Evolutionary placement of Xanthomonadales based on conserved protein signature sequences

Ania M. Cutiño-Jiménez a, Marinalva Martins-Pinheiro b, Wanessa C. Lima b, Alexander Martín-Tornet c, Osleidy G. Morales a, Carlos F.M. Menck b,*

a Department of Biology, Facultad de Ciencias Naturales, Universidad de Oriente, Ave. Patricio Lumumba s/n, Santiago de Cuba, CP 90 500, Cuba
b Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, Ave. Prof. Lineu Prestes, 1374, São Paulo, SP, Brazil
c Department of Informatics, Centro de Ingeniería Genética y Biotecnología (CIGB), Ave. 31 e/158 y 190, Playa, P.O. Box 6162, Habana 10600, Cuba

Article history:
Received 14 May 2009
Revised 11 September 2009
Accepted 21 September 2009
Available online xxxx

Keywords:
Xanthomonadales
Xanthomonas
Xylella
DNA repair
INDELs

ARTICLE INFO

ABSTRACT

Xanthomonadales comprises one of the largest phytopathogenic bacterial groups, and is currently classified within the gamma-proteobacteria. However, the phylogenetic placement of this group is not clearly resolved, and the results of different studies contradict one another. In this work, the evolutionary position of Xanthomonadales was determined by analyzing the presence of shared insertions and deletions (INDELs) in highly conserved proteins. Several distinctive insertions found in most of the members of the gamma-proteobacteria are absent in Xanthomonadales and groups such as Legionellales, Chromati-ales, Methylococcales, Thiotrichales and Cardiobacteriales. These INDELs were most likely introduced after the branching of Xanthomonadales from most of the gamma-proteobacteria and provide evidence for the phylogenetic placement of the early gamma-proteobacteria. Moreover, other proteins contain insertions exclusive to the Xanthomonadales order, confirming that this is a monophyletic group and provide important specific genetic markers. Thus, the data presented clearly support the Xanthomonadales group as an independent subdivision, and constitute one of the deepest branching lineage within the gamma-proteobacteria clade.

1. Introduction

With the advent of the genomic era, new molecular approaches to identify evolutionary relationships among organisms have come to light, and since then extensive work has been done in bacterial taxonomic classification, mainly focusing in the use of conserved protein sequences to reconstruct evolutionary histories. However, the use of single gene or protein sequences to reconstruct organism evolution has been contested (Korbel et al., 2002; Ochman, 2001), mainly due to inherent methodological problems during phylogenetic reconstruction and analysis or to systematic errors from the sequences. Moreover, phylogenetic trees based on rRNA or protein sequences are unable to determine how certain divisions within Bacteria are related to each other and how they have evolved from a common ancestor (Gupta and Griffiths, 2002). New sequence-based approaches for assessing evolutionary relationships have reduced some of the biases in earlier methodology that can lead to the misinterpretation of phylogenetic results. Detection of signature sequences in the primary structure of a protein is a reliable and intuitive method for examining evolutionary relationships. Specific changes in protein residues, such as amino acid substitutions or specific deletions or insertions, observed in all members of one or more taxa but not in others, make it possible to establish common ancestry and identify major groups that share the same signature (Brocchieri, 2001; Gupta, 1997).

