Dickeya species relatedness and clade structure determined by comparison of recA sequences

Neil Parkinson,1 David Stead,1 Janice Bew,1 John Heeney,1 Leah Tsror (Lahkim)2 and John Elphinstone1

Correspondence
Neil Parkinson
n.parkinson@csl.gov.uk
1Central Science Laboratory, Sand Hutton, York YO41 1LZ, UK
2Agricultural Research Organisation, Gilat Research Centre, MP Negev 85280, Israel

Using sequences from the recA locus, we have produced a phylogeny of 188 Dickeya strains from culture collections and identified species relatedness and subspecies clade structure within the genus. Of the six recognized species, Dickeya paradisiaca, D. chrysanthemi and D. zeae were discriminated with long branch lengths. The clade containing the D. paradisiaca type strain included just one additional strain, isolated from banana in Colombia. Strains isolated from Chrysanthemum and Parthenium species made up most of the clade containing the D. chrysanthemi type strain, and the host range of this species was extended to include potato. The D. zeae clade had the largest number of sequevars and branched into two major sister clades that contained all of the Zea mays isolates, and were identified as phylogtypes PI and PII. The host range was increased from six to 13 species, including potato. The recA sequence of an Australian sugar-cane strain was sufficiently distinct to rank as a new species-level branch. In contrast to these species, Dickeya dadantii, D. dianthicola and D. dieffenbachiae were distinguished with shorter branch lengths, indicating relatively closer relatedness. The recA sequence for the type strain of D. dadantii clustered separately from other strains of the species. However, sequence comparison of three additional loci revealed that the D. dadantii type strain grouped together with the six other D. dadantii strains that were sequenced. Analysis of all four loci indicated that the D. dadantii strains were most closely related to D. dieffenbachiae. Three further branches (DUC-1, -2 and -3) were associated with these three species, which all diverged from a common origin and can be considered as a species complex. The large clade containing the D. dianthicola type strain comprised 58 strains and had little sequence diversity. One sequevar accounted for the majority of these strains, which were isolated nearly exclusively from eight hosts from Europe. Isolation of this sequevar on multiple occasions from Dianthus and (more recently) potato demonstrates that this lineage has become established in these species. The D. dadantii clade comprised 11 sequevars, and the known host range of the species was extended from eight to 19 species. New hosts included several ornamental species and potato. The clade DUC-1 was made up exclusively of potato strains originating from Europe, which had identical sequences, whilst DUC-2 strains were isolated mostly from a variety of monocotyledonous species. A single strain from Aglaonema sp. made up DUC-3. A single sequevar constituted the D. dieffenbachiae clade. The phylogenetic method described will provide a simple means for identification to the species and intraspecies level, which will support efforts to control these pathogens based on monitoring and surveillance.

INTRODUCTION

Plant diseases caused by Dickeya species result in significant losses to crops and ornamentals around the world. Disease symptoms include soft rots as well as wilts resulting from vascular infections. The genus Dickeya was formed from the transfer of Pectobacterium chrysanthemi and Brenneria paradisiaca as Dickeya chrysanthemi and D. paradisiaca, together with the identification of four newly described species, Dickeya dadantii, D. dianthicola, D. dieffenbachiae and D. zeae (Samson et al., 2005). The reader is referred to
the comprehensive review of the complex classification and taxonomic history of plant-pathogenic pectolytic enterobacteria in the latter report, and specifically to the earlier characterization studies of Young et al. (1978), Nassar et al. (1994) and Hauben et al. (1998). The study defining the current Dickeya classification used a wide range of characterization techniques including analysis of 16S rRNA gene sequences and DNA–DNA reassociation kinetics, as well as determination of a range of phenotypic traits including biochemical and serological reactions.

