Phylogenetic analysis of *Xanthomonas* species based upon 16S–23S rDNA intergenic spacer sequences

Edmilson R. Gonçalves and Yoko B. Rosato

**INTRODUCTION**

The xanthomonads constitute a group of bacteria with diverse physiological traits and phytopathological specializations. Mainly on the basis of studies of DNA–DNA reassociation, *Xanthomonas campestris* was split into 16 genomic species and, at the time of writing, the genus *Xanthomonas* contained 20 genomic species (Vauterin et al., 1995). Analysis of 16S rDNA sequences (Hauben et al., 1997) has revealed a very small degree of divergence among the xanthomonads, with just three phylogenetic lineages being detected. The largest lineage included 15 *Xanthomonas* species, indicating that the high degree of conservation of the...
The high level of similarity in the 16S rDNA within Xanthomonas prompted us to investigate the usefulness of another molecular phylogenetic marker, the 16S–23S rDNA intergenic spacer (ITS), for distinguishing between these species. PCR-RFLP of ITS sequences from several Xanthomonas species has established a consistent degree of variation among them (Gonçalves & Rosato, 2000). ITS sequences are located between the 16S and 23S ribosomal subunits and, in most Gram-negative bacteria, contain a tRNA^Ala and a tRNA^Asp in their sequence (for a review, see Gürtler & Stanisich, 1996). ITS sequences have been under less intensive selection pressure and are considered to be 10 times as variable as 16S rDNA (Leblond-Bourget et al., 1996).

In the present report, the complete sequences of 18 ITS regions from 17 Xanthomonas genomic species described by Vauterin et al. (1995) are determined. These data, together with other sequences available in the GenBank database, were used to assess the diversity of the ITS sequences and to infer phylogenetic relationships within the genus.

**METHODS**

**Bacterial strains and culture medium.** The following type and reference strains of Xanthomonas described by Vauterin et al. (1995) were acquired from the culture collection of the Laboratorium voor Microbiologie, Universiteit Gent, Belgium: X. axonopodis pv. axonopodis LMG 538^T, X. axonopodis pv. broni LMG 947^T, X. campestris pv. campestris LMG 568^T, X. cassavae LMG 673^T, X. codiacei LMG 8678^T, X. cucurbitae LMG 690^T, X. fragariae LMG 708^T, X. hortorum pv. hederae LMG 733^T, X. hyacinthi LMG 739^T, X. melonis LMG 8670^T, X. pisi LMG 847^T, X. sacchari LMG 471^T, X. theicola LMG 8684^T, X. translucens pv. translucens LMG 876^T and X. vasica pv. holicola LMG 736^T. Strains Xanthomonas alboflexuosa ICMP 196, X. arboricola pv. juglandis ICMP 35^T and X. axonopodis pv. passiflorae ICMP 3151 were provided by the International Collection of Microorganisms from Plants, Auckland, New Zealand. All strains were grown on NYG medium (Turner et al., 1984) at 30 °C.

**DNA extraction.** An aliquot (1-5 ml) of bacterial suspension, grown overnight, was washed with 500 μl TAE buffer (50 mM Tris/HCl, pH 7.8, 50 mM EDTA, 150 mM NaCl), centrifuged and resuspended in 450 μl TAE buffer. SDS (50 μl; 10%, w/v) and proteinase K (100 μg) were added and the mixture was incubated for 1 h at 55 °C. Proteins and cell debris were extracted with phenol/chloroform and the aqueous suspension was then dialysed against TE buffer (10 mM Tris/HCl, pH 7.8, 1 mM EDTA) for 48 h. The DNA concentration was estimated by comparison with a known concentration of λ DNA in agarose gel electrophoresis (Sambrook et al., 1989).

**PCR and cloning.** Amplification of the ITS region was carried out using primers 16S uni 1330 (5'-GGTTCCCGCCGCTTG-TGACACAC-3') and 235 uni 322 anti (5'-GTTTCTTCTTCACCCTCCCTC-3'), described by Honeycutt et al. (1995). All PCRs were performed in a total volume of 25 μl using 30–50 ng DNA, 1.25 mM MgCl₂, 100 μM dNTPs, 0.5 mM each primer and 0.5 U Taq polymerase (Pharmacia). Cycling conditions were: one cycle of 94 °C for 30 s and 40 cycles of 94 °C for 30 s, 65 °C for 30 s and 72 °C for 1 min. PCR products were cloned using a pGEM-T easy kit (Promega).