There are numerous mechanisms by which specific insertions and deletions (herein referred to as INDELs) may be formed, such as, for example, DNA recombination, expansion of repetitive DNA sequences and insertion sequence (IS)-mediated events. As they are, in general, the product of one specific event, INDELs were proposed as reliable signature sequences with certain advantages over traditional phylogenetic analyses based on gene or protein sequences (Griffiths and Gupta, 2007; Gupta, 1998, 2006; Gupta and Mok, 2007). Phylogenetic relationships assume constant mutation frequencies and this may be incorrect over long periods of time, ultimately leading to the identification of incorrect species relationships (Philippe et al., 2005). In contrast, conserved INDELs of defined sizes are not greatly affected by such differences in evolutionary rates (Gupta, 1998).
Based on a 16S rRNA tree, the proteobacteria phylum has been divided into five phylogenetically distinct groups (gamma, beta, alpha, delta and epsilon subdivisions or classes) (Olsen et al., 1991; Stackebrandt and Woese, 1984). These subdivisions constitute well-defined clades, their member species being clearly distinguishable from the other and from the other bacterial divisions (Eisen, 1995; Ludwig and Schleifer, 1994; Stoffels et al., 1999). In support of the use of signature sequences as evolutionary descriptors, all proteobacterial subdivisions contain specific INDELS in some of their protein sequences, confirming the evolutionary relationships inside the proteobacteria group (Gupta, 2000, 2005, 2006; Gupta and Mok, 2007).

The gamma subdivision is a well-defined group that probably diverged from the rest of proteobacteria as early as 500 million years ago (Clark et al., 1999). It is one of the largest groups within Bacteria and includes at least 14 different subgroups, or orders (Acidithiobacillales, Aeromonadales, Alteromonadales, Cardiacobacterales, Chromatiales, Enterobacteriales, Legionellales, Methylococcales, Oceanospirillales, Pasteurellales, Pseudomonadales, Thiotrichales, Vibrionales and Xanthomonadales), where we find free-living and commensal species, intracellular symbionts and plant and animal pathogens, as well as species of medical and agricultural relevance. In one of the most basal group of gamma-proteobacteria, the order Xanthomonadales, resides many significant pathogenic bacteria affecting humans, animals and plants. This clade, though it contains a single family (Xanthomonadaceae), comprises a diverse set of species differing both in their phenotypic traits and habitat. The group has considerable economic impact on agriculture, worldwide, and more than 350 different plant diseases caused by xanthomonads have been reported (Leys et al., 1984; Swings and Civerolo, 1993). Xanthomonas and Xylella are genera within Xanthomonadales for which complete genome sequences are available for several species and strains (da Silva et al., 2002; Lee et al., 2005; Qian et al., 2005; Simpson et al., 2000; Thieme et al., 2005; Van Sluys et al., 2003). The genus Xanthomonas comprises a diverse group of Gram-negative, obligate aerobic, non-fermentative rods, in which all reported strains are plant-associated and most are reported as being pathogenic to a particular plant host (Van Sluys et al., 2002). Similarly, Xylella fastidiosa (Xf) is responsible for causing economically important diseases in grapevines, citrus fruits and many other plant species. Foremost among these are Pierce's disease (PD) of grapevines, citrus variegated chlorosis (CVC), and the leaf scorch diseases of almonds (ALS) and oleanders (OLS) (Davis et al., 1978; Purcell and Hopkins, 1996). Although the gamma subdivision is a well-defined clade, the classification of Xanthomonadales inside the group is quite problematic. The 16S rRNA phylogenetic tree, by far the most frequently used sequence to assess phylogenetic placement of a given bacterial group, places Xanthomonadales at the root of the gamma-proteobacteria (Lima et al., 2008). However, very often phylogenetic trees of highly conserved genes place this group in the same branch as the beta subdivision (Martins-Pinheiro et al., 2004) or even as an outgroup of the beta/gamma-proteobacteria (Schneider et al., 2007). Moreover, these studies are complicated by the high frequency of genes that are proposed to be acquired by LGT, especially within the Xanthomonas genus. Due to frequent LGT, Xanthomonadales represents an extreme case of genomic mosaicism that prevents it from being assigned to any one of the major proteobacterial clades (Comas et al., 2006; Lima et al., 2005, 2008). On considering the disagreements on the phylogenetic placement of Xanthomonadales, the use of other reliable, non-phylogenetic approaches may allow us to better position this group within the proteobacteria.

In this work, we were able to identify several specific INDELS, based on a sequence alignment of several highly conserved proteins, mostly those linked to DNA metabolism, that confirm the Xanthomonadales as a monophyletic, early-branching group within the gamma-proteobacteria subdivision.