In recent years, phylogenies based on comparison of protein-coding genes have provided greater taxonomic resolution than those based on 16S RNA gene sequences and DNA–DNA hybridization analysis. Several loci have been used and found reliably to indicate species or subspecies relatedness in several groups of plant-pathogenic bacteria, including *Balstonia solanacearum*, *Pseudomonas syringae* and *Xanthomonas* (Fegan & Prior, 2005; Sarkar & Guttmann, 2004; Parkinson et al., 2007). A single-locus analysis of approximately 203 strains has identified new diversity within *Xanthomonas* and resolved the classification of many poorly characterized strains (Parkinson et al., 2009). A recent phylogenetic analysis using concatenated *atpD*, *carA* and *recA* loci (Young & Park, 2007) resolved the relatedness between many plant-pathogenic genera and species of the *Enterobacteriaceae*. However, within Dickeya, sequences were reported only for *D. chrysanthemi* and *D. paradisiaca*. Our current study using the *recA* locus (Waleron et al., 2002) will be a first phylogenetic analysis to use an open reading frame to indicate relatedness between all species within the genus. To determine Dickeya population sequence diversity and subspecies clade structure, we sequenced the *recA* locus from 188 strains obtained from culture collections. Currently, diagnostic identification of Dickeya strains is problematic, even to the species level, as no reliable, rapid technique is available. This study also aims to provide a rapid means of identification to the species and subspecies level.

**METHODS**

**Bacterial cultures.** Strains used are listed in Supplementary Table S1 (available in IJSEM Online) and were from four collections: the National Collection of Plant Pathogenic Bacteria (NCPBP), Central Science Laboratory, Sand Hutton, York, UK, the Collection Française de Bactéries Phytopathogènes (CFBP), Beaucouzé, France, the Culture Collection of the Plant Protection Service (PD), Wageningen, Netherlands, and the collection of the Institute for Plant Protection (IPO), DLO Research Institute for Plant Protection (IPO-DLO), Wageningen, Netherlands. Recently characterized Dickeya strains from Israeli potatoes have various identifiers and were supplied by one of the authors (L.T.). Laboratory strains from the CSL diagnostic clinic are prefixed ‘207’. Supplementary Table S1 also lists the plant hosts, country of origin and accession date.

**Sequencing and phylogenetic analysis.** The protocol for PCR amplification of the *recA* locus was based on the method described by Waleron et al. (2002). Final concentrations in the PCR mixture were 2.5 mM MgCl₂, 300 nM each primer, 0.2 mM dNTPs and 1 U Taq polymerase, made up in a final volume of 25 μL. PCR cycling parameters were initial denaturation at 95 °C for 3 min followed by 34 cycles of 95 °C for 1 min, 47 °C for 1 min and 72 °C for 2 min, with a final elongation step at 72 °C for 5 min. The PCR product (approx. 700 bp) was purified using a commercially available kit (Wizard PCR and Gel Clean Up; Promega). The product was sequenced in both directions using a commercial sequencing service (MWG Ltd) using forward and reverse sequencing primers. Sequences were clipped to a standard start and finish position to a length of 481 nt. The 5’ start sequence begins CGG(T/C)/GG(T/C) and the 3’ sequence ends at GTAA(C/T)/G. New conserved primers DicCONFPCR1 and Dic.CON.RPCR1 (Supplementary Table S2) are recommended, which were designed based on sequences produced in this study. The PCR conditions described above were used except that the annealing temperature was increased to 55 °C and the extension time at 72 °C was reduced to 1 min. After purification, both strands of the PCR product were sequenced using the sequencing primers listed in Supplementary Table S2. The sequences were processed to the same start and finish positions as described above. Sequence alignment was done using CLUSTAL W (Thompson et al., 1994). Distance estimation was done using the algorithm described by Tajima & Nei (1984). Phylogenetic trees were generated using the neighbour-joining method (Saitou & Nei, 1987). The circular phylogeny was produced from CLUSTAL W using the Interactive Tree Of Life (iTOL) online tool (Letunic & Bork, 2007).

