Phylogenetic analysis of Xanthomonas species by comparison of partial gyrB gene sequences

Neil Parkinson,¹ Valentine Aritua,² John Heeney,¹ Claire Cowie,¹ Janice Bew¹ and David Stead¹

¹Central Science Laboratory, Sand Hutton, York YO41 1LZ, UK
²National Agricultural Biotechnology Centre, Kawanda Agricultural Research Institute, PO Box 7065, Kampala, Uganda

The genus Xanthomonas currently comprises 27 species with validly published names that are important crop and horticultural pathogens. We have constructed a phylogram from alignment of gyrB sequences for all xanthomonad species, both to indicate inter-species relatedness and as an aid for rapid and accurate species-level identification. The phylogeny indicated a monophyletic group, with X. albilineans and X. sacchari as the most ancestral species. Three species, X. hyacinthi, X. translucens and X. theicola, formed an early-branching group. Three clades were supported by high bootstrap values: group 1 comprised X. cucurbitae, X. cassavae and X. codiae; group 2 comprised X. arboricola, X. campestris, X. populi, X. hortorum, X. gardneri and X. cynarae; group 3 contained the remaining species, within which two further clades, supported by a 100 % bootstrap value, were identified. Group 3A comprised X. axonopodis, X. euvesicatoria, X. perforans and X. melonis, together with X. alfalfae, X. citri and X. fuscans, whose names were recently validly published. Group 3B contained the monocot pathogens X. vasicola and X. oryzae. Two recently identified species, X. cynarae and X. gardneri, were poorly discriminated and were related closely to X. hortorum. Three species, X. perforans, X. euvesicatoria and X. alfalfae, had identical gyrB sequences. Partial sequencing of a further five genes from these species found only minor sequence differences that confirmed their close relatedness. Although branch lengths between species varied, indicating different degrees of genetic distinctiveness, the majority (n=21) were well-differentiated, indicating the utility of the method as an identification tool, and we now use this method for routine diagnosis of xanthomonad species.

INTRODUCTION

Classification of species within the genus Xanthomonas, which was hitherto based on biochemical characteristics (Van den Mooter & Swings, 1990), underwent major revision after a comprehensive spectrophotometric DNA–DNA hybridization study of the genus (Vauterin et al., 1995), which resulted in the recognition of 20 constituent species. Following this study, names of seven further species have been validly published: X. cynarae (Trebaol et al., 2000), X. euvesicatoria, X. perforans, X. gardneri (Jones et al., 2004; Euzéby, 2006), X. citri, X. fuscans and X. alfalfae (Gabriel et al., 1989; Euzéby, 2007; Schaad et al., 2005, 2006, 2007).

The GenBank/EMBL/DDBJ accession numbers for the gyrB sequences of xanthomonad species determined in this study are EU007516–EU007542 and DQ676938.

Supplementary tables showing type strains of all xanthomonad species used in the study, PCR and sequencing primers and a similarity matrix indicating all inter-species percentage nucleotide identities are available with the online version of this paper.

The complexity of the genus and resource requirements for comprehensive pairwise analyses required for DNA–DNA hybridization analysis make this technique unsuitable for routine identification in most diagnostic laboratories. A rapid characterization method based on analysing cell-wall fatty acids using GC was developed and applied to the identification of plant pathogens (Stead, 1989, 1992). Whilst this method can be very useful for typing of some xanthomonads, it does not provide the necessary resolution or reproducibility to be useful for all species. More recently, a major and comprehensive analysis of the genus was completed (Rademaker et al., 2005) that compared amplicon profiles produced by using PCR primers designed to repeated sequences dispersed in bacterial genomes (Martin et al., 1992; Louws et al., 1994; Thwaites et al., 1999). This study utilized all three commonly used repetitive PCR methods (REP, BOX and ERIC) to produce a high-resolution identification scheme, and encompassed many of the pathovars of X. campestris that could not be resolved in the earlier study by Vauterin...
et al. (1995) using DNA–DNA hybridization. Comparison of gel profiles, however, still presents problems outside reference laboratories, and inter-laboratory standardization can be problematic. DNA sequencing studies of the 16S rRNA gene and the 16S–23S intergenic region of Xanthomonas species (Hauben et al., 1997; Gonçalves & Rosato, 2002) have indicated the phylogenetic position of the genus within the Gammaproteobacteria. Whilst the 16S–23S phylogeny had greater resolution than that produced from the 16S study, most of the species within the genus fell into a single group, and neither method could be used reliably to differentiate the majority of species.

