MINIREVIEW

SPECIES AND INFRA-SPECIES PHYLOGENETIC DISCRIMINATION OF PSEUDOMONAD AND XANTHOMONAD PATHOGENS OF STONE FRUIT AND NUTS

N. Parkinson and J. Elphinstone

Food and Environment Research Agency (Fera), Sand Hutton, York, YO41 1LZ, UK

SUMMARY

Recent phylogenetic analysis of xanthomonad and pseudomonad plant pathogens based on comparison of nucleotide sequences from protein coding loci have proved to reliably classify strains according to core-genome relatedness. Increasingly larger strain populations have been analysed from both pathogen groups, which has defined their diversity in terms of phylogroups, clades and sequevars. The level of discrimination from a single locus is sufficient to facilitate rapid and cost effective identification to species and subspecies level. Analysis of 56 pseudomonad pathogens including all the pathovar reference strains isolated from Prunus, hazel and walnut using the rpoD locus, has placed these pathogens within the phylogenetic structure defined previously using seven loci. The large majority of the strains grouped within phylogroups 1-3. Within Xanthomonas these pathogens are restricted to X. arboricola. The development of sequence reference data-bases to which isolates encountered in the diagnostic laboratory can be compared, will provide opportunities to improve risk assessment, diagnosis, monitoring and surveillance of these important pathogens.

DEVELOPMENT OF PHYLOGENETICS

Classical amplicon profile or ‘fingerprint’-based techniques generated using primers which bind to multiple sites in purified DNA have provided a means for species and sub-species identification of Xanthomonas and the Pseudomonas syringae group of plant pathogens (Rade-maker et al., 2005). However, whilst these methods can provide useful information for diagnostic identification when isolates are similar to reference strains or pathovars, these techniques have limited application in identifying true taxonomic relatedness and can be difficult to apply reliably across different laboratories. The low discrimination power of other classical (biochemical, serological) techniques also often prevent refined classification and identification. Consequently, many strains of both major groups were identified simply as undetermined pathovar and the relatedness of these strains to known reference types was unknown. Use of DNA hybridisation provides a means for determining relatedness of strains based on comparison of genomic similarity measured using DNA/DNA hybridisation kinetics. However, this method is technically difficult and expensive and has therefore generally only been applied to relatively small studies of a limited number of (reference) strains. Phylogenetic studies produced by comparison of 16S RNA nucleotide sequences have proved remarkably successful in elucidating relatedness down to genus level, but the conserved nature of this locus often limits discrimination of taxa below this level. Most recently, comparison of nucleotide sequences from protein coding loci (so-called housekeeping genes) have proved successful in phylogenetic discrimination of taxa at the species and infra-species level. Since the introduction of improved sequencing technologies (around 2000) using fluorochrome labelled nucleotides, sequencing of protein coding loci has rapidly gained acceptance. Today, the technique forms an important component of assessing population diversity and strain relatedness across most bacterial groups, including Xanthomonas and Pseudomonas, representatives of which are important stone fruit and nut pathogens that cause significant economic losses (Janse, 2006, 2010). Nucleotide sequences offer good levels of reproducibility which has reduced problems of inter-laboratory comparison of strains sometimes encountered using profile methods. Moreover, sequence data have facilitated the analysis of increasingly large populations. In addition to clarifying strain relatedness and diversity within the two groups, these studies have provided important reference data sets to compare strains. Perhaps the most surprising finding emerging from these studies is the absence of a clear relationship between a lineage and host range, with closely related strains being pathogenic to unrelated and disparate hosts. This suggests pathovars are diversifying broadly across lineages within both xanthomonad and pseudomonad groups. This potential for pathovar diversification to unrelated hosts is illustrated...
within xanthomonads isolated from India. Early published pathology studies of these pathovar type strains have confirmed their ability to produce disease in a wide taxonomic range of weed and minor crop species. Whilst extensive differential pathology studies were done establishing specific pathovar status, the taxonomic position of the pathovars was ambiguous, and for expediency these poorly classified strains were classified as *X. campestris*. Analysis of the 56 pathovar type strains from this region using a 530 bp gyrB sequence revealed two principle lineages, one of which comprised of 34 pathovars which had identical sequences, indicating extensive pathovar/host diversification of a sequenar (Parkinson et al., 2009). The simplicity of the technique and the degree of discrimination afforded has enabled unambiguous identification of hitherto difficult taxa and resolved the taxonomic position of many species. Some species have been found to be closely related whereas other previously unidentified strains have been found to represent new lineages. The advantages of the technique for identification and diagnosis of these pathogens have led to its current introduction into phyndiagnostic laboratories around the world.

