacteriaceae are carried by a large range of insect species in potato fields. Appl Environ Microbiol. 2018; 84: e00281-18.

Enjoy *OBM Genetics* by:

1. Submitting a manuscript
2. Joining in volunteer reviewer bank
3. Joining Editorial Board
4. Guest editing a special issue

For more details, please visit:
http://www.lidsen.com/journals/genetics