2. Methodology

2.1. Phylogenetic analyses

Protein sequences were aligned by using Clustal X 2.0.9 program (Larkin et al., 2007), and regions of the alignments that were ambiguous, hypervariable or containing gaps were excluded from subsequent analysis (GeneDoc program, Nicholas et al., 1997). ProtTest was used to assess the best-fit amino acid substitution model for maximum likelihood-based tree reconstruction (Abascal et al., 2005). Maximum likelihood (ML) trees were set up with RAxML 7.0.4 (Stamatakis, 2006), with the WAG (Whelan–Goldman) model and parameters for invariable sites (+I) and gamma-distributed rate heterogeneity (+G, 4 categories). One-thousand bootstrap replicates were executed and bootstrap values drawn up on the best-scoring ML-tree. Trees were visualized using the TreeView program (Page, 1996), and were arbitrarily rooted at midpoint (although trees should be fundamentally viewed as unrooted). The concatenated tree was based on the alignment of the concatenated sequences of 11 evolutionarily conserved genes: arginyl-tRNA synthetase; LigA; DNA gyrase subunit A; Hsp70; isoleucyl-tRNA synthetase; DNA polymerase I; DNA polymerase III subunit alpha; DNA polymerase subunit epsilon; RecA; ribosomal L2 and S3 proteins. The proteins employed in phylogenetic analyses are shown in Table S01 (Supplementary material).

2.2. Identification of conserved signature sequences

DNA repair and replication-related protein sequences were obtained by similarity searches using the BlastP program (Altschul et al., 1997) as implemented in the NCBI server. The list of organisms used in this study, as well as the seed sequences used to perform the similarity searches, is shown in Table S02. Sequences were aligned using Clustal X (Thompson et al., 1997), and alignment parameters suggested by Hall (2005) were implemented (gap opening: 35.00 and gap extension: 0.75). The identification of potential INDELS was performed as proposed before (Gupta, 1998). Briefly, this methodology is based on the study of insertions and deletions present in a set of orthologous proteins. INDELS are detected as regions of defined length and sequence, and flanked by highly conserved regions (which excludes misleading INDEL identification due to improper alignment or sequencing errors) and found at precisely the same position in orthologs from different species (INDELS must be positional homologs). All sequence alignments are available upon request.

3. Results

Phylogenetic analyses of proteobacterial groups were conducted with the concatenated sequences of 11 conserved proteins, including the sequences employed in the search for conserved INDELS (Fig. 1). The use of concatenated sequences for tree generation gave more consistent trees and better bootstrap support than single genes. In this concatenated tree, the Xanthomonadales form a monophyletic group within the same clade as the beta-proteobacteria and at the base of the gamma-proteobacteria, close to other orders that are also difficult to correctly classify, such as the Legionelales, Chromatiales, Methylococcales, Thiotrichales and Vibrionales (Fig. 1). Phylogenetic trees for individual sequences (including 16S rRNA, RecA, TopA and all the 11 genes employed in the concatenated tree; see MM for the list) were also generated. As a general trend, proteobacterial groups are clearly

distinguished (betas and gammas always forming a monophyletic group with high bootstrap support; data not shown). Nevertheless, depending on the sequence analyzed, the gamma-proteobacteria may or may not form a monophyletic group (as in the case of TopA, where the group is split into at least two, Fig. 1, in Supplementary material) and the branching order for betas, gammas and Xanthomonadales, as seen in the topology obtained for the concatenated tree, is supported in most of the cases, but not always (7 out of 11 single gene trees have the same topology as the concatenated tree; data not shown).