Recently, DNA sequencing of genes encoding conserved proteins involved in essential cell processes that constitute the ‘core genome’ has been used to produce phylogenies that are more resolving than 16S and 16S–23S studies. Increasingly, this approach is being adopted for bacterial identification, as the method provides a convenient and rapid means of providing information on the relatedness of species and strains. This study aimed to produce a xanthomonad phylogeny based on partial sequence alignments of the gyrB gene, both to indicate xanthomonad relatedness and as a simple and reliable means of species-level identification.

As published DNA–DNA hybridization values are available for species comparisons (Vauterin et al., 1995), it was possible to relate percentage gyrB sequence identities to these values for several species comparisons, and to produce a scatter graph to clarify the relationship between the two methods for quantifying DNA sequence similarity. During the study, it was found that X. euvesicatoria, X. perforans and X. alfalfae had identical gyrB sequences. To address this anomaly and to clarify relatedness between these species, partial sequences from an additional five genes were analysed. Concatenated sequences from these loci were produced and compared in a phylogeny along with similar sequences derived from genome sequences for X. axonopodis pv. citri, X. oryzae and X. campestris.

METHODS

Bacterial strains. Type strains of all xanthomonad species used in the study were obtained from international culture collections (mostly the National Collection of Plant Pathogenic Bacteria; NCPPB) and their reference numbers, together with alternative strain references from other collections, are indicated in Supplementary Table S1, available in IJSEM Online. Species are as indicated in the phylogram (Fig. 1).

 gyrB locus and PCR amplification. Primers were designed from xanthomonad alignments of gyrB sequences available from GenBank and corresponded to the C-terminal region of the protein. In relation to the gyrB gene from X. euvesicatoria (strain 85-10; GenBank accession no. CAJ23349), the region amplified corresponded to aa 603–778, of a total of 832 residues for this protein. DNA, prepared from suspensions of freshly grown xanthomonad cultures, in water was adjusted to an A600 of 0.1 using a spectrophotometer, and then heated in a microfuge tube for 6 min at the highest temperature setting (100 °C) in a dry block heater. After cooling, the prepared DNA suspensions were frozen at −20 °C prior to use, when 1 μl preparation was added to 24 μl PCR reagent mix. Centrifuging of the boiled cells or other purification treatment was not required.

As xanthomonads can have very high G + C contents, a blend of Taq polymerase and a proof-reading polymerase was used for all amplifications (Long PCR Enzyme mix; Fermentas). The PCR buffer included magnesium and was as supplied by the enzyme supplier. The reaction components comprised: forward primer (10 pmol μl−1), 0.75 μl; reverse primer (10 pmol μl−1), 0.75 μl; 10 × PCR buffer (including 15 mM MgCl2), 2.5 μl; dNTP mix (2.5 mM each), 2.0 μl; Long PCR Enzyme mix, 0.125 μl; prepared DNA, 1.0 μl; water, 17.9 μl. The final concentrations of reagents were: MgCl2, 1.5 mM; dNTPs, 0.2 mM; primers, 0.3 μM each; enzyme mix, 0.75 units (all per 25 μl reaction volume). PCR and sequencing primers are shown in Supplementary Table S2, available in IJSEM Online. We routinely used standard primers that produced large yields of PCR product for all species except X. albilineans and related xanthomonads; for these species, the alternative conserved primer set was used. The conserved primers were designed from accrued gyrB alignments from this study and can be used for gyrB amplification from all species. The PCR cycling conditions were 2.5 min at 94 °C, then 30 s at 94 °C, 45 s at 50 °C and 1 min at 68 °C for 34 cycles, then 7 min at 68 °C to end.

