Molecular Phylogenetics and Evolution 54 (2010) 524–534

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Evolutionary placement of Xanthomonadales based on conserved protein signature sequences

Ania M. Cutiño-Jiménez [a], Marinalva Martins-Pinheiro [b], Wanessa C. Lima [b,1], Alexander Martín-Tornet [c], Osleidys 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 INFO

Article history:
Received 14 May 2009
Revised 11 September 2009
Accepted 21 September 2009
Available online 26 September 2009

Keywords:
Xanthomonadales
Xanthomonas
Xylella
DNA repair
INDELs

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 Legionelalles, Chromatiales, 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).

* Corresponding author. Address: Ave. Prof. Lineu Prestes, 1374, Institute of Biomedical Sciences, ICB II/USP, Cidade Universitária, São Paulo, CEP 05508-900, Brazil. Fax: +55 11 3091 7357.
E-mail address: cfmmenck@usp.br (C.F.M. Menck).
[1] Present address: Department Cell Physiology and Metabolism, University of Geneva, R. Michel Servet 1, 1211, Geneva, CH, Switzerland.

1055-7903/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2009.09.026

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 one 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, Cardiobacteriales, 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 (Leyns 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 Pierces'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 Legionelalles, Chromatiales, Methylococcales, Thiotrichales and Cardiobacteriales (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