Phylogenetic studies of the *P. syringae* complex using protein coding loci including gyrB rpoD rpoB and argK loci identified relatedness of nearly all the species type strains and some of the pathovar type strains (Sawada et al., 1999; Yamamoto et al., 2000; Ait Tayeb et al., 2005; Hwang et al., 2005). The relatedness of the *P. syringae*-complex species identified in these studies agreed with the 'genomospecies' groupings defined by (Gardan et al., 1999). This latter study used DNA/DNA hybridisation analysis and ribotyping to discriminate nine genomospecies. One of these genomospecies (G. sp. 2) grouped together *P. fuscereectae*, *P. savastanoi*, the almond pathogen *P. amygdali* and *P. meliae*. More recently a more extensive phylogenetic analysis using seven loci analysed more than 50 strains (Sarkar and Guttman, 2004) whereas the other species type strains were found to be polyphyletic within *P. amygdali*, *P. meliae* and *P. savastanoi*. A smaller population set was analysed using four loci (Young et al., 2008) which provided higher resolution of strains and confirmed the overall phylogenetic structure of the genus.

FORMAL BINOMIAL CLASSIFICATION AND PHYLOGROUP

In order to relate the formal classification to the *Pseudomonas* phylogroup scheme a 578 nucleotide region of the *rpoD* gene, that encompassed the same locus used by Sarkar and Guttman, was used to produce a phylogeny which included all species and pathovar type strains (N. Parkinson, unpublished information). This phylogram found that *P. fuscereectae*, *P. tremae*, *P. amygdali*, *P. meliae* and *P. savastanoi* are polyphyletic within phylogroup 3 (as defined in the study by Sarkar and Guttman, 2004) whereas the other species type strains each grouped into separate phylogroups.

SINGLE LOCUS DISCRIMINATION OF STONE FRUIT AND NUT PATHOGENS

The 578 bp *rpoD* phylogeny (Fig. 1) indicates stone fruit and nut pathogen diversity within the *P. syringae* complex. Since this locus encompassed the region used in the earlier multi-locus study (Sarkar et al., 2006) it was possible to assign strains to their phylogroup scheme by construction of a combined phylogeny utilising the 450 bp sequences available from GenBank originating from the previous study. All the GenBank sequences grouped together into groups relating to the phylogroups defined using all seven loci. The phylogenetic structure of *Prunus* and nut pathogens identified, using the *rpoD* locus, reveals considerable diversity within phylogroups 1-3. Some of the pathogens grouped with phylogroup 4 or 5 strains, or with the
clade containing *P. caricapapayae* (identified as phylogroup 7). Both *Prunus* and hazelnut pathogens occur in different phylogroups (Table 1) and would have evolved independently from each other. This confirms the diversity of strains pathogenic to hazel which has been previously identified by multi locus sequence analysis (Wang et al., 2007). There is some evidence that strains may possess cross pathogenicity since strains from different *Prunus* species group together in the same clades, though the host range of *Prunus* strains would be required to be verified through experimentation. As well as attribution of species and pathovars to the major phylogroups the 578 nucleotide *rpoD* locus provides sub-clade and seqevar information. Within phylogroup 1 two major clades are discriminated one containing *P. syringae* pv. *morsprunorum* and the other

---

**Fig. 1.** Phylogenetic structure of *P. syringae*-complex stone fruit and nut pathogens produced by neighbour-joining analysis of a 578 bp *rpoD* locus. *Prunus* and nut hosts are indicated next to NCPPB culture collection references. Pathovar and species type strains are indicated by PTS and STS. Phylogroups are as defined by Sarkar and Guttman (2004) except for phylogroups 6 (*P. viridiflava*) and phylogroup 7 (*P. caricapapayae* clade) allocated here (see text for details).
both *P. syringae* pv. *persicae* and *P. syringae* pv. *avii*.