These results confirm that the evolutionary history of these bacteria is quite complex and underline the need to use novel tools to investigate the taxonomic position of Xanthomonadales. Therefore, we sought additional protein signature sequences, the INDELs. The assumption behind this approach is that when a conserved INDEL (conserved in terms of length, sequence and residual position, and

Fig. 1. Maximum likelihood tree of the proteobacteria based on the concatenated sequences of 11 conserved proteins. Branch lengths are proportional to the number of amino acid substitutions per site. Numbers at the nodes represent the percentage of bootstrap support (only values > 50% are shown). See Table S01 for the accession codes, and Table S02 for abbreviations of species names.
flanked by a region sharing high levels of sequence similarity over all organisms analyzed) is found in an alignment, the simplest and most parsimonious explanation for its presence is that it was introduced only once during the course of evolution into a hypothetical ancestral organism and then kept in all its progeny. Thus, INDELs shared by two organisms indicate a common ancestry. To assess the phylogenetic relationship of the Xanthomonadales group within the proteobacteria, we selected evolutionarily conserved proteins with a role in DNA metabolism processes. The proteins selected participate in informational processes and their function is dependent upon interaction with other proteins, thereby minimizing the chance of LGT (Jain et al., 1999). Table 1 summarizes the major INDELs that characterize distinct proteobacterial groups, especially the Xanthomonadales. Each protein containing an INDEL used in this study is further described below.

3.1. Gamma-proteobacteria specific signature sequences

3.1.1. DNA polymerase III, alpha subunit

The DNA polymerase III holoenzyme is a complex that contains 10 different types of subunits. The alpha subunit, encoded by the polC gene, is a 130-kDa polypeptide endowed with 5 to 3 DNA polymerase activity but none for proofreading (Patel and Loeb, 2001; Steitz, 1999). A signature sequence gene that consists of a 5-amino acid insert is shared by homologs from gamma-proteobacteria, but is absent from Xanthomonadales, Methylococcales, Legionellales, Cardiobacteriales, part of the Thiotrichales (Francisella novicida), and from other proteobacterial subdivisions (Fig. 2). As with many other INDELs, this separates the Xanthomonadales (and other groups) from the core group of gamma-proteobacteria.

### Table 1

<table>
<thead>
<tr>
<th>Protein Description</th>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA polymerase III alpha subunit</td>
<td>5-amino acid length insertion (Fig. 2)</td>
<td>Enterobacteriales, Vibrionales, Pasteurellales, Pseudomonadales, Alteromonadales, Aeromonadales, Oceanospirillales, Chromatiales orders and Thiomicrospira crunogena (from Thiotrichales group)</td>
</tr>
<tr>
<td>DNA Topoisomerase I</td>
<td>29-41-amino acid length insertion (Fig. 25)</td>
<td>Enterobacteriales, Vibrionales, Pasteurellales, Pseudomonadales, Alteromonadales, Aeromonadales, Oceanospirillales and Pseudomonadales</td>
</tr>
<tr>
<td>DNA polymerase I</td>
<td>29-41-amino acid length insertion (Fig. 25)</td>
<td>Enterobacteriales, Vibrionales, Pasteurellales, Pseudomonadales, Alteromonadales, Aeromonadales, Oceanospirillales and Pseudomonadales</td>
</tr>
<tr>
<td>DNA Topoisomerase I 4–5-amino acid insertion (Fig. 3)</td>
<td>7-amino acid length insertion (Fig. 35)</td>
<td>Enterobacteriales, Vibrionales, Pasteurellales, Pseudomonadales, Alteromonadales, Aeromonadales, Oceanospirillales and Pseudomonadales orders</td>
</tr>
<tr>
<td>DNA polymerase III epsilon subunit</td>
<td>1-amino acid length insertion (Fig. 35)</td>
<td>Pseudomonadales family (representing by Pseudomonas and Azotobacter)</td>
</tr>
<tr>
<td>DNA Topoisomerase I 4–5-amino acid length insertion (Fig. 45)</td>
<td>57–65-amino acid length insertion (Fig. 4)</td>
<td>Enterobacteriales, Pasteurellales, Vibrionales, Aeromonadales, Cardiobacteriales and members from Shewanellaceae family (Alteromonadales)</td>
</tr>
<tr>
<td>MutS</td>
<td>5-amino acid length insertion (Fig. 45)</td>
<td>Enterobacteriales, Pasturellellales, Vibrionales, Aeromonadales, Cardiobacteriales and members from Shewanellaceae family (Alteromonadales)</td>
</tr>
<tr>
<td>RecA</td>
<td>2-amino acid length insertion (Fig. 55)</td>
<td>Enterobacteriales, Pasturellellales, Vibrionales, Aeromonadales, Cardiobacteriales and members from Shewanellaceae family (Alteromonadales)</td>
</tr>
<tr>
<td>DNA Topoisomerase III alpha subunit</td>
<td>4-amino acid length insertion (Fig. 75)</td>
<td>Xanthomonadales</td>
</tr>
<tr>
<td>DNA ligase NAD-dependent</td>
<td>57–65-amino acid length insertion (Fig. 4)</td>
<td>Xanthomonadales</td>
</tr>
<tr>
<td>MutS</td>
<td>5-amino acid length insertion (Fig. 45)</td>
<td>Xanthomonadales</td>
</tr>
<tr>
<td>RecA</td>
<td>2-amino acid length insertion (Fig. 55)</td>
<td>Xanthomonadales</td>
</tr>
<tr>
<td>DNA polymerase III alpha subunit</td>
<td>4-amino acid length insertion (Fig. 75)</td>
<td>Xanthomonadales</td>
</tr>
</tbody>
</table>