Purity and yield of PCR product for sequencing were checked by running 5 μl reaction mixture on a 1.5% agarose gel and post-staining using ethidium bromide, according to standard protocols (Thwaites et al., 1999). The remaining mixture was purified by using a commercial kit to remove residual primers and other reaction components (Wizard PCR Clean-up kit; Promega). After elution of bound DNA in 50 μl water, the equivalent of approximately 12 μg DNA (usually 5–10 μl purified product) was tranferred to a fresh microfuge tube and lyophilized prior to sequencing from a commercial service (MWG Ltd), using a Sanger sequencing-based fluorescent-label PCR method. The sequences, provided in Fasta format files, were cropped further to provide a 530 nt sequence, corresponding to nt 1807–2336 in the gyrB sequence from X. euvesicatoria strain 85-10. The cropped sequences all start with ATCACCGG; the end of the sequence finishes (with some variants) with AGCTGTG.

Sequencing of X. euvesicatoria, X. perforans and X. alfalfae using additional gene loci. PCR amplification and sequencing primers for a further five loci [sigma factor 70 (rpoD), aconitate hydratase B (acnB), phosphoglucoisomerase (pgi), glyceraldehyde-3-phosphate dehydrogenase (gap) and phosphofructokinase (pFKA)] were designed from the X. euvesicatoria genome sequence (strain 85-10) available from GenBank (accession no. CAJ23349) (Thieme et al., 2005). PCR conditions, cycling parameters and sequencing protocol were the same as used above for the gyrB analysis. PCR and sequencing primers for multi-locus sequence typing of X. euvesicatoria, X. perforans and X. alfalfae are shown in Supplementary Table S3, available in IJSEM Online.

Concatenated sequences were compared with the same loci derived from GenBank for X. campestris and X. axonopodis pv. citri (da Silva et al., 2002), as well as for X. oryzae (Lee et al., 2005).

Sequence alignment and phylogenetic analysis. All loci were sequenced using both forward and reverse primers. Unprocessed gyrB sequences were cropped to a common start and finish sequence to produce 530 nt sequences that were aligned by using the CLUSTAL W program (Thompson et al., 1994), available at http://www.ebi.ac.uk/clustalw. Phylogenies were rooted by using the gyrB sequence from a laboratory isolate that, by 16S rRNA gene sequencing, was related most closely to the genus Stenotrophomonas. The TREESON program (Van de Peer & De Wachter, 1993) was used for phylogenetic analysis.
Distance estimation was done by using two methods available in the program (Jukes & Cantor, 1969; Tajima & Nei, 1984). Tree construction used the neighbour-joining method (Saitou & Nei, 1987).

RESULTS

Xanthomonas gyrB phylogeny

The phylogeny indicates a monophyletic classification, in agreement with previous studies comparing 16S rRNA gene and 16S–23S intergenic sequences (Fig. 1). Sequence alignments revealed the absence of any insertions or deletions, consistent with a single origin for the genus. Of 530 nucleotide positions, 55.3% were invariant. Inter-species sequence similarities ranged between 75 and 100%; a similarity matrix indicating all inter-species percentage nucleotide identities is provided in Supplementary Table S4, available in IJSEM Online.

The two most ancestral species within the phylogeny were X. albilineans and X. sacchari. The relative phylogenetic positions of these species, and of the clade containing X. theicola, X. translucens and X. hyacinthi, were supported by low bootstrap values. To account for this ambiguity in the relationship between these basal species, which occurs often in the most ancestral regions of phylogenies (Sarker & Guttman, 2004), they are designated ‘early-branching species’ in the phylogram. In contrast, three major clades representing the remaining species were supported by high bootstrap values (>77%). Group 1 comprised X. cassiniae, X. codiaeae and X. cucurbitae. Group 2 comprised X. arboricola, X. campestris, X. populi, X. hortorum, X. gardneri and X. cynarae, which are all species indigenous to Europe. Group 3 comprises a large group of the remaining species,