Among the 203 xanthomonad pathovars and poorly characterised strains analysed using the 530 bp *gyrB* locus (Parkinson *et al.*, 2009), tree pathogens were found to be largely restricted to *X. arboricola*, and stone fruit and nut pathogens were restricted to this species (Fig. 2). Hazel and walnut strains were discriminated as distinct sequevars. Two pathogens of *Prunus domestica* occurred as distinct sequevars.

### Table 1. Phylogroup attribution for pseudomonad pathogens of *Prunus* and nut identified using the *rpoD* locus.

<table>
<thead>
<tr>
<th>Host</th>
<th>Common name</th>
<th>Pathogen phylogroup</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Prunus americana</em></td>
<td>wild cherry</td>
<td>PG 2</td>
</tr>
<tr>
<td><em>P. amygdali</em></td>
<td>almond</td>
<td>PG 3</td>
</tr>
<tr>
<td><em>P. armeniaca</em></td>
<td>apricot</td>
<td>PG 2</td>
</tr>
<tr>
<td><em>P. avium</em></td>
<td>bird cherry</td>
<td>PG 1+2</td>
</tr>
<tr>
<td><em>P. cerasifera</em></td>
<td>cherry plum</td>
<td>PG 1</td>
</tr>
<tr>
<td><em>P. cerasus</em></td>
<td>sour cherry</td>
<td>PG 1+2</td>
</tr>
<tr>
<td><em>P. domestica</em></td>
<td>plum</td>
<td>PG 1</td>
</tr>
<tr>
<td><em>P. mume</em></td>
<td>Japanese apricot</td>
<td>PG 2</td>
</tr>
<tr>
<td><em>P. persica</em></td>
<td>peach</td>
<td>PG 1+2</td>
</tr>
<tr>
<td><em>P. salicina</em></td>
<td>Japanese plum</td>
<td>PG 1</td>
</tr>
<tr>
<td>Corylus avellana</td>
<td>hazel nut</td>
<td>PG 2</td>
</tr>
<tr>
<td>Castanea sativa</td>
<td>sweet chestnut</td>
<td>PG 3</td>
</tr>
<tr>
<td>Juglans regia</td>
<td>walnut</td>
<td>PG 2</td>
</tr>
</tbody>
</table>

### BENEFITS OF SINGLE LOCUS SEQUENCE ANALYSIS IN DIAGNOSIS OF PATHOGENS

Sequence analysis using multiple loci (MLSA) has the advantage of providing greater phylogenetic information compared to a single locus. Strains are resolved to a high level of discrimination and the relative placement of the major branches indicating evolutionary history is produced with greater certainty. However, the use of MLSA in a diagnostic laboratory, at least as a first-level identification tool is costly, and increases the time needed to determine identification compared to a single locus analysis. It has been demonstrated that the 578 *rpoD* locus and 530 bp *gyrB* locus provide identification to sub-clade and sequevar level of pathogens within *Xanthomonas* and the *P. syringae* complex. Furthermore, analysis of large population sets is facilitated using a single locus. These studies are beginning to define the taxonomic position of previously unclassified strains and enable an improved view of diversity and relatedness within these two major plant pathogen groups. Diagnostic protocols are currently being re-evaluated for the inclusion of sequence-based diagnosis and identification into their testing schemes across European diagnostic laboratories. The description of new pathovars can require considerable resources which may restrict the study of pathogen diversity and obscure the presence of new pathogens. Sequence-based classification will provide a means for identifying both these potential new
ACKNOWLEDGEMENTS

This research was funded by Defra. Helpful discussions with expert specialists under the EU COST 873 and EUPHRESO projects are acknowledged. The license number for working with the quarantine bacteria is PHL 251.

REFERENCES


