International Journal of Systematic and Evolutionary Microbiology (2016), **66**, 5575–5599 **DOI** 10.1099/ijsem.0.001485

Genome-based phylogeny and taxonomy of the '*Enterobacteriales*': proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae*, *Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov.

Mobolaji Adeolu,† Seema Alnajar,† Sohail Naushad and Radhey S. Gupta

Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada

Correspondence
Radhey S. Gupta
gupta@mcmaster.ca

Understanding of the phylogeny and interrelationships of the genera within the order '*Enterobacteriales*' has proven difficult using the 16S rRNA gene and other single-gene or limited multi-gene approaches. In this work, we have completed comprehensive comparative genomic analyses of the members of the order '*Enterobacteriales*' which includes phylogenetic reconstructions based on 1548 core proteins, 53 ribosomal proteins and four multilocus sequence analysis proteins, as well as examining the overall genome similarity amongst the members of this order. The results of these analyses all support the existence of seven distinct monophyletic groups of genera within the order '*Enterobacteriales*'. In parallel, our analyses of protein sequences from the '*Enterobacteriales*' genomes have identified numerous molecular characteristics in the forms of conserved signature insertions/deletions, which are specifically shared by the members of the identified clades and independently support their monophyly and distinctness. Many of these groupings, either in part or in whole, have been recognized in previous evolutionary studies, but have not been consistently resolved as monophyletic entities in 16S rRNA gene trees. The work presented here represents the first comprehensive, genome-scale taxonomic analysis of the entirety of the order '*Enterobacteriales*'. On the basis of phylogenetic analyses and the numerous identified conserved molecular characteristics, which clearly distinguish members of the order '*Enterobacteriales*' and the seven reported clades within this order, a proposal is made here for the order *Enterobacterales* ord. nov. which consists of seven families: *Enterobacteriaceae, Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov.

INTRODUCTION

The order '*Enterobacteriales*' is a large and diverse group of Gram-negative, facultatively anaerobic, non-spore-forming, rod-shaped bacteria within the class *Gammaproteobacteria*. Members of this group inhabit a number of different ecological niches and have been found in soil, water and in association with living organisms including plants, insects, animals and humans (Brenner & Farmer, 2005). Many members of the order '*Enterobacteriales*' have been implicated as pathogens in humans and animals, such as the species *Escherichia coli, Salmonella enterica*, and *Yersinia pestis*, and as economically devastating phytopathogens, such as members of the genera *Dickeya, Pectobacterium, Brenneria, Erwinia* and *Pantoea* (Bonn & van der Zwet, 2000; Coutinho & Venter, 2009; Croxen & Finlay, 2010; Hauben *et al.*, 1998; Livermore, 2012; Tyler & Triplett, 2008). At the time of writing, the order

†These authors contributed equally to this work.

Abbreviations: CSI, conserved signature insertion/deletion; MLSA, multilocus sequence analysis.

Seventy-five supplementary figures and five supplementary tables are available with the online Supplementary Material.

'E*nterobacteriales*' contains 60 genera with validly published names (www.namesforlife.com; Parte, 2014) including the recently described genus *Chania* (Ee *et al.*, 2016) and an additional genus which has been recently described but the name is not yet validly published ['*Atlantibacter*' (Hata *et al.*, 2016)]. Most genera within the order 'E*nterobacteriales*', encompassing over 250 species, are placed within the sole family with a validly published name within the order, *Enterobacteriaceae*; making the family *Enterobacteriaceae* one of the most taxonomically diverse bacterial families currently recognized (www.namesforlife.com; Parte, 2014). A number of distinct groupings of genera within the family *Enterobacteriaceae* are well known (viz. the groupings of the genera *Salmonella, Citrobacter* and *Escherichia/Shigella*, and the genera *Dickeya, Pectobacterium* and *Brenneria*, the close associations between the genera *Xenorhabdus* and *Photorhabdus*, the genera *Erwinia* and *Pantoea*, and the genera *Obesumbacterium* and *Hafnia*) (Goodrich-Blair & Clarke, 2007; Naushad *et al.*, 2014; Octavia & Lan, 2014; Samuel *et al.*, 2004; Zhang & Qiu, 2015; Zhang *et al.*, 2016), but these groupings are not recognized as unique taxonomic units.

The biochemical diversity and the large number of organisms within the order '*Enterobacteriales*' has made biochemical descriptions of the order and its constituent subgroups difficult (Brenner & Farmer III, 2005; Octavia & Lan, 2014). Our current understanding of the phylogeny and interrelationships of the members of the order '*Enterobacteriales*' is primarily based on the 16S rRNA gene (Francino *et al.*, 2006; Hauben *et al.*, 1998; Naum *et al.*, 2008; Spröer *et al.*, 1999). However, the 16S rRNA gene has low discriminatory power and interrelationships of the members of the order '*Enterobacteriales*' are poorly resolved in 16S rRNA-gene-based phylogenetic trees (Hauben *et al.*, 1998; Naum *et al.*, 2008; Octavia & Lan, 2014). Additionally, the branching of the genera and species within '*Enterobacteriales*' in 16S rRNA-gene-based phylogenies shows considerable stochasticity depending on the algorithms used and the organisms analysed (Naum *et al.*, 2008; Octavia & Lan, 2014). Most concerning, comprehensive 16S rRNA gene phylogenetic trees for the order '*Enterobacteriales*' and other members of the class *Gammaproteobacteria* suggest that the order '*Enterobacteriales*' exhibits polyphyletic branching and does not form a coherent monophyletic grouping (Brenner & Farmer, 2005; Octavia & Lan, 2014; Yarza *et al.*, 2008; Yilmaz *et al.*, 2013). A number of alternative genes have been employed in phylogenetic analysis of the order '*Enterobacteriales*' in order to gain additional insight into the interrelationships of the members of the order, such as *gyrB* (Dauga, 2002; Fukushima *et al.*, 2002), *dnaJ* (Pham *et al.*, 2007), *oriC* (Roggenkamp, 2007) and *recA* (Tailliez *et al.*, 2010). More recently, multiple gene/protein-based multilocus sequence analysis (MLSA) studies have been conducted to further elucidate the phylogeny of the order '*Enterobacteriales*' including studies based on the genes *tuf* and *atpD* (Paradis *et al.*, 2005), the genes *atpD, carA* and *recA* (Young & Park, 2007), the genes *gapA, gyrA* and *ompA* (Naum *et al.*, 2011), the genes *rpoB, gyrB, dnaJ* and *recA* (Hata *et al.*, 2016), the genes *fusA, pyrG, rplB, rpoB* and *sucA* (Ee *et al.*, 2016), and, most commonly, the genes *gyrB*, *rpoB*, *atpD* and *infB* (Brady *et al.*, 2008, 2013, 2014b; Glaeser & Kämpfer, 2015; Zhang & Qiu, 2015). These studies have led to a significant number of reclassifications within the order '*Enterobacteriales*' and have alleviated many of the issues related to polyphyletic genera within the order. However, no family-level divisions within the order '*Enterobacteriales*' have thus far been proposed.