![Phylogenetic tree diagram of Xanthomonas gyrB]

**Fig. 1.** gyrB phylogeny of the genus *Xanthomonas* produced by using the neighbourhood-joining method. Sequences were analysed by using the method of Jukes & Cantor (1969). Percentage bootstrap values >50% from 500 samplings are indicated. All species are type strains, except for *X. hortorum* pv. *pelargonii*. The phylogeny was rooted by using a laboratory isolate that, by 16S rRNA gene sequencing, was related most closely to the genus *Stenotrophomonas*. Bar, 0.1 substitutions per site.
and contained two further clades (3A and 3B) that were supported by a high bootstrap value (100%). Clade 3A comprised *X. perforans*, *X. euvesicatoria*, *X. alfalfa*, *X. melonis*, *X. axonopodis*, *X. fuscans*, *X. citri* and *X. bromi*, whilst clade 3B comprised the monocot pathogens *X. vasicola* and *X. oryzae*.

Most of the reference type species (*n*=21) were distinguished clearly according to branch length. Exceptions were *X. perforans*, *X. euvesicatoria* and *X. alfalfa*, which had identical sequences. The relatedness of these species was investigated further by partial sequencing of an additional five loci and these results are described below. The type strains of *X. hortorum* (pv. *hederae*), *X. gardneri* and *X. cynarae* were distinguished only by very short branch lengths compared with other species. The *pelargonii* pathovar of *X. hortorum*, which was used in DNA–DNA hybridization studies to support *X. cynarae* and *X. gardneri* as distinct species, was separated from the type strain of *X. hortorum* with a branch length similar to those distinguishing other species of *Xanthomonas*, and was supported by a high bootstrap value (97%).

**Relationship between gyrB sequences and reported DNA–DNA hybridization values for paired species comparisons**

The percentage gyrB sequence similarity and published DNA–DNA hybridization values (Vauterin et al., 1995) for 24 species comparisons (Fig. 2) were selected for plotting on a scatter graph. As the reported hybridization values were means from groups of strains, those species comparison values with high standard errors (SEM) were avoided. The paired sequence identity and DNA–DNA hybridization values were plotted as a scatter plot by using the SigmaPlot software package. The scatter plot (Fig. 2) indicates a clear correlation between the published DNA–DNA hybridization data and gyrB percentage sequence identity for the same-paired species comparisons, supported by a Pearson correlation coefficient of 0.755.

**Relatedness between X. euvesicatoria, X. perforans and X. alfalfa*, determined by using additional gene loci**

Concatenated *rpoD*, *acnB*, *pgi*, *gap* and *pfkA* sequences were cropped to a total length of 2720 nt. The CLUSTAL W sequence alignment indicated scores of 99% sequence identity for all three comparisons of *X. euvesicatoria, X. perforans* and *X. alfalfa*. The unrooted phylogeny produced from the alignment (Fig. 3) indicated that these three species were most similar to *X. citri*, in agreement with the gyrB phylogeny. Branch lengths between the three species were very short, indicating that none of the additional five loci sequenced produced a species-level differentiation. The close sequence similarity found for all five additional loci confirms a close level of core-genome relatedness for *X. euvesicatoria, X. perforans* and *X. alfalfa*.

**DISCUSSION**

Until recently, phylogenetic analysis of plant-pathogenic bacteria based on protein-encoding genes has been used to analyse diversity within a species. Endoglucanase and *hrpB* genes have provided a detailed description of sequvar diversity within *Ralstonia solanacearum*, which identified five major phylotypes associated with distinct geographical regions (Poussier et al., 2000; Fegan & Prior, 2005). Seven loci were analysed in a large multi-locus sequence study of *Pseudomonas syringae* (Sarkar & Gutman, 2004; Sarkar et al., 2006) that provided high-resolution analyses of lineages within five major clades identified within this species. Sequence analysis revealed that recombination accounted for a very small proportion of nucleotide divergence compared with other bacteria. This was suggested to be attributable to adaptation to specific host plants, which restricted contact with other lineages and minimized genetic exchange. This feature may have contributed to a phylogeny that reflects very accurately the evolutionary history of this species. As multiple loci were analysed, it was possible to compare phylogenies that revealed close agreement between six of the seven loci examined.