Three additional loci, aconitase (*acnB*), phosphofructokinase (*pfkA*) and RNA polymerase (*rpoD*), were used for sequence comparison of the *D. dadantii* type strain and previously characterized *D. dadantii* strains. Primers used for PCR and sequencing are listed in Supplementary Table S2 and were designed from sequences from a strain of *D. dadantii* (referred to as strain 3937), currently being sequenced at the Institute for Genomic Research (TIGR) and the Genome Evolution Laboratory at the University of Wisconsin. The PCR mixture was the same as described above. PCR parameters were denaturation at 95 °C for 2.5 min, followed by 35 cycles of 95 °C for 30 s, 55 °C for 45 s and 72 °C for 45 s, with a final extension step of 10 min at 72 °C. All three loci produced amplicons of approximately 700 bp. Sequences were clipped to common start and finish sequences [pfkA (5’-TTTTTC; AAAAATC-3’), rpoD (5’-ATCAACC; CGACCCA-3’) and acnB (5’-CCGGG; GTTTCG-3’)] to produce respective lengths of 554, 554 and 583 bases. Phylogenies were produced as described for the *recA* locus.

**RESULTS**

*recA* sequences for the *Dickeya* type strains shared between 82 and 94% sequence identity (Supplementary Table S3). GenBank accession numbers are listed in Supplementary Table S4. Three phylogenies are presented. Fig. 1 is a circular phylogram of all the strains, which should be read in conjunction with an A3-sized version of a conventional phylogeny that is presented as Supplementary Fig. S1. A summary of the phylogeny, including type strains of all species and unclassified clades, together with *D. dadantii* strains, is provided in Fig. 2. Taxa within Dickeya grouped separately from the other members of the *Enterobacteriaceae* (members of the genera *Pantoaea*, *Pectobacterium* and *Erwinia*) that were sequenced.

The type strains of *D. paradisiaca*, *D. zeae* and *D. chrysanthemi* were differentiated from all other species with long branch lengths. The type strain of *D. paradisiaca*
shared an identical sequence with another banana isolate, also from Colombia, and these strains were the sole representatives of a single branch. The *D. zeae* clade contained all the maize strains and split into two principal subclades. One clade contained all maize isolates from the USA, which we denote phylotype PI, whilst the sister clade is denoted PII (Supplementary Fig. S1). Both had a high proportion of sequence variants, totalling 21 in all. Though the most common host was maize, this pathogen was also isolated from several other plants, and both subclades contained Australian potato isolates. Of the six strains from *Chrysanthemum* species, five grouped in a clade with the *D.*
chrysanthemi type strain; other hosts were potato, tomato and Parthenium species. All three Parthenium isolates had identical sequences, including the pathovar reference strain for this host. The branch length for the sequence of strain NCPPB 569, an Australian sugar-cane isolate, was similar to those of the three species discussed above.

The recA sequence of the D. dadantii type strain (CFBP 1269) grouped on its own, separate from the other previously characterized strains attributed to this species (CFBP 3697, CFBP 3780, CFBP 3855, CFBP 4152, CFBP 6467 and CFBP 4153; Samson et al., 2005). To resolve this discrepancy, we compared the type strain with these strains using three other loci. The phylogenies for acnB, pfkA and rpoD (Fig. 2) each placed the type strain in clades containing these previously characterized strains. These three loci each reliably distinguished the D. dadantii strains from D. dieffenbachiae. The pfkA phylogeny indicates the relative branch lengths that separate D. dadantii strains from D. dieffenbachiae and D. dianthicola. Based on these phylogenies, the recA clade containing the D. dadantii strains was attributed to D. dadantii (Figs 2 and 3). The D. dadantii clade comprised a high proportion of sequevars (11 from 31 strains), which were isolated from a very diverse range of plant hosts (19 species) including the crop plants potato, Ipomoea sp. (sweet potato) and Musa sp. (banana) as well as nine ornamental plant species, including the species type strain, from Pelargonium capitatum. The D. dieffenbachiae type strain grouped in a clade with identical sequences made up of Dieffenbachia sp. isolates from Germany and the USA, in addition to a tomato strain.