The increasing prevalence and ubiquity of genome sequencing technology has led to an increasing wealth of publically available genome sequence data. Currently, there are over 14 000 genomes from 54 genera with valildy published names within the order '*Enterobacteriales*' available in the NCBI genome database (http://www.ncbi.nlm.nih.gov/genome). These genome sequences are enabling the increasing use of robust and reliable core genome-based phylogenetic reconstructions in '*Enterobacteriales*' research (Husník *et al.*, 2011; Wattam *et al.*, 2014; Zhang & Qiu, 2015; Zhang *et al.*, 2016), which have been shown to mitigate the effects of recombination or lateral gene transfer and provide greater resolving power than phylogenetic trees based on single genes/proteins (Ciccarelli *et al.*, 2006; Gao *et al.*, 2009; Rokas *et al.*, 2003; Wu *et al.*, 2009). Genome sequence data is also enabling the detection of conserved molecular characteristics shared by evolutionarily related groups of organisms. One particular class of conserved molecular characteristics, which have recently been utilized to great effect in prokaryotic taxonomy are conserved signature insertions/deletions (CSIs) present in widely distributed proteins (Gupta, 2014, 2016; Naushad *et al.*, 2014). CSIs are insertions or deletions (indels) that are uniquely present in a related group of organisms. The most parsimonious explanation of the presence of the CSI in a related group of organisms is the existence of a common ancestor in which the genetic change leading to the CSI occurred, and which was subsequently inherited by all of its various desce dents. Thus, CSIs represent synapomorphic characteristics and they provide reliable evidence, independent of phylogenetic trees, that the species from the groups in which they are found are specifically related to each other due to common ancestry. Recently, on the basis of CSIs and other molecular characteristics, the taxonomy of a number of important prokaryotic groups, ranging from genus to phylum level taxa, has been revised (Campbell *et al.*, 2015; Gupta, 2016; Gupta *et al.*, 2015a, b, 2016; Naushad *et al.*, 2014, 2015b; Sawana *et al.*, 2014).

In our earlier work, a limited number of CSIs and unique proteins, referred to as conserved signature proteins, were identified that were distinctive characteristics of either all *Gammaproteobacteria* or were commonly shared by members from certain orders of *Gammaproteobacteria* which reliably grouped together in phylogenetic trees reconstructed in this work (Gao *et al.*, 2009; Gupta, 2000). We have also previously completed comprehensive studies in order to identify large numbers of CSIs utilized to reclassify members within the gammaproteobacterial orders *Pasteurellales* and *Xanthomonadales* (Naushad & Gupta, 2012, 2013; Naushad *et al.*, 2015a, b). In the present study, we

have extended our earlier work on *Gammaproteobacteria* by carrying out comprehensive phylogenetic and comparative genomic studies on members of the order '*Enterobacteriales*' to examine their evolutionary relationships and taxonomy. Using whole genome sequences of 179 representative genome sequenced members of the order '*Enterobacteriales*', we have reconstructed a highly robust phylogenetic tree based on 1548 shared core proteins, as well as phylogenetic trees based on 53 ribosomal proteins and four MLSA proteins, and to identify conserved molecular characteristics that can be used to determine the interrelationships within the order '*Enterobacteriales*'. Here we present five CSIs which are unique characteristics of all '*Enterobacteriales*' and an additional 66 CSIs which are specific for seven main groups of genera within the order '*Enterobacteriales*' identified in our phylogenetic trees. The 71 CSIs identified in this work, when combined with previously discovered CSIs (Naushad *et al.*, 2014) and the highly robust phylogenetic trees reconstructed here, provide for a comprehensive understanding of interrelationships within the order '*Enterobacteriales*' and form the basis for a novel taxonomic framework. On the basis of the phylogenetic analyses and the identified conserved molecular characteristics presented here, we propose a division of the order '*Enterobacteriales*' into seven novel families.

METHODS

Phylogenetic and genomic analyses of the order '*Enterobacteriales*'. Three phylogenetic trees were produced in this work utilizing 179 representative genome-sequenced members of the order '*Enterobacteriales*' (Table S1, available in the online Supplementary Material) and four members of the families *Pasteurellaceae* and *Vibrionaceae* as outgroups. Representative genomes for the genus *Plesiomonas* and the endosymbiotic genera *Buchnera* and *Wigglesworthia* were not included in the phylogenetic trees shown in the main figures due to the potential for phylogenetic artifacts caused by long branch attraction effects (Bergsten, 2005; Philippe *et al.*, 2005), but they are shown in the respective supplemental figures for each phylogenetic tree. A core genome phylogeny was produced based on the concatenated sequences of 1548 core proteins. The core protein families used in the core genome phylogeny were identified using the UCLUST algorithm (Edgar, 2010) to identify protein families which shared at least 50 % sequence identity and 50 % sequence length. The 1548 identified protein families which were present in at least 80 % of the input genomes were used in the phylogenetic analysis. The 53 ribosomal proteins were identified using HMMer 3.1 (Eddy, 2011) based on profile hidden Markov models (Table S2) obtained from the Pfam database (Finn *et al.*, 2016). The four MLSA proteins (viz. GyrB, RpoB, AtpD and InfB) were identified using HMMer 3.1 (Eddy, 2011) based on amino acid sequences from *Escherichia coli* K12 (Blattner *et al.*, 1997) (Table S2) obtained from the UniProt database (UniProt Consortium, 2015). In each case, each identified protein family was individually aligned using Clustal Omega (Sievers *et al.*, 2011), trimmed using Gblocks 0.91 b (Castresana, 2000) with relaxed parameters (Talavera & Castresana, 2007), and concatenated with the other proteins in its dataset. The concatenated alignments were 458 971, 5930 and 3535 aligned amino acids long for the core protein, ribosomal protein, and MLSA protein datasets, respectively. Maximum-likelihood trees based on these concatenated alignments were reconstructed using FastTree 2 (Price *et al.*, 2010) employing the Whelan and Goldman model of protein sequence evolution (Whelan & Goldman, 2001) and RAxML 8 (Stamatakis, 2014) using the Le and Gascuel model of protein sequence evolution (Le & Gascuel, 2008). SH-like statistical support values (Guindon *et al.*, 2010) for each branch node in the final phylogenetic trees were calculated using RAxML 8 (Stamatakis, 2014). The resultant phylogenetic trees were drawn using MEGA 6 software package (Tamura *et al.*, 2013). This process was completed using an internally developed software pipeline. A manuscript for this pipeline is currently in preparation and the pipeline will be available for public use on Gleans.net once released. We have also utilized the protein families identified by the USearch algorithm (Edgar, 2010) for our core- protein-based phylogenetic tree to calculate the proportion of shared protein families in each pair of genomes in our dataset.