Another 4–5-amino acid INDEL in this protein is absent in most of the gamma-proteobacteria, but present in Xanthomonadales, Legionellales, Methyllococcales, Chromatiales, Cardiobacteriales, Thiotrichales (Francisella novicida), as well as proteobacteria of the beta and alpha subdivisions, and a small number of delta-proteobacteria (Fig. 3). Curiously, this INDEL is absent from most of the delta subdivision and is completely absent from the epsilon subdivision. This heterogeneity suggests that delta-proteobacteria may not be a monophyletic group, as part of it is more closely related to alpha-proteobacteria, as already previously suggested by Gupta (2005). This insertion is also absent outside of the proteobacteria clade (represented here by some firmicutes and chlamydialae, Fig. 3). Interestingly, these inferences from signature sequence are supported by phylogenetic analyses based on Topoisomerase I (Fig. 1S): the beta-proteobacteria, most of the deep branching gamma-proteobacteria groups, and Geobacter branch together and independently from the other proteobacteria. Thus, the
Fig. 2. Partial sequence alignment of the DNA polymerase III epsilon subunit showing a 5-amino acid insert specific for the core gamma-proteobacteria groups, but absent in Xanthomonadales, several deep branching bacteria of this class, and all other bacteria. Dots (.) in this and other alignments denote identity with the amino acid on the top line and gaps are indicated by empty spaces. Accession numbers are given after each species name. See Table S02 for abbreviations of species names. The numbers on the left indicate the amino acid position for each protein sequence. Sequence information for only representative species is shown, but the sequences from other available species from these groups behave similarly.
The evolutionary history of this gene is not typical and indicates an early divergence of the two groups (one of them with the insertion). The possibility that this gene underwent an ancient LGT event cannot be excluded.

The atypical grouping for Thiotrichales (represented here by Francisella novicida and Thiomicrospira crunogena, which are discriminated here by both signatures) is in agreement with earlier work by the group of Gupta, who found the two genera sepa-

Fig. 3. Partial sequence alignment of Topoisomerase I, showing a 4–5-amino acid insert is absent from the core gamma-proteobacteria, but present in Xanthomonadales and most of the other deep branching bacteria of this class, as well as in the beta and alpha-proteobacteria, but only a few of the delta-proteobacteria. For most of the delta-proteobacteria and also other bacteria (the Firmicutes/Chlamydiae group are presented as outgroups), the insert is absent. Details as described in Fig. 2.
rated in protein signature and phylogenetic analyses (Gao et al., 2009).