Most recently, phylogenies have been developed that encompass species within higher-level taxa, including the
genus *Pseudomonas* and the family *Enterobacteriaceae* (Ait Tayeb et al., 2005; Paradis et al., 2005). Both of these studies revealed good correlation between established reference species and their phylogenetic identification. The two genes used to study the family *Enterobacteriaceae* (Paradis et al., 2005), elongation factor Tu (*tuf*) and F-ATPase β-subunit (*atpB*), produced phylogenies of similar resolution. However, comparison of the two loci used for analysis of the genus *Pseudomonas* (Ait Tayeb et al., 2005) revealed that the *rpoB* sequences produced a phylogeny of approximately three times higher resolution than that produced by using the alternative *rrs* gene, indicating that choice of locus can affect the degree of strain discrimination afforded by a phylogeny.

The *gyrB* sequences used in this study to produce the phylogeny (Fig. 1) were sufficiently distinct to discriminate the majority of established species in the genus *Xanthomonas*. The phylogram indicates that the genus is a monophyletic group derived from a common ancestor that acquired the ability to exploit plant tissue as an environmental niche. Four of the five species that branch at the base of the phylogeny, and which probably represent early-evolving species, are pathogens of monocot hosts, i.e. *X. albilineans*, *X. sacchari*, *X. hyacinthi* and *X. translucens*, and it is possible that the genus first arose as a monocot pathogen. Earlier studies using the 16S or 16S–23S loci and it is possible that the genus first arose as a monocot pathogen. Earlier studies using the 16S or 16S–23S loci (Hauben et al., 1997) and the 16S–23S intergenic loci (Hauben et al., 1997; Gonçalves & Rosato, 2002) also found *X. albilineans*, *X. hyacinthi*, *X. theicola*, *X. translucens* and *X. sacchari* to be the most ancestral species. The *gyrB* phylogeny identifies three major inter-species groups that are well supported by bootstrap analysis (groups 1–3), whilst a fourth comprises early-branching species of uncertain relatedness. Extensive species diversification has occurred within groups 2 and 3, which constitute the majority of the species within the genus. A bootstrap-value threshold of 70% is often used as a guideline to indicate a high level of confidence in phylogenogram branches, and the separation of groups 2 and 3 is well supported with a bootstrap value of 78%. Group 3 is the largest species-level group, and the most basal members of this clade comprised *X. fragariae*, *X. pisi* and the tomato pathogen *X. vesicatoria*. Clades 3A and 3B were supported by a 100% bootstrap value. Clade 3A comprises both monocot and eudicot pathogens, including economically important pathogens of bean, citrus, tomato and pepper. Clade 3B comprises the sorghum and rice pathogens *X. vasicola* and *X. oryzae*. DNA–DNA hybridization data are considered to be the ‘gold standard’ to measure relatedness between species. The scatter graph clearly indicates a correlation between DNA hybridization and the *gyrB* phylogeny, which is supported by the high Pearson correlation coefficient (0.755), and this forms the basis for the general agreement between the two methods of distinguishing xanthomonad species. The deviation from the calculated trend line for many comparisons may be attributable to experimental error, inherent in DNA–DNA hybridization studies. The presence of genetic elements associated with the flexible genome, especially those of large size or that occur as multiple copies, including phage, transposable elements, insertion sequences and plasmids, could all potentially affect DNA reassociation kinetics and hybridization values. These elements are a further potential source of deviation between DNA–DNA hybridization- and nucleotide sequence-based measures of relatedness.