**DNA sequencing.** Plasmids containing the ITS amplified fragments were extracted using a concert rapid plasmid extraction kit (Qiagen).

**Table 1.** Length and G + C content of ITS sequences from Xanthomonas species

<table>
<thead>
<tr>
<th>Strain</th>
<th>Spacer length (bp)</th>
<th>G + C content (mol %)</th>
<th>GenBank accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IT51</td>
<td>IT52</td>
<td>IT53</td>
</tr>
<tr>
<td>X. albilineans ICMP 196</td>
<td>146</td>
<td>14</td>
<td>223</td>
</tr>
<tr>
<td>X. arboricola pv. juglandis ICMP 35^T</td>
<td>109</td>
<td>19</td>
<td>226</td>
</tr>
<tr>
<td>X. axonopodis pv. axonopodis LMG 538^T</td>
<td>113</td>
<td>18</td>
<td>232</td>
</tr>
<tr>
<td>X. axonopodis pv. passiflorae ICMP 3151^T</td>
<td>113</td>
<td>18</td>
<td>229</td>
</tr>
<tr>
<td>X. broni LMG 947^T</td>
<td>112</td>
<td>19</td>
<td>228</td>
</tr>
<tr>
<td>X. campestris pv. campestris LMG 568^T</td>
<td>112</td>
<td>19</td>
<td>225</td>
</tr>
<tr>
<td>X. cassavae LMG 673^T</td>
<td>113</td>
<td>19</td>
<td>228</td>
</tr>
<tr>
<td>X. codiacei LMG 8678^T</td>
<td>113</td>
<td>79</td>
<td>228</td>
</tr>
<tr>
<td>X. cucurbitae LMG 690^T</td>
<td>110</td>
<td>19</td>
<td>219</td>
</tr>
<tr>
<td>X. fragariae LMG 708^T</td>
<td>114</td>
<td>19</td>
<td>226</td>
</tr>
<tr>
<td>X. hortorum pv. hederae LMG 733^T</td>
<td>113</td>
<td>19</td>
<td>226</td>
</tr>
<tr>
<td>X. hyacinthi LMG 739^T</td>
<td>103</td>
<td>75</td>
<td>166</td>
</tr>
<tr>
<td>X. melonis LMG 8670^T</td>
<td>112</td>
<td>75</td>
<td>230</td>
</tr>
<tr>
<td>X. pisi LMG 847^T</td>
<td>110</td>
<td>75</td>
<td>227</td>
</tr>
<tr>
<td>X. sacchari LMG 471^T</td>
<td>106</td>
<td>15</td>
<td>218</td>
</tr>
<tr>
<td>X. theicola LMG 8684^T</td>
<td>105</td>
<td>77</td>
<td>208</td>
</tr>
<tr>
<td>X. translucens pv. translucens LMG 876^T</td>
<td>103</td>
<td>76</td>
<td>208</td>
</tr>
<tr>
<td>X. vasica pv. holicola LMG 736^T</td>
<td>113</td>
<td>19</td>
<td>227</td>
</tr>
</tbody>
</table>
miniprep system (Gibco BRL) and sequencing reactions were done with an ABI Prism BigDye terminator cycle-sequencing kit (Applied Biosystems) using both forward and reverse M13 primers. The sequences were read with an ABI 373 automatic sequencer (Perkin Elmer). All sequencing procedures were repeated at least twice for each strain.

**Analysis of sequence data.** ITS sequences were aligned using the clustal w program (Thompson et al., 1994) prior to calculation of the similarities. Evolutionary distances were obtained by the p-distance method. Topology of the phylogenetic tree was assessed by the neighbour-joining method (Saitou & Nei, 1987) and bootstrap values were obtained from 2000 replicates using the mega (Molecular Evolutionary Genetic Analysis) package, version 1.01 (Kumar et al., 1993). The full sequence alignments are available as supplementary material in IJSEM Online (http://ijs.sgmjournals.org/).

**Nucleotide sequence accession numbers.** The 18 ITS sequences determined in this work were deposited in the GenBank database under the accession numbers shown in Table 1. These sequences were compared to those of X. campestris pv. badrii (strain not specified; L42974), X. campestris pv. gardneri XCGA (AF123093), Xanthomonas oryzae H8319 (AB026287) and Xanthomonas vesicatoria XV1111 (AF123088) available in GenBank. The sequence of Stenotrophomonas maltophilia ATCC 13637T (L28166) was used as an outgroup reference strain.