The large, well-defined clade containing the D. dianthicola type strain has little sequence diversity. The large subclade containing the type strain comprised a single sequevar of 49 strains that were isolated from potato and species of Dianthus, Chrysanthemum, Sedum, Begonia, Dahlia and Kalanchoe. A sister clade was made up of two sequevars, which originated from potato and species of Cichorium and Dianthus.

chrysanthemi type strain; other hosts were potato, tomato and Parthenium species. All three Parthenium isolates had identical sequences, including the pathovar reference strain for this host. The branch length for the sequence of strain NCPPB 569, an Australian sugar-cane isolate, was similar to those of the three species discussed above.

The recA sequence of the D. dadantii type strain (CFBP 1269) grouped on its own, separate from the other previously characterized strains attributed to this species (CFBP 3697, CFBP 3780, CFBP 3855, CFBP 4152, CFBP 6467 and CFBP 4153; Samson et al., 2005). To resolve this discrepancy, we compared the type strain with these strains using three other loci. The phylogenies for acnB, pfkA and rpoD (Fig. 2) each placed the type strain in clades containing these previously characterized strains. These three loci each reliably distinguished the D. dadantii strains from D. dieffenbachiae. The pfkA phylogeny indicates the relative branch lengths that separate D. dadantii strains from D. dieffenbachiae and D. dianthicola. Based on these phylogenies, the recA clade containing the D. dadantii strains was attributed to D. dadantii (Figs 2 and 3). The D. dadantii clade comprised a high proportion of sequevars (11 from 31 strains), which were isolated from a very diverse range of plant hosts (19 species) including the crop plants potato, Ipomoea sp. (sweet potato) and Musa sp. (banana) as well as nine ornamental plant species, including the species type strain, from Pelargonium capitatum. The D. dieffenbachiae type strain grouped in a clade with identical sequences made up of Dieffenbachia sp. isolates from Germany and the USA, in addition to a tomato strain.

The large, well-defined clade containing the D. dianthicola type strain has little sequence diversity. The large subclade containing the type strain comprised a single sequevar of 49 strains that were isolated from potato and species of Dianthus, Chrysanthemum, Sedum, Begonia, Dahlia and Kalanchoe. A sister clade was made up of two sequevars, which originated from potato and species of Cichorium and Dianthus.
The recA sequences from these three species all diverged from a common ancestor, along with three further branches, not assigned to a species, which can be considered as a species complex. The three unassigned clades were designated D. dieffenbachiae/D. dadantii/D. dianthicola species complex unassigned clades 1–3 (DUC-1 to -3). Strains comprising DUC-1 shared identical sequences and were made up exclusively of potato isolates from Europe. Five of six strains that constituted DUC-2 were from monocotyledonous hosts and were isolated from around the world. The single strain representing DUC-3 was from Aglaonema sp. from St Lucia.

**DISCUSSION**

The recA phylogeny confirmed that Dickeya constitutes a monophyletic group, distinct from species of Pantoea, Pectobacterium and Erwinia, related plant-pathogenic members of the Enterobacteriaceae. Branch lengths varied between the species; those previously classified as members of other genera, D. paradisiaca and D. chrysanthemi, together with D. zeae, had long branch lengths, and these species are not closely related. The sequence of the Australian sugar-cane strain NCPPB 569 is sufficiently distinct for it to be considered as a member of a potential novel species. The distinctiveness of this strain is further supported by 16S rRNA gene sequence data (N. Parkinson, unpublished). A different isolate from the same host and country was unclustered in the previous Dickeya study (Samson et al., 2005), and these strains should be compared in future studies. Future studies aimed at assessing the placement of this strain within a novel species should encompass analysis of further loci, together with additional supporting characterization data.

Interestingly, the recA sequence from the type strain of D. dadantii was distinct from those of other strains previously characterized as D. dadantii. The three additional loci sequenced to resolve this discrepancy each placed the type strain within a clade containing the D. dadantii strains (Fig. 3). We have therefore attributed the recA clade containing these strains to this species. The discrepancy between the loci may have arisen as a result of a recombination event involving the recA locus that occurred in the evolutionary history of the type strain. Such recombination events have been found to occur rarely within strains of Pseudomonas syringae. In this species, the low incidence of recombination was suggested to be due to the narrow host ranges of pathovars, which minimized the possibilities for genetic exchange (Sarkar & Guttmann, 2004; Sarkar et al., 2006).