3.1.3. DNA polymerase I

In addition to 5′ to 3′ DNA polymerase activity, DNA polymerase I exhibits 3′ to 5′ exonuclease activity, which mediates proofreading, and 5′ to 3′ nick translation during DNA repair (Patel et al., 2001). A 7–8-amino acid insertion is present in homologs from the core gamma-proteobacteria, but is not found in homologs from Xanthomonadales, Chromatiales, Methyllocccales, Legionellales, Cardiobacteriales and Thiotrichales, neither in homologs from the beta, alpha, delta and epsilon subdivisions (Fig. 3S, Supplementary material).

Moreover, a small 1-amino acid insertion in the Pseudomonadales order is also observed. This INDEL is present as well in the Oceanospirillales Chromohalobacter salexigens, thus indicating a close relationship with the Pseudomonadales, as already demonstrated in previous studies (Quillaguaman et al., 2004). The absence of both INDELS in two Acinetobacter species (belonging to the Pseudomonadales group) points to an ancient divergence of these bacteria within the Pseudomonadales.

3.1.4. DNA polymerase III, epsilon subunit

The DNA polymerase III epsilon subunit protein, encoded by the dnaQ gene, is involved in DNA replication through its proofreading 3′ to 5′ exonuclease activity (Patel and Loeb, 2001; Steitz, 1999). A signature sequence consisting of a 4–5-amino acid insert is present in dnaQ homologs of the gamma-proteobacteria, but absent in homologs from Xanthomonadales, Methyllocccales, Legionellales, Chromatiales, Thiotrichales, Oceanospirillales, Pseudomonadales and Alteromonadales, as well as in the other proteobacterial groups (Fig. 4S, Supplementary material). This indicates a late origin for these bacteria within the Pseudomonadales.

3.2. Xanthomonadales specific signature sequences

3.2.1. NAD-dependent DNA ligase LigA

Two distinct families of DNA ligase are found, one that is typical of Bacteria and uses NAD+ as a cofactor, and the other that uses ATP and is common to Archaea and Eukaryotes; both families of DNA ligase, however, are present in Xanthomonas (Wilkinson et al., 2001; Martins-Pinheiro et al., 2004). The NAD-dependent DNA ligase bears a distinctive 65-amino acid insertion, only found in the Xanthomonadales group (Fig. 4). Recursive BLAST searches using this striking long INDEL as a seed sequence were performed, but only sequences from the Xanthomonadales group showed similarity.

3.2.2. Mismatch repair protein MutS

The MutS protein is involved in the repair of mismatches in DNA. This protein forms a dimer (MutS2) that recognizes the mismatched base on the daughter-strand and binds the mutated DNA with feeble ATPase activity (Kunkel and Erie, 2005). This protein reveals a conserved 5-amino acid insertion, exclusive to members of the Xanthomonadales group (Fig. 5S, Supplementary material).

3.2.3. RecA

RecA plays an essential role in at least three distinct cellular processes: homologous recombination, gene control of the SOS regulon and DNA damage induced mutagenesis (Cox, 2003; Eisen, 1995). Moreover, this protein has been widely used for phylogenetic reconstruction in order to assess relationships within Bacteria (Eisen, 1995; Martins-Pinheiro et al., 2004). A 2-amino acid insertion in a highly conserved region has been shown to distinguish the Xanthomonadales order (Fig. 6S, Supplementary material).

Other interesting signatures were also observed, such as a 4-amino acid insertion exclusive to Xanthomonadales in the alpha subunit of DNA polymerase III (Fig. 7S, Supplementary material), a 4-amino acid insertion in valyl-tRNA synthetase exclusive to Xanthomonadales and a 37-amino acid insertion, also observed in valyl-tRNA synthetase, present in all gamma and beta-proteobacteria, but absent in the remaining groups, previously described (Gupta, 2000).