**RESULTS AND DISCUSSION**

**Structure and similarity of ITS sequences among Xanthomonas spp.**

ITS fragments from 18 strains representing 17 genomic groups of Xanthomonas and X. axonopodis pv. passiflorae were amplified using primers that bind to conserved regions of the 16S and 23S rRNA genes. A unique PCR product of approximately 1·1 kb was generated for all Xanthomonas strains. Nucleotide sequences of the PCR products were determined and the region spanning the ITS was aligned to other known Xanthomonas sequences deposited in GenBank. Sequence similarity analysis was performed using all these Xanthomonas species and S. maltophilia. The number of nucleotides ranged from 492 to 578 and extensive similarity (63–99% mean 83±10·4%) was found among the Xanthomonas species. This result contrasts with the mean level of similarity of 98±2% found in the 16S rDNA sequences (Hauben et al., 1997). The lowest level of ITS similarity was for X. hyacinthi (63±83·2%), followed by X. theicola and X. translucens (70±1–89·1%). The 16S–23S rDNA ITS sequences showed greater (approx. ninefold) diversity (16·2%) than the 16S rDNA sequences (1·8%). This result is similar to that reported by Leblond-Bourguet et al. (1996) for the genus Bifidobacterium, where the evolutionary rate of the 16S–23S rDNA ITS sequence was 10 times that of the 16S rDNA sequence.

Analysis of ITS sequences revealed the presence of tRNA\(^{Ala}\) and tRNA\(^{Leu}\) genes (Fig. 1) interspersed within the sequence, thus confirming previous results for X. albilineans (Honeycutt et al., 1995). Both tRNAs were highly conserved in all Xanthomonas species and S. maltophilia, with only small differences in size and few nucleotide changes. tRNA\(^{Ala}\) was 75–79 bases in size and the predicted anticodon was UGC. Nine Xanthomonas species showed identical sequences; in the other species, a total of 19 changes (15 substitutions and 4 deletions) of single bases was observed. Most of these alterations (10/19) occurred at positions 4 and 69. X. albilineans, X. hyacinthi, X. sacchari, X. translucens and X. theicola, herein referred to as the hyacinthi group, showed a group-specific substitution from A to T at position 69. tRNA\(^{Leu}\) varied from 73 to 79 bases in size and the anticodon was GAU. The sequence was slightly more variable than tRNA\(^{Ala}\) and, overall, 56

---

**Fig. 1.** Overall structure of the 16S–23S rDNA ITS of Xanthomonas showing the location of the tRNA genes and the nucleotide sequence of ITS2. Conserved nucleotide blocks are indicated in bold; consensus nucleotides are indicated by asterisks.

http://ijs.sgmjournals.org
base changes were observed. The most frequent sites of nucleotide change were at positions 17, 59 and 68–69. *X. axonopodis* showed the largest number of nucleotide changes (10), which were either single or contiguous (2–3 bases).

The tRNAs split the ITS into three regions (ITS1, ITS2 and ITS3). Distinct ITS regions from *Xanthomonas* spp. showed variable degrees of divergence (Table 1). ITS1 was the most conserved region and the mean sequence similarity among the *Xanthomonas* species was 87.1 ± 9.2%. *X. arboricola*, *X. campestris* pv. *gardneri*, *X. cucurbitae*, *X. pisi* and *X. vesicatoria* showed identical ITS1 sequences. Sequence differences in the other species consisted mainly of changes in single nucleotides. The deletion of 1–3 contiguous bases was seen in the *hyacinthi* group (positions 40, 88–90, 116–118 or 125–127). This group also showed specific base substitutions not found in other *Xanthomonas* species. For example, at positions 30–31, AG replaced the consensus sequence GT. Other single base changes, also unique to this group, were found at positions 60, 93, 97, 104 and 111.