Identification of conserved signature indels. Conserved signature indels were identified as detailed by Gupta (2014) using protein sequences found in the genomes of *Shimwellia Blattae* DSM 4481^T (Brzuszkiewicz *et al.*, 2012), *Providencia stuartii* MRSN 2154 (Clifford *et al.*, 2012), *Pragia fontium* 24613 (Snopková *et al.*, 2015) and *Dickeya zeae* Ech586 (Pritchard *et al.*, 2013) as the starting points. BLAST (Altschul *et al.*, 1997) searches were conducted on each of the protein sequences in these genomes that were >75 amino acids in length against the NCBI non-redundant database. From the results of the BLAST searches, 15–20 homologues belonging different genera of '*Enterobacteriales*' and 6–8 species from other orders/classes of proteobacteria were selected. The selected sequences were aligned using CLUSTAL X 2.1 (Jeanmougin *et al.*, 1998). The alignments were then visually inspected for the presence of insertions or deletions that were flanked on both sides by at least 5–6 conserved amino acid residues in the neighbouring 30–40 amino acids. Gaps that were of a variable length or that were not flanked by conserved residues were not investigated further. Detailed BLAST searches were then carried out on short sequence segments containing the indel and the flanking conserved regions (60–100 amino acids long) and compared against the top 500 BLAST hits to determine the specificity of the indels. In some cases, an additional BLAST search was conducted to include a more diverse representation of the '*Enterobacteriales*' species involving 1000 alignments, or excluding overrepresented species. SIG_CREATE and SIG_STYLE (available on Gleans.net) were then used to create Signature files for identified CSIs that were specific to the order '*Enterobacteriales*' or one of its subgroups as described by Gupta (2014). Due to the large number of genome sequences available for the order '*Enterobacteriales*', the sequence alignment files presented here contain sequence information for only a limited number of species. However, unless otherwise indicated, homologues of all members of the specified groups displayed similar sequence characteristics.

RESULTS

Phylogenetic and genomic analyses of the order '*Enterobacteriales*'

Phylogenetic analyses of the order '*Enterobacteriales*'. In this work, we have produced three phylogenetic trees for 179 representative members of the order '*Enterobacteriales*', encompassing 49 genera with validly published names within the order: one tree based on 1548 core proteins, another based on 53 ribosomal proteins, and a third based on four MLSA proteins (Figs 1a–c and S1–S3). The 1548 core-protein-based phylogeny produced for this work, covering a majority of the diversity present within the order, represents one of the most comprehensive genome-based phylogenetic trees for the order '*Enterobacteriales*' produced to date. Additionally, a 16S rRNA gene-based phylogenetic tree of the '*Enterobacteriales*', produced as part of the All-

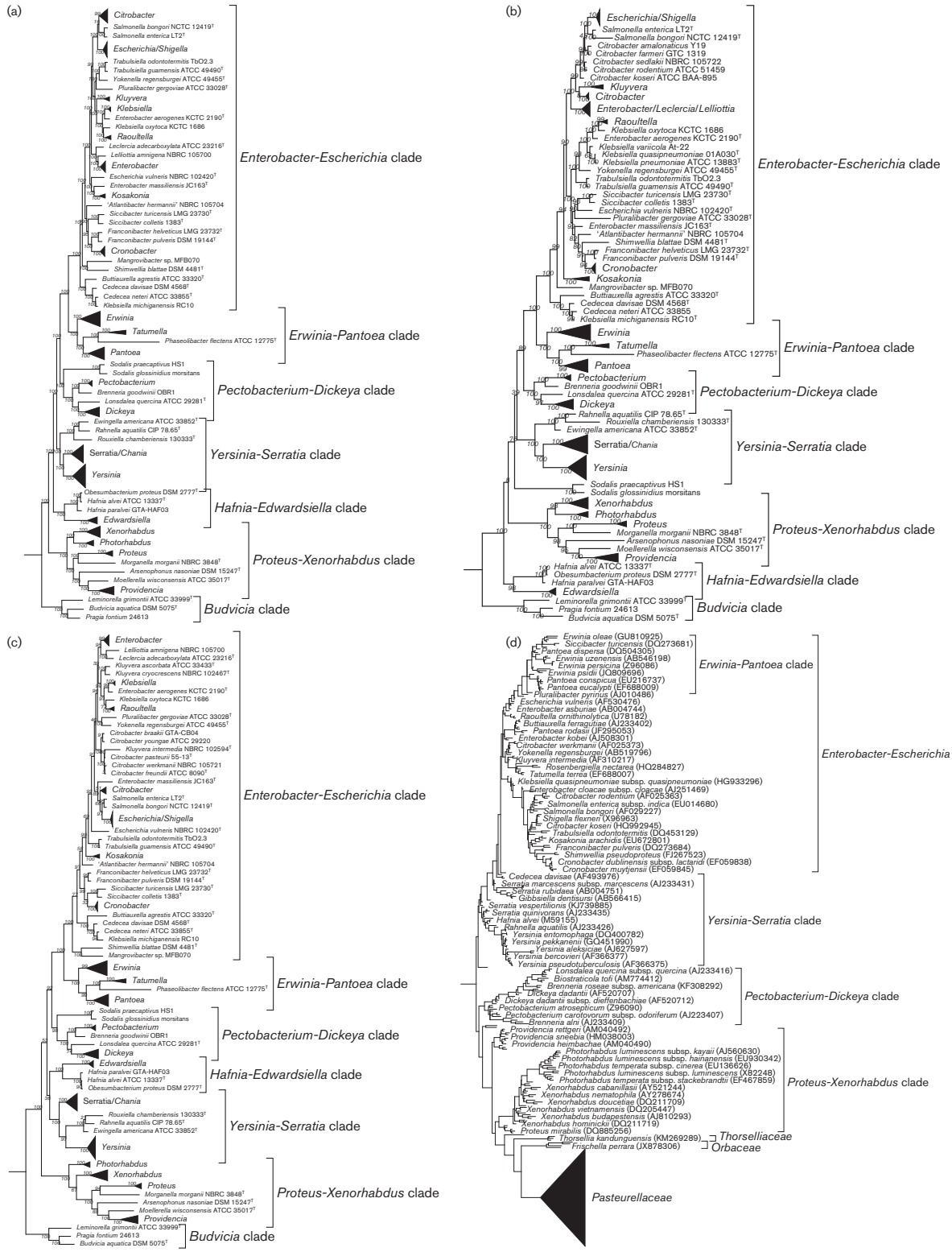

Fig. 1. Maximum-likelihood phylogenetic trees for 179 representative genome sequenced members of the order '*Enterobacteriales*' spanning 49 genera with validly published names based on (a) 1548 core protein families, (b) 53 ribosomal proteins, and (c) four MLSA proteins (GyrB, RpoB, AtpD and InfB). The seven major clades identified within the order '*Enterobacteriales*' are labelled. SH-like statistical support values are shown at branch nodes. Versions of these phylogenetic trees showing all 179 tree leaves are provided in Figs S1–S3. (d) A phylogenetic tree based on the 16S rRNA gene

reproduced from the All-Species Living Tree project release 123 (Yarza *et al.*, 2008; Yilmaz *et al.*, 2013). The closest analogues to the clades observed in the protein-based phylogenetic trees are labelled where possible. All branches within the order '*Enterobacteriales*' are shown, but not all leaves are labelled; a version of this phylogenetic tree showing labels for all branches in the phylogenetic tree is provided in Fig. S4.