4. Discussion

An accurate classification of xanthomonads is not only important for the basic knowledge of the phylogeny of this group, but also has practical implications for global regulation of severe diseases caused by these plant pathogens. Misidentification and/or unclear classification of xanthomonads may result in unnecessary quarantine regulations and economic loss. In previous studies using different gene and protein sequences, the classification of Xanthomonadales was contradictory to the placement of these bacteria as gamma-proteobacteria. For example, when analyzing the phylogeny of the RecA protein (Martins-Pinheiro et al., 2004) or 16S rRNA (Lima et al., 2005, 2008), independent results placed the Xanthomonadales in a branch close to the beta subdivision (as also shown herein, Fig. 1). Different genes and different methods may sometimes result in contradicting trees. These problems are typical in molecular phylogeny studies, even in those cases in which LGT can be ruled out. Other authors have previously highlighted the incongruity between gene and presumed organism phylogeny (Poptsova and Gogarten, 2007).

Comas et al. (2006) analyzed whether incongruities apparent in protein and/or gene trees from Xanthomonadales could have resulted from phylogenetic noise and/or LGT. Phylogenetic noise affects basal groups whose positions may have changed due to limitations in phylogenetic reconstruction methods. They concluded that extensive transfer among Xanthomonadales and alpha, beta and even gamma-proteobacteria is the main source of incongruity in the proteobacterial tree. The results of Lima et al. (2008) corroborate this hypothesis through the identification of genome islands as products of early and late LGT events, thus confirming the mosaic structure of Xanthomonas genomes. Phylogenetic trees of proteins not normally subject to LGT may help to solve this problem. However, as observed for 16S rRNA, RecA and other phylogenetic reconstructions (Figs. 1 and 1S), the Xanthomonadales group appears as paraphyletic with respect to the beta-proteobacteria, branching at the base of the gamma-proteobacteria clade.

The identification of specific signature sequences is a reliable and consistent approach for determining relative branching and taxonomic relationships, and has been used with success to determine both broad and specific evolutionary relationships (Gupta, 1997, 1998, 2000; Gupta and Griffiths, 2002). The main idea of these studies is based on the assumption that these signatures, mainly INDELS, are unique events within certain genes, and provide powerful tools for inferring evolutionary relationships. However, some of these INDELS do present some variation, which appears as several small gaps within the insertion presented in Fig. 2S. Although most of these gaps are clearly specific to certain clades, their origin indicates that the region of the protein affected by the INDEL has some structural flexibility due to low selective pressures.
In the present work, the search for INDELs was applied mainly to uncover taxonomic relationships of the Xanthomonadales order within the proteobacteria, by using a set of conserved proteins related to core metabolic functions such as DNA repair and replication. The main results of these analyses are summarized in a schematic genetic evolutionary flow of beta and gamma-proteobacteria, as shown in Fig. 5.

Interestingly, most of the described INDELs distinguish the gamma-proteobacteria group from other proteobacteria groups, as these INDELs are absent in beta, alpha or delta/epsilon subdivisions; however, the same profile of absence is shared by early divergent groups inside the gamma-proteobacteria, notably Xanthomonadales, Chromatiales, Methylococcales, Legionellales, Cardiobacteriales and Thiotrichales. This strongly suggests a close relationship among these groups, in agreement with the phylogenetic trees shown in Figs. 1 and 1S. In fact, these INDELs were most likely to have been introduced after these groups branched from the rest of the gamma-proteobacteria. The profile of presence in gamma-proteobacterial species and absence in other proteobacterial groups indicates that these signatures constitute a synapomorphy for this subdivision.