The phylogenies from the three additional loci sequenced confirmed the relatively close relatedness between D. dadantii strains and D. dieffenbachiae that was found using the recA locus. The clades containing strains of D. dianthicola, D. dadantii and D. dieffenbachiae, together with DUC-1, -2 and -3, were distinguished with shorter branch lengths. All of these lineages diverged from a common ancestral recA sequence origin and can be considered as a species complex. The phylogeny groups DUC-1 and DUC-3 strains together and places DUC-1 as a sister clade to D. dieffenbachiae; however, the branches are short and are supported only by low bootstrap values. Since recombination could potentially have contributed to the distinctiveness of DUC-1, -2 and -3 recA sequences, it will be important to analyse further loci and to provide additional characterization data to resolve their relatedness to species within the complex with confidence.

Sequence diversity within species implies that strains have been maintained in the environment over a relatively long period of time, allowing sequence variation to accumulate. Greatest sequence diversity occurred within the two major D. zeae clades identified as phylotypes PI and PII. Whilst maize was the most common host for this species, its host range was considerably extended from six to 13 species; new hosts included potato (from Australia and Papua New Guinea), Triticum aestivum (wheat) and Brassica chinensis (Chinese cabbage) as well as ornamental species. All the maize isolates were found within these two clades. Only PI contained strains isolated from maize in the USA, which were represented by five sequvars that may reflect a geographical centre for these sequvars. Significant sequence diversity was also found within D. chrysanthemi and D. dadantii. The host range of the former species was increased to include potato, carrot and species of Kalanchoe and Euphorbia. The host range of D. dadantii was extended from eight to 19 species; new hosts included potato (with isolates from Peru, Brazil and the Netherlands), as well as many ornamental species.

In contrast to this sequence diversity, strains within D. dianthicola, D. dieffenbachiae and DUC-1 had little or no sequence variation. Sequence similarity implies a relatively recent common origin. Most of the 58 strains from the D. dianthicola clade had identical recA sequences. Repeated isolation from potato and species of Dianthus indicates that this lineage has become established in these two hosts, though the sequvar was also found in a further five hosts. The range of hosts demonstrates interspecies dissemination at the sequvar level. A geographical basis for the occurrence of D. dianthicola is suggested, since nearly all strains have been isolated from Europe. Culture collection accession dates indicate that the lineage was isolated during or before 1956 from Dianthus caryophyllus; later accessions were from Dahlia sp. (1962), Begonia bertainii (1969) and potato (1975), suggesting that this lineage was first established in ornamental species. A distinct D. dianthicola sequvar was also found in potato from the UK and the Netherlands. Further research providing higher-resolution strain identification is needed to confirm whether direct transmission to new hosts has occurred within intensive production systems, which is suggested by the identical recA sequences. A further example of dissemination at the sequvar level in potato is provided by the clade DUC-1. This sequvar has so far been found exclusively in potato since 2004 (Tsror (Lahkim) et al., 2006, 2009), and has
been isolated from at least three European countries. The restricted host range, geographical origin and short time period since emergence of the sequevar suggest clonal dissemination within potato. Asymptomatic tuber infection may hinder efforts to control this pathogen in potato. The host ranges of \textit{D. zeae}, \textit{D. dadantii} and \textit{D. chrysanthemi} were all extended in this study to include potato, which underlines the threat posed by members of the genus to this important crop.

Whilst the \textit{recA} locus identified DUC-1, -2 and -3 as distinct lineages, there was insufficient sequence information to attribute these lineages confidently to taxa within the species complex. Current procedures for identifying \textit{Dickeya} species either lack resolution or use methods that are technically difficult. The \textit{recA} classification established in this study will provide a convenient means of identifying strains to the species and subspecies level, which will support control measures based on monitoring and surveillance of these important broad-host-range pathogens.

**ACKNOWLEDGEMENTS**

This work was supported by the British Potato Council and Plant Health Division of DEFRA. The licence number for working with quarantine strains is PHL 251C/5574 (02/2007).

**REFERENCES**