Major differences in size were found in the ITS2 region. Most *Xanthomonas* species had 19 nt (named ITS2-S), but five species (*X. codiaeai*, *X. hyacinthi*, *X. melonis*, *X. theicola* and *X. translucens*) exhibited a longer sequence, of 78–85 bases (ITS2-L), caused either by the addition of a long stretch of 51–56 nt and two smaller (CCT and GCA CA) sequences or by the loss of these sequences from the other species (Fig. 1). Although ITS2-S was a shorter sequence, formation of a small secondary hairpin structure could be predicted (see *X. axonopodis*; Fig. 2). This structure may be a putative target for RNase III during processing of tRNA^{Ala} and tRNA^{Ile} in most xanthomonads, where the sequence is present. A more complex secondary structure was predicted for ITS2-L, found also in *X. campestris* pv. *badrii* (Figs 1 and 2). The presence of ITS2-L in six xanthomonad species and *S. maltophilia* is suggestive of a common origin for this sequence among these organisms. The ITS2-L sequence of *X. campestris* pv. *badrii* described by Honeycutt *et al.* (1995) showed a 66-base stretch that shared > 60% identity with the corresponding sequence from *S. maltophilia*, but was distinct from the sequences in other *Xanthomonas* species. Interestingly, most of the ITS2-L sequences shared a low level of similarity. Alignment of these sequences revealed a block of 14 highly conserved nucleotides close to tRNA^{Ala} (see Fig. 1; in bold). This conservation may be related to the formation of secondary structures. When ITS2-L was present, different secondary structures could be predicted (Fig. 2). A longer stem, which provided greater stability to the structure, was found in *X. campestris* pv. *badrii*, *X. codiaeai*, *X. hyacinthi,* and also X. theicola.
ITS phylogenetic tree of *Xanthomonas* spp.

**Fig. 3.** Neighbour-joining tree based on the 165–235 rDNA ITS sequences showing relationships between species of the genera *Xanthomonas* and *Stenotrophomonas*. Distances were determined by the p-distance method using aligned sequences by **CLUSTAL W**. Numbers on branches indicate the percentage of the number of times that the node was supported by 2000 replicates of bootstrap analysis (only values greater than 80% are shown). Bar, 0·01 changes per nucleotide.

*X. theicola* and *X. translucens*, whereas *X. melonis* and *S. maltophilia* had a slightly shorter stem. The relevance of the conserved block of 14 nt is not clear, since its participation in the formation of the predicted secondary structure varied. The four upstream bases take part in the first loop in most cases (arrows in Fig. 2), but the role of the remaining 10 bases is unknown.

The addition of 51–56 nt in the longer ITS2 sequence represents a remarkable difference in the ITS sequence. The sharp discontinuity in the ITS sequence caused by this interposing sequence suggests that horizontal transfer of the whole sequence occurred in the past. As there was great diversity in the five sequences, except for the conservation of the 14 upstream bases, the transfer event involved was probably not recent. Despite the great diversity in ITS2-L sequences, a putative secondary structure was predicted in all strains. Another possibility to explain the presence of the ITS2-L sequence would be a recombination process between different rDNA operons, which could lead to divergence between genes in different species. A conclusive determination of the number of rDNA copies in *Xanthomonas* is lacking, but at least two copies of unknown sequence have been described in *X. campesstris* pv. *campestris* (Lin & Tseng, 1997). Conservation of nucleotide blocks at the flanking ends of the spacer and rearrangements in the ITS2 region are suggestive of a recombination process. ITS sequences of several bacterial groups are now under intense investigation and the availability of a larger number of sequences for comparison would help to elucidate the importance of these findings.

ITS3, the longest sequence, varied from 166 to 232 bases and was less conserved than ITS1, with a mean similarity of 82·3±11·5% among *Xanthomonas* species. Several changes affecting 1–5 bases were found in most sequences and greater divergence was observed for the *hyacinthi* group. *X. hyacinthi* showed a large deletion of 44 contiguous bases at positions 103–146.

**The G+C content of ITS sequences**

The total G+C content in *Xanthomonas* was 63·3–69·7 mol% (for a review, see Swings & Civerolo, 1993) whereas, in the ITS sequence, the range was 50·4–57·1 mol%. The low G+C content of the ITS sequences contrasted with that of the 16S rDNA (69 mol%), indicating that the selective pressure was not identical in coding and non-coding regions. The similarities of the tRNA<sup>Ile</sup> and tRNA<sup>Ala</sup> were also higher (98·05±2·8 and 93·4±3·6, respectively) than those of the 16S rDNA. The G+C content of tRNA<sup>Ala</sup> agreed with the general G+C content within the species (64·5±0·3 mol%), whereas the tRNA<sup>Ala</sup> had a lower G+C content (59·6±2·0 mol%). There is no explanation for this lower G+C content in the ITS sequences, but similar situations occur in other prokaryotes like *Salmonella typhimurium* and *Escherichia coli* (Syvanen et al., 1994).