A detailed phylogenomic analysis of the gamma-proteobacteria was recently published (Gao et al., 2009). Basically, a phylogenetic tree based on concatenated protein sequences from conserved genes placed Xanthomonadales as a deep branch (as well as Thiotrichales, Cardiobacteriales, Legionellales, Chromatiales and Methylococcales) in the phylogenies of gamma-proteobacteria. Although the beta-proteobacteria were not included in those analyses, a specific 2-amino acid INDEL, in AIACR-transformylase (PurH), could differentiate all the other proteobacteria (including beta-proteobacteria) from Xanthomonadales and gamma-proteobacteria. This INDEL was included in the schematic genetic evolutionary flow shown in Fig. 5. Moreover, four conserved proteins, of unknown functions, were identified as exclusive to the gamma-proteobacteria, including Xanthomonadales. These results are good indications that Xanthomonadales share important features with the gamma-proteobacteria, but similar searches for genes shared exclusively between Xanthomonadales and beta-proteobacteria would be important to understand the relationships among these groups. In fact, the INDEL found for topoisomerase I (Fig. 3) and its interesting phylogeny (Fig. 1S) place the Xanthomonadales (as well as Legionellales, Chromatiales, Methylococcales, Cardiobacteriales and Thiotrichales) as a deep branch in the phylogenies of gamma-proteobacteria.

[Fig. 4. Partial sequence alignment of the DNA ligase NAD-dependent protein, showing a 65-amino acid insert specific for Xanthomonadales. Details as described in Fig. 2.]
riales and some of the Thiotrichales, Francisella genus) in the same branch of beta-proteobacteria, but, in this case, independent of gamma-proteobacteria. This can be explained by a duplication of this gene, followed by loss of one of the paralogues. The evolution of the two paralogues place the Xanthomonadales as an outgroup from the core gamma-proteobacteria. An alternative explanation is that this insert was introduced in a common ancestor of the various alpha (and a few delta), beta and gamma-proteobacteria, followed by a second genetic event leading to loss of this insert from a common ancestor of the Enterobacteriales, Pasteurellales, Vibrionales, Alteromonadales Aeromonadales, Oceanospirillales, Pseudomonadales and some from Thiotrichales group (Thiomicrospira genus).

On the other hand, the INDELs exclusively shared by homologs from Xanthomonadales, and absent in any homolog from other proteobacteria, are suggestive of the independent evolution of this group and provide molecular signatures to distinguish and define it. These signatures are likely recent molecular events in the Xanthomonadales, which occurred after its divergence from the main branches of gamma- and beta-proteobacteria. In view of the highly conserved and well-defined nature of these signatures, they clearly define the Xanthomonadales as monophyletic. Combining the data of INDEL and phylogeny, these bacteria are more likely to have evolved independently soon after, or concomitant to, the divergence of the beta and gamma-proteobacteria.

5. Conclusions

The alignment of conserved protein sequences has led to the identification of conserved INDELs, useful for assessing phylogenetic placement of the Xanthomonadales group. The data presented here delineate the evolution of the gamma lineage and provide further evidence for the early gamma placement, indicating that Xanthomonadales constitute one of the deeper branching lineage within gamma-proteobacteria. Moreover, the results confirm the monophyly of Xanthomonadales and constitute interesting genetic markers to identify these bacteria.

Acknowledgments

The authors thank Dr. Radhey S. Gupta (McMaster University, Ontario, Canada) for helpful discussion. The authors are also in debt to the anonymous reviewers for their comments, criticisms and excellent suggestions. Financial support was obtained from FAPESP (São Paulo, Brazil) and CNPq (Brasília, Brazil). During this work, A.M.C.J. received a fellowship from CAPES (Brasília, DF, Brazil), and M.M.P. and W.C.L. from FAPESP.

Appendix A. Supplementary data

ML phylogenetic tree of topoisomerase I (Fig. 1S) and partial protein alignments (Figs. 2S–7S). List of proteins numbers used in phylogenetic analyses (Table S01) and of organisms used to perform similarity searches (Table S02). Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2009.09.026.
References


Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of Xylella fastidiosa. J. Bacteriol. 185, 1018–1026.