The *gyrB* phylogeny clearly offers greater resolution than those produced previously from the 16S rRNA gene or 16S–23S intergenic loci (Hauben et al., 1997; Gonçalves & Rosato, 2002), and has resolved *X. citri* and *X. fuscans*, which were previously classified as pathovars of *X. axonopodis*. These two species, whose names were recently validly published, were identified by using a DNA–DNA hybridization technique (Schaad et al., 2005) that may be more discriminating than the spectrophotometric assay used previously (Vauterin et al., 1995). The *gyrB* phylogeny supports the distinctiveness of these two species.

The inability of the *gyrB* phylogeny to discriminate *X. euvesicatoria*, *X. perforans* and *X. alfalfae* was investigated by partial sequencing of a further five genes. None of these five loci was able to distinguish any of the three species with branch lengths consistent with species-level differentiation, confirming their close core-genome relatedness (Fig. 3). Similarity between *X. euvesicatoria* and *X. alfalfae* has been found recently in the REP-PCR study reported by...
Rademaker et al. (2005), which placed both species in the same group, designated 9.2. Relatively high DNA–DNA hybridization values (58%) between X. perforans and X. euvesicatoria have been reported (Jones et al., 2004). DNA–DNA hybridization data differ from sequence analysis in that the former are a measure of whole-genome DNA similarity, which may include elements of the flexible genome, e.g. phage, transposable elements and plasmids, that are often variable within strains of a species. These components could affect DNA reassociation kinetics and lead to differential DNA–DNA hybridization values. A large 190 or 155 kb plasmid has been reported in 75% of X. euvesicatoria strains (Wichmann et al., 2005). The genome-sequenced strain of X. euvesicatoria, 85-10, contains a large plasmid and three others, totalling 241,686 bp and representing 4.66% of the genome, which, taking into account that most plasmids are present in more than a single copy, constitutes a significant proportion of the genome. Whilst the plasmid contents of X. perforans and X. alfalfae are unknown, it is possible that the presence of plasmids in X. euvesicatoria may have contributed to DNA–DNA hybridization values and this may explain the disparity between the DNA hybridization and gyrB sequence data. The close core-genome relatedness of these three species that infect pepper, tomato and alfalfa suggests that evolution of the pathogens to enable utilization of different hosts may be proceeding relatively rapidly, before substantial substitutions in the core genome have had time to accumulate.

Two other species, X. cynarae and X. gardneri, could not be distinguished from each other or from the type strain of X. hortorum by using the gyrB phylogeny. The reference strain of X. hortorum used for comparison in the DNA–DNA hybridization study to discriminate X. cynarae from X. hortorum (Trébaol et al., 2000) was the pathotype reference strain of X. hortorum pv. pelargonii, rather than the species type strain, which belongs to X. hortorum pv. hederae. The study by Vauterin et al. (1995) recognized heterogeneity within X. hortorum and reported DNA–DNA hybridization values of 77% between these two pathovars. The gyrB phylogeny placed X. hortorum pv. pelargonii as a distinct lineage with a branch length close to that differentiating other species, indicating that this pathotype was a poor choice to represent the species in the DNA–DNA hybridization studies. Consequently, X. cynarae may not have been differentiated adequately from X. hortorum by using DNA–DNA hybridization. The same X. hortorum pv. pelargonii reference strain was also used to distinguish X. gardneri as a separate species from X. hederae (Jones et al., 2004). We could not find reported DNA–DNA hybridization data comparing X. cynarae and X. gardneri. The phylogeny illustrates how the greater resolution of the gyrB phylogeny compared with DNA–DNA hybridization can clarify xanthomonad relatedness.

One issue that may need to be addressed in the future, if a phylogenetic approach is used for classification, is to decide what level of sequence divergence should be used to delineate a species. The phylogeny could distinguish 24 of the currently accepted species and clearly there is excellent potential for gyrB sequencing as a convenient and accurate means of indicating xanthomonad relatedness at the species and, possibly, strain or pathovar level, and we now use the method for routine xanthomonad identification. We are currently sequencing the gyrB locus for all of the xanthomonad pathovar type strains to provide a comprehensive phylogeny of the genus Xanthomonas.

ACKNOWLEDGEMENTS

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

REFERENCES