Species Living Tree project release 123 (Yarza *et al.*, 2008; Yilmaz *et al.*, 2013), is shown in Figs 1d and S4.

The branching pattern of the main groups within the order '*Enterobacteriales*' in the genome-based tree, the ribosomal protein tree, and the MLSA-based phylogenetic tree are highly consistent. In each of the phylogenetic trees, the members of the order '*Enterobacteriales*' form seven main groups/clades which are labelled in the phylogenetic tree figures. The first group, referred to as the *Enterobacter-Escherichia* clade, is the largest group within the order '*Enterobacteriales*' and consists of the genera '*Atlantibacter*', *Buttiauxella, Cedecea, Citrobacter, Cronobacter, Enterobacter, Escherichia, Franconibacter, Klebsiella, Kluyvera, Kosakonia, Leclercia, Lelliottia, Mangrovibacter, Pluralibacter, Raoultella, Salmonella, Shigella, Shimwellia, Siccibacter, Trabulsiella* and *Yokenella*. The *Erwinia-Pantoea* clade, which is present in a monophyletic grouping with the *Enterobacter-Escherichia* clade, consists of the genera *Erwinia, Pantoea, Phaseolibacter* and *Tatumella*. The *Pectobacterium-Dickeya* clade consists of the genera *Brenneria, Dickeya, Lonsdalea, Pectobacterium* and *Sodalis*, the *Yersinia-Serratia* clade consists of the genera *Chania*, *Ewingella, Rahnella, Rouxiella, Serratia* and *Yersinia*, the *Hafnia-Edwardsiella* clade consists of the genera *Edwardsiella, Hafnia* and *Obesumbacterium*, the *Proteus-Xenorhabdus* clade consists of the genera *Arsenophonus, Moellerella, Morganella, Photorhabdus, Proteus, Providencia* and *Xenorhabdus*, and lastly, the *Budvicia* clade consists of the genera *Budvicia, Leminorella* and *Pragia*. Apart from one exception, the genera within the order '*Enterobacteriales*' consistently branch together within the clades described above as distinct monophyletic groupings in the phylogenetic trees. The sole exception to these groupings is observed in the ribosomal-protein-based phylogenetic tree wherethe two representative members of the genus *Sodalis*, which are early branching members of the *Pectobacterium-Dickeya* clade in other phylogenetic trees, branch outside of the *Pectobacterium-Dickeya* clade, exhibiting no branching affinity for any of the main clades within the order '*Enterobacteriales*' in the ribosomal-protein-based phylogenetic tree. The early branching of the genus *Sodalis* from other members of the *Pectobacterium-Dickeya* clade in the genome- and MLSA-based phylogenetic trees, and the lack of branching affinity of the genus *Sodalis* to any main clade within the order '*Enterobacteriales*' in the ribosomal-protein-based phylogenetic tree, may be a result of the endosymbiotic adaptations of the genus *Sodalis* which have led to significant genome degradation and genetic divergence from its closest relatives (Toh *et al.*, 2006).

The genera *Buchnera, Plesiomonas* and *Wigglesworthia* exhibit atypical branching characteristics and are not included in the main figures, but the results for them are presented in Figs S1b, S2b and S3b. The endosymbiotic genera *Buchnera* and *Wigglesworthia* possess extremely long branches and form a monophyletic cluster. However, the monophyletic clustering of *Buchnera* and *Wigglesworthia* is potentially a consequence of long branch attraction artefacts, compositional bias due to their small A+T-rich genomes, and rooting (Bergsten, 2005; Herbeck *et al.*, 2005; Husník *et al.*, 2011; Philippe *et al.*, 2005; Williams *et al.*, 2010). The genera *Buchnera* and *Wigglesworthia* branch between the *Enterobacter-Escherichia* and the *Erwinia-Pantoea* clades in both the genome- and ribosomal-protein-based phylogenetic trees (Figs S1b and S2b), but branch earlier, after the *Budvicia* clade, in the MLSA-based phylogenetic tree. In contrast to these two genera, the genus *Plesiomonas* forms an early diverging outgroup of the order '*Enterobacteriales*' in the genome- and MLSA-based phylogenetic trees Figs S1b and S3b), and branches between the *Vibrionaceae* and *Pasteurellaceae* members in the ribosomal-protein-based phylogenetic tree (Fig. S2b). It is of interest to note that *Plesiomonas* has historically been difficult to place in a specific taxonomic group due to its atypical phenotypic characteristics and highly recombinant genome (Janda *et al.*, 2016; Salerno *et al.*, 2007). The genus *Plesiomonas* was originally placed within the family *Vibrionaceae* before transfer to the family *Enterobacteriaceae* (Janda, 2005; Ruimy *et al.*, 1994).

The genera within the '*Enterobacteriales*' in the 16S rRNA gene-based phylogenetic tree (Figs 1, S1d and S4) exhibit extensive polyphyly and many of the clades identified in the genome-, ribosomal protein-, and MLSA-based phylogenetic trees are poorly resolved or unsupported in the 16S rRNA gene-based phylogenetic tree. Similar to the genome-, ribosomal protein-, and MLSA-based phylogenetic trees, a monophyletic grouping of the genera within the *Enterobacter-Escherichia* clade and the *Erwinia-Pantoea* clade is observed in the 16S rRNA gene-based phylogenetic tree. However, the members of the *Erwinia-Pantoea* clade branch within the *Enterobacter-Escherichia* clade in the 16S rRNA gene-based phylogeny instead of branching as two distinct, but related groups. In the 16S rRNA gene-based phylogenetic tree, the *Yersinia-Serratia* clade and the *Hafnia-Edwardsiella* clade, as well as the genus *Budvicia* from the *Budvicia* clade, form a highly intermixed, paraphyletic outgroup of the *Enterobacter-Escherichia* and *Erwinia-Pantoea* clades (simply labelled as the *Yersinia-Serratia* clade in Fig. 1d). The *Pectobacterium-Dickeya* clade forms a distinct, monophyletic grouping in the 16S rRNA gene-based phylogenetic tree that is largely consistent with the branching seen in the genome-, ribosomal protein- and MLSA-based phylogenetic trees. The members of the *Proteus-*

Xenorhabdus clade cluster together in a paraphyletic grouping. Notably, the earliest branching members of the order '*Enterobacteriales*' in the genome-, ribosomal protein- and MLSA-based phylogenetic trees (viz. the *Proteus-Xenorhabdus* and *Budvicia* clades) and the members of the *Pectobacterium-Dickeya* clade exhibit closer affinity to other families within the class *Gammaproteobacteria* (viz. *Pasteurellaceae*, *Orbaceae* and *Thorselliaceae*) than to the other members of the *Enterobacteriaceae*, making the order '*Enterobacteriales*' polyphyletic in the 16S rRNA gene-based phylogenetic tree.

Genome relatedness of the members of the order '*Enterobacteriales*'. The gold standard technique in microbial classification is the DNA–DNA hybridization methodology (Gevers *et al.*, 2005; Goris *et al.*, 2007). Recently, *in silico* measures of genome to genome relatedness have been used in classification as replacements for the DNA–DNA hybridization procedure (Auch *et al.*, 2010; Konstantinidis & Tiedje, 2005; Rosselló-Mora, 2006). Here we utilize a measure of genome to genome relatedness with applications for phylogeny and classification, the proportion of shared protein families in a pair of genomes, that has alternately been referred to as Percentage of Conserved Proteins (Qin *et al.*, 2014) and Alignment Fraction (Varghese *et al.*, 2015) in prior studies (Fig. 2). This measure of genome to genome relatedness is particularly useful at higher taxonomic ranks because of its large dynamic range which extends from >60 % for closely related organisms (Qin *et al.*, 2014; Varghese *et al.*, 2015) to <1 % for distantly related organisms (Ciccarelli *et al.*, 2006; Dagan & Martin, 2006). The seven main groups of genera observed in our phylogenetic trees (Fig. 1) exhibit distinctly higher genome to genome relatedness to each other than to other groups of genera in our analysis of shared protein families (Fig. 2). Additionally, the proportion of shared protein content also supports the general branching order observed in the phylogenetic trees

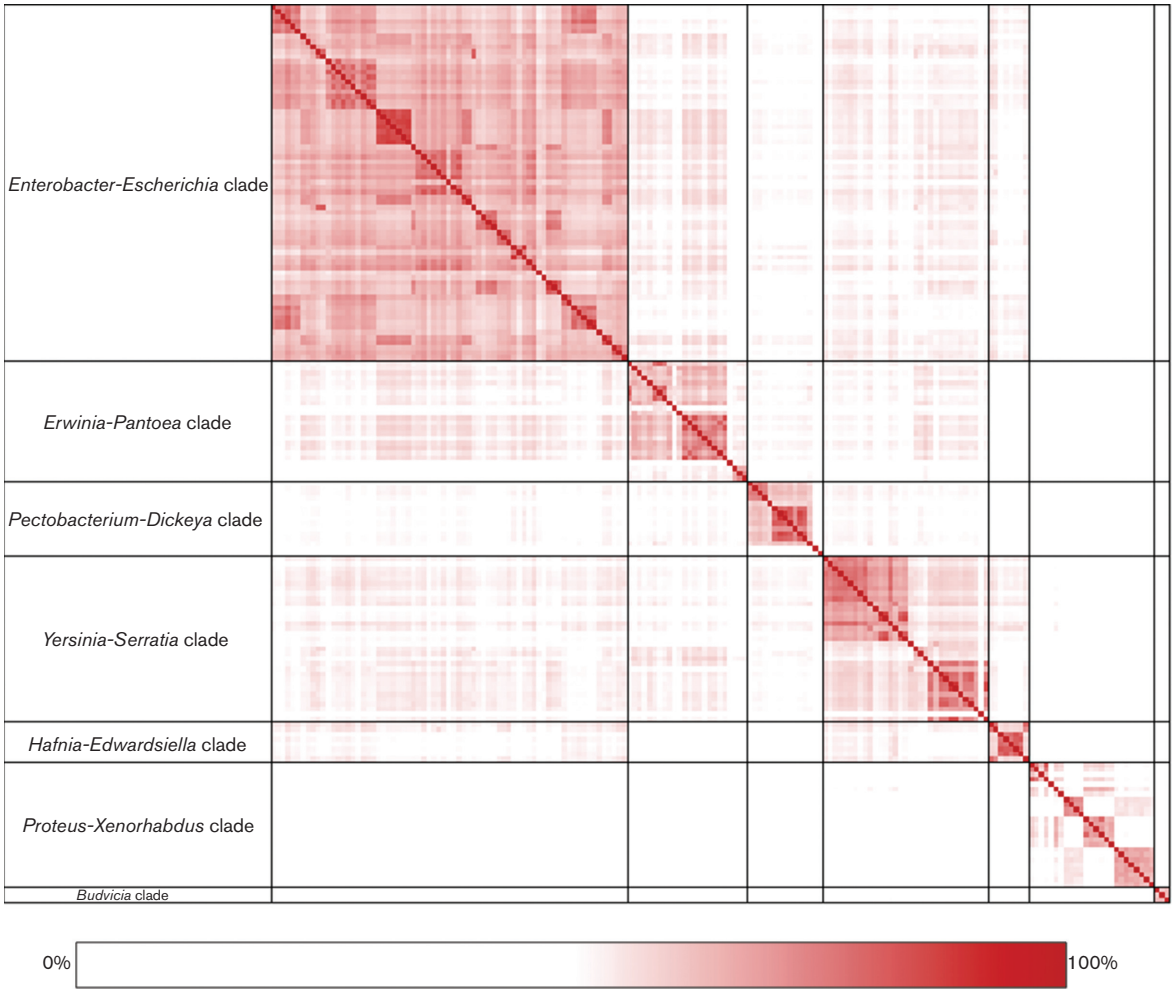

Fig. 2. A matrix of the percentage of shared protein families in the 179 genomes of members of the order '*Enterobacteriales*' analysed in this study. Genome pairs that share more protein families are shaded more darkly. The numerical values underlying this matrix are provided in Table S3, and a presence/absence matrix for each of the shared protein families identified in this work is provided in Table S4.

with the *Enterobacter-Escherichia, Erwinia-Pantoea, Yersinia-Serratia, Hafnia-Edwardsiella* and *Pectobacterium-Dickeya* clades exhibiting a higher proportion of shared protein families with each other than to the early branching *Proteus-Xenorhabdus* and *Budvicia* clades.

Identification of conserved signature indels

Molecular characteristics which are unique to the order '*Enterobacteriales*'. In this work, we have completed a comprehensive comparative analysis of the publically available genomes from members of the order '*Enterobacteriales*' in order to identify discrete markers of common evolutionary ancestry in the form of CSIs. We have identified 69 CSIs which are distinctive characteristics of the order '*Enterobacteriales*' and its main constituent clades. Five of these CSIs are a shared, distinguishing characteristic of the members of the order '*Enterobacteriales*' in its entirety. An example of one such CSI, consisting of a single amino acid (aa) insertion in the L-arabinose isomerase protein, is shown in Fig. 3. This insertion is present in homologues from all sequenced members (>150) from the order '*Enterobacteriales*' and is absent in homologues from all other bacteria (top 1000 BLAST hits examined). More detailed information for this CSI is shown in Fig. S5. Four additional CSIs, which are distinguishing characteristics of the members of the order '*Enterobacteriales*', were identified in elongation factor P-like protein YeiP, peptide ABC transporter permease, pyrophosphatase and a hypothetical protein; sequence alignments for these CSIs are shown in Figs S6–S9 and some properties of these CSIs are briefly summarized in Table 1. The unique shared presence of these CSIs in all of the '*Enterobacteriales*', but in no other bacteria, except for one or two isolated exceptions provides evidence, independent of the phylogenetic trees, that the order '*Enterobacteriales*' is monophyletic in nature and these CSIs are distinguishing characteristics of this large group of bacteria. Homologues from the genera *Buchnera* and *Wigglesworthia* were not identified in any of the five proteins containing CSIs shared by all '*Enterobacteriales*', while homologues from the genus *Plesiomonas* were only identified for the peptide ABC transporter permease (Fig. S8) and pyrophosphatase (Fig. S9). In both cases, the genus *Plesiomonas* did not share the CSI shared by all other '*Enterobacteriales*'.

Molecular characteristics distinguishing the main clades within the order '*Enterobacteriales*'. The main focus of this study is on the identification of unique shared characteristics, which can be used to distinguish the main groups within the order '*Enterobacteriales*'. We have identified a total of 66 CSIs which distinguish the seven main groups of genera within the order '*Enterobacteriales*', observed in the phylogenetic trees, from each other and from all other bacteria. A number of additional CSIs distinguishing the *Pectobacterium-Dickeya* clade were identified in a previous study (Naushad *et al.*, 2014), and the specificities were re-examined in this work. The identified CSIs which distinguish each of the seven main clades of the order '*Enterobacteriales*' are described below.

Clade 1: The *Enterobacter-Escherichia* clade. The members of the genera *Salmonella, Citrobacter, Escherichia* and *Shigella* are a well-recognized and highly researched grouping of genera within the order '*Enterobacteriales*' (Fukushima *et al.*, 2002; Gordienko *et al.*, 2013; Nataro *et al.*, 2011; Samuel *et al.*, 2004). *Escherichia coli*, in particular, is one of the most important model organisms in microbiology and has been highly studied and sequenced (Blattner *et al.*, 1997; Chaudhuri & Henderson, 2012; Gordienko *et al.*, 2013; http://www.ncbi.nlm.nih.gov/genome). These genera and their closest relatives (viz. *Enterobacter, Cronobacter, Klebsiella*, etc.) are the largest grouping of genera within the order '*Enterobacteriales*'. This grouping of genera, labelled the *Enterobacter-Escherichia* clade, is clearly observed in our genome-, ribosomal protein- and MLSA-based phylogenetic trees and an association between these genera is also seen in 16S rRNA gene-based phylogenies (Figs 1 and S1–S4). We have identified 21 CSIs which are shared, distinguishing characteristics of the members of the *Enterobacter-Escherichia* clade in our phylogenetic trees, providing evidence that the members of the *Enterobacter-Escherichia* clade form a coherent phylogenetic grouping. An example of a unique, characterizing CSI which is shared by the members of the *Enterobacter-Escherichia* clade is depicted in Fig. 4(a). The CSI consists of a 3 aa insert in the protein NADH:ubiquinone-oxidoreductase (subunit M), which is present in all of the sequenced species/homologues belonging to this group, and absent in other homologues from the '*Enterobacteriales*'. More detailed information for this signature is shown in Fig. S10 and the sequence alignments for the 20 other signatures depicting the different identified CSIs which are also distinguishing characteristics of the *Enterobacter-Escherichia* clade are shown in Figs S11–S30 and their properties are briefly summarized in Table 2.

Clade 2: The *Erwinia-Pantoea* clade. The genera *Erwinia* and *Pantoea* are a well-studied grouping of bacteria containing a number of insect and plant pathogens (Coutinho & Venter, 2009; Zhang & Qiu, 2015). These genera and their closest relatives, labelled the *Erwinia-Pantoea* clade in our phylogenetic trees, exhibit a close association with the members of the *Enterobacter-Escherichia* clade. In our genome-, ribosomal protein- and MLSA-based phylogenetic trees, the members of the *Erwinia-Pantoea* clade branch as a distinct subgroup in a monophyletic grouping with the *Enterobacter-Escherichia* clade and branch within the *Enterobacter-Escherichia* clade in 16S rRNA gene-based phylogenetic trees (Fig. 1 and Figs S1–S4). We have identified 12 CSIs that are unique distinguishing characteristics of the *Erwinia-Pantoea* clade and an additional 6 CSIs that are shared characteristics of both the *Enterobacter-Escherichia* and *Erwinia-Pantoea* clades. An example of each type of CSI is shown here. The first CSI consists of a single amino acid insertion in the protein glutamate-cysteine ligase that is uniquely present in all sequenced members (>20) of the *Erwinia-Pantoea* clade (Fig. 4b), while the second CSI consists of a single amino acid insertion in the protein cysteine synthase A that is uniquely present in homologues from

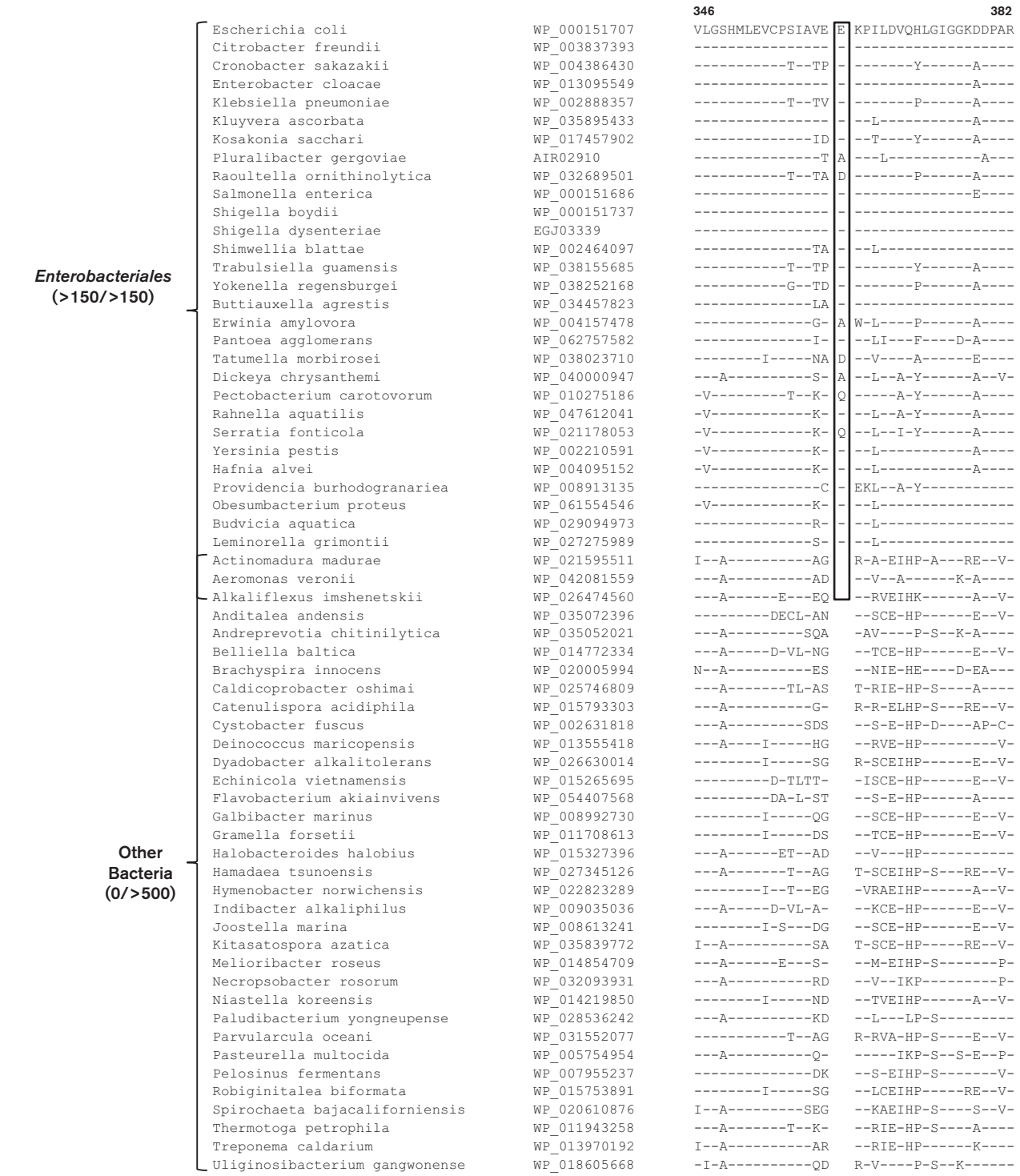

Fig. 3. A partial sequence alignment of the protein L-arabinose isomerase containing a single amino acid insert (boxed) that is exclusively found in all '*Enterobacteriales*' members and is absent in other bacteria. Sequence information for a limited number of '*Enterobacteriales*' and other bacteria are shown here, but unless otherwise indicated similar CSIs were detected in all members of the indicated group and not detected in any other species in the top 500–1000 BLAST hits. Dashes (–) in the alignments indicate identity with the residue in the top sequence. GenBank accession numbers for each sequence are indicated in the second column. Additional CSIs specific for '*Enterobacteriales*' are presented in Table 1 and Figs S5–S9.

Table 1. Summary of conserved signature indels specific for all members within the order '*Enterobacteriales*'

Protein name	GenBank accession number	Figure number	Indel size	Indel position
L-arabinose isomerase	WP_000151707	Fig. 3; Fig. S5	1 aa ins	346–382
Elongation factor P-like protein YeiP	WP_001610470	Fig. S6	1 aa ins	89–129
Hypothetical protein	ACI70584	Fig. S7	6 aa ins	143–185
Peptide ABC transporter permease	WP_000552295	Fig. S8	3 aa ins	157–198
Pyrophosphatase	WP_000640873	Fig. S9	1 aa ins	105–148

members of both the *Enterobacter-Escherichia* and *Erwinia-Pantoea* clades (Fig. 5a). In both cases, similar insertions were not identified in any other related protein homologues from other organisms. More detailed information for these two CSIs as well sequence alignments for the 16 other CSIs, which are specific for either the *Erwinia-Pantoea* clade or supporting a grouping of the *Enterobacter-Escherichia* and *Erwinia-Pantoea* clades are shown in Figs S31–S48 and their properties are briefly summarized in Table 3.

It is of much interest that of the 12 CSI-containing proteins that are distinguishing characteristics of the *Erwinia-Pantoea* clade, homologues for three of them were detected in *Buchnera aphidicola* (Figs 4b, S31, S36 and S41). In each case, *Buchnera aphidicola* shared the characteristic CSIs identified in the CSI-containing proteins with the members of the *Erwinia-Pantoea* clade. Additionally, *Buchnera aphidicola* homologues were identified for two proteins containing CSIs shared by both the *Enterobacter-Escherichia* and *Erwinia-Pantoea* clades (Figs 5, S43 and S45). These results provide reliable evidence that support previous assertions that *Buchnera aphidicola* is specifically related to the members of the *Erwinia-Pantoea* clade (Husník *et al.*, 2011). Homologues for most of the CSI-containing proteins shared by the *Erwinia-Pantoea* clade or the *Enterobacter-Escherichia* clade were not found in *Wigglesworthia glossinidia* and, in the few cases where they were found (Figs S24 and S36), *Wigglesworthia glossinidia* did not share the CSI with either of the two clades. However, *Wigglesworthia glossinidia* was found to specifically share a CSI in a ribonucleotide reductase stimulatory protein which is a distinguishing characteristic of both the *Enterobacter-Escherichia* and *Erwinia-Pantoea* clades (Fig. S46). This CSI supports the view that *Wigglesworthia glossinidia* is also specifically related to either the *Erwinia-Pantoea* clade or the *Enterobacter-Escherichia* clade, though it is likely a more distant relative of either clade than *Buchnera aphidicola*.

Clade 3: The *Pectobacterium-Dickeya* clade. The members of the genera *Dickeya, Pectobacterium* and *Brenneria* are important plant-pathogenic bacteria (Hauben *et al.*, 1998; Ma *et al.*, 2007; Young & Park, 2007; Zhang *et al.*, 2016). *Dickeya, Pectobacterium*, and *Brenneria* branch with the genera *Lonsdalea* and *Sodalis* in our genome- and MLSA-based phylogenetic trees (Fig. 1a, c), in a grouping referred to as the *Pectobacterium-Dickeya* clade. However, the genus *Sodalis* does not branch with the other members of this clade in our ribosomal protein-based phylogenetic tree (Fig. 1b). Here we describe four CSIs which are shared by *Brenneria, Dickeya, Lonsdalea, Pectobacterium* and *Sodalis* providing independent evidence of the unique shared ancestry of this group of species. An example of one of these CSIs, consisting of a 2-aa insertion in a hypothetical protein that is uniquely present in homologues from *Brenneria, Dickeya, Lonsdalea, Pectobacterium* and *Sodalis* and absent in all other bacterial groups is shown in Fig. 5(b). More detailed information for this CSI is shown in Fig. S49. In earlier work, we have reported 10 CSIs which, at that time, were indicated to be specific for the genera *Dickeya, Pectobacterium* and *Brenneria* (Naushad *et al.*, 2014). A re-examination of these CSIs has shown that two of these previously identified CSIs (in a two-component sensor histidine kinase protein and flagellar motor protein MotB) were found in all members of the *Pectobacterium-Dickeya* clade. However, the remaining eight CSIs identified in our earlier work (and not described here) (Naushad *et al.*, 2014) were either not found in homologues from *Sodalis* or the homologues of these proteins were not detected in members of the genus *Sodalis*, and thus they are specific for a subclade of the enlarged *Pectobacterium-Dickeya* clade described here. Sequence alignments for the three other CSIs which are distinguishing characteristics of the *Pectobacterium-Dickeya* clade are shown in Figs S50–S52 and their properties are briefly summarized in Table 4.

Clade 4: The *Yersinia-Serratia* clade. The genus *Yersinia* contains the causative agent of the plague, a disease which led to one of the most devastating pandemics in human history. Consequently, the members of the genus *Yersinia* are the subjects of significant research interest (Eppinger *et al.*, 2010; Morelli *et al.*, 2010; Parkhill *et al.*, 2001; Perry & Fetherston, 1997). In our genome-, ribosomal protein- and MLSA-based phylogenetic trees (Fig. 1a–c), the members of the genus *Yersinia* are part of a distinct group which contains the genera *Chania*, *Ewingella, Rahnella, Rouxiella*and *Serratia*, referred to as the *Yersinia-Serratia* clade. We have identified three CSIs which are shared, distinguishing characteristics of the members of the *Yersinia-Serratia* clade, providing independent evidence that the members of these genera shared a unique common ancestor. One example of such a CSI, shown in Fig. 6(a), consists of a single aa insertion in the TetR family transcriptional regulator protein found in homologues from the members of the *Yersinia-Serratia* clade. More detailed information for this signature

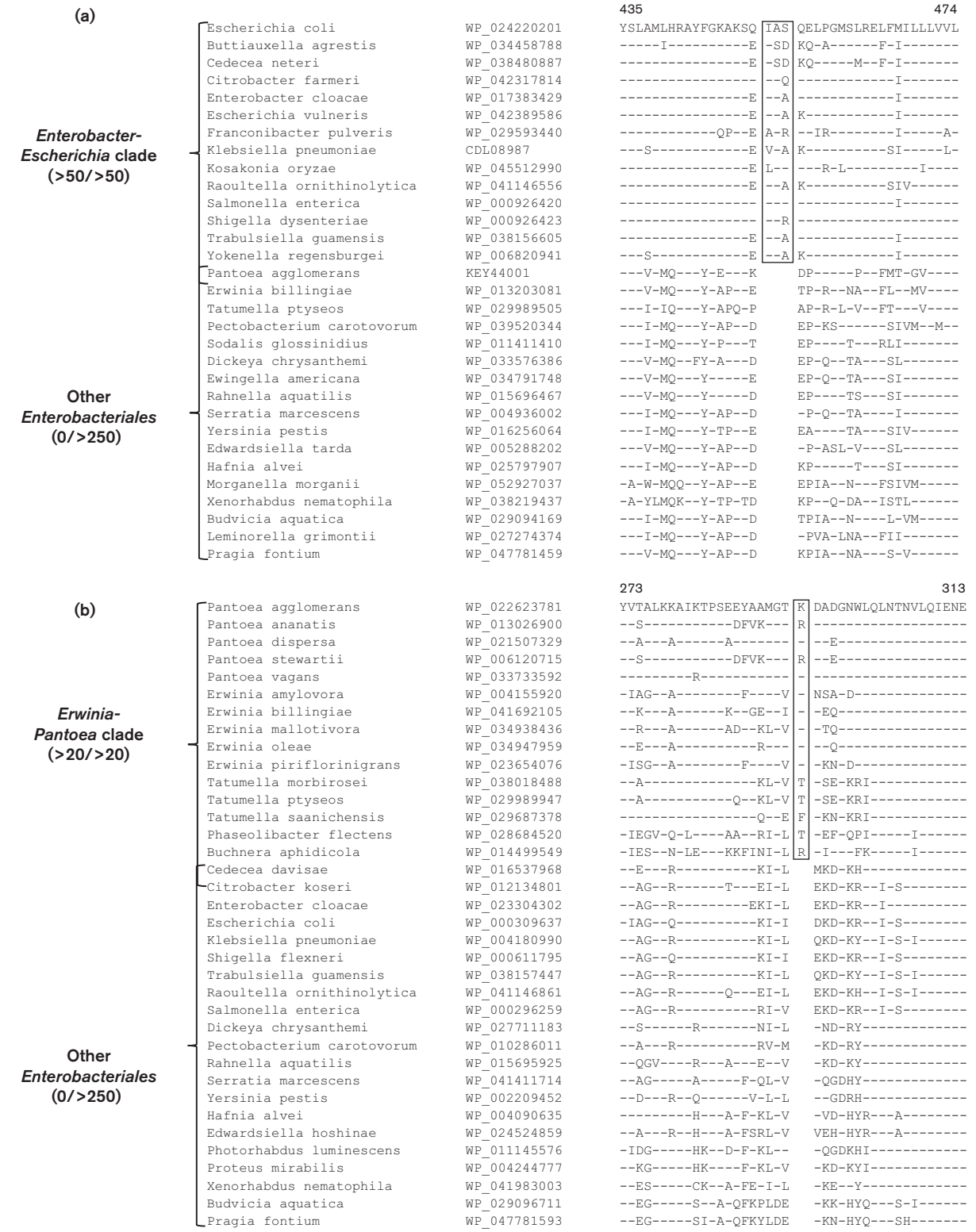

Fig. 4. Partial sequence alignments of (a) the protein NADH:ubiquinone oxidoreductase (subunit M) containing a three amino acid insert (boxed) that is exclusively found in all members within the *Enterobacter-Escherichia* clade, and (b) the protein glutamate – cysteine ligase containing a single amino acid insert (boxed) exclusive to members within the *Erwinia-Pantoea* clade. Additional CSIs specific for these clades are presented in Table 2 and Figs S10–S30 for the *Enterobacter-Escherichia* clade and Table 3 and Figs S31–S42 for the *Erwinia-Pantoea* clade.

Table 2. Summary of conserved signature indels specific for the members of the *Enterobacter-Escherichia* clade

Protein name	GenBank accession number	Figure number	Indel size	Indel position
NADH: ubiquinone oxidoreductase subunit M	WP_024220201	Fig. 4a; Fig. S10	3 aa ins	435–474
Twitching motility protein PilT	CAR94647	Fig. S11	4 aa del	32–82
2, 3-dihyroxybenzoate-AMP ligase	WP_001589860	Fig. S12	1 aa del	126–184
ATP/GTP-binding protein	CTV70932	Fig. S13	1 aa del	56–96
Multifunctional fatty acid oxidation complex subunit alpha	WP_032330678	Fig. S14	1 aa ins	548–586
S-formylglutathione hydrolase	WP_000421369	Fig. S15	2 aa ins	187–230
Aspartate-semialdehyde dehydrogenase	WP_001289176	Fig. S16	1 aa ins	165–201
Epimerase	WP_009430590	Fig. S17	1 aa del	198–233
Membrane protein	WP_000912606	Fig. S18	2 aa del	158–185
Formate hydrogenlyase subunit 7	CAA35552	Fig. S19	5 aa del	208–245
Glutathione S-transferase	WP_000779789	Fig. S20	1 aa del	134–168
Major facilitator superfamily transporter	WP_032237477	Fig. S21	1 aa ins	243–281
Peptide ABC transporter ATP-binding protein	WP_001572064	Fig. S22	1 aa ins	283–325
Major facilitator superfamily transporter	WP