**Phylogenetic analysis of ITS sequences**

Phylogenetic relationships among the *Xanthomonas* species and *S. maltophilia* were assessed by the p-distance method and the ITS phylogenetic tree was constructed using the neighbour-joining method (Fig. 3). The *Xanthomonas* species clustered together...
and were well separated from *S. maltophilia*, used as an outgroup strain. Six major clusters with bootstrap values higher than 98% were formed. Cluster I contained most *Xanthomonas* species (*X. arboricola*, *X. axonopodis*, *X. bromii*, *X. campestris*, *X. cassavae*, *X. codiaeai*, *X. cucumbitaec*, *X. fragariae* *X. horitorum*, *X. melonis*, *X. oryzae*, *X. pisi*, *X. vasicola* and *X. vesicatoria*). Although *X. codiaeai* and *X. melonis* had longer ITS sequences because of the presence of ITS2-L sequences, they were not differentiated from the other members of cluster I, since the pairwise comparison was performed by partial deletion. Some well-accepted species within the *Xanthomonas* genus (*X. oryzae* and *X. fragariae*) were included in cluster I, with no clear differentiation between them. *X. campestris pv. gardneri*, recently suggested as a novel *Xanthomonas* species, ‘*Xanthomonas gardneri*’ (Jones et al., 2000), and herein accepted as a pathovar of *X. campestris*, and most of the species suggested by Vauterin et al. (1995) and Schaad et al. (2000) were also included in cluster I. The other clusters (II–VI) included *X. albilineans*, *X. hyacinthi*, *X. sacchari*, *X. theicola* and *X. translucens* as representative strains of distinct clusters. Of this group, *X. hyacinthi*, *X. theicola* and *X. translucens* were remarkably different from the cluster I species, including substitution of bases in ITS1 (positions 60, 93, 97, 104 and 111) and the presence of the ITS2-L sequence. The G+C content was higher in these species (55.4–57.1 mol%) than in cluster I (50.4–53.5 mol%). The distinction of *X. hyacinthi*, *X. theicola* and *X. translucens* based on their G+C content was similar to that reported for their 16S rDNA sequences (Hauben et al., 1997).

The topology of the ITS tree showed considerable similarities to the 16S rDNA phylogenetic tree obtained by Hauben et al. (1997). These authors found three major clusters: cluster 1, comprising most *Xanthomonas* species; cluster 2, represented by *X. sacchari*; and cluster 3, including *X. albilineans*, *X. hyacinthi*, *X. theicola* and *X. translucens*. In the ITS tree, the species included in cluster 3 of the 16S rDNA tree were split into four distinct clusters (III–VI). Differentiation of these clusters was not easy when based on pathological characteristics, an unreliable taxonomic marker. The ITS sequence of *X. hyacinthi* showed the highest level of divergence. The symptoms caused by *X. hyacinthi* were the first description (in the 1870s) of a plant disease caused by a bacterium, restricted almost exclusively to The Netherlands (see Swings & Civerolo, 1993). Therefore, it has been maintained in isolation in a specific region with no worldwide dispersion. *X. sacchari* and *X. albilineans* infect the same host (*Saccharum officinalis*), but there is a clear distinction between the species (Vauterin et al., 1992). With the exception of *X. theicola*, all the species belonging to clusters II–VI are pathogenic to monocotyledonous plants. The grouping of these species has also been reported on the basis of 16S rDNA sequences (Hauben et al., 1997).

Phylogenetic relationships and major lineages among *Xanthomonas* species could be delineated with higher resolution with ITS than from the 16S rDNA tree, even though the large cluster, containing most of the *Xanthomonas* species, remained unresolved. Other approaches showing higher discriminative resolution could be used to reveal the distinctiveness of each genomic group, as described previously (Gonçalves & Rosato, 2000; Rademaker et al., 2000).

**ACKNOWLEDGEMENTS**

This work was supported by FAPESP (Proc. 90/7067–1). CNPq sponsored a studentship to E.R.G. The authors thank L. O. S. Berian (IB, SBF), J. Rodrigues Neto (IB, SBF) and R. P. Leite (IAPAR) for providing the *Xanthomonas* strains and S. Hyslop for revising the English.

**REFERENCES**


Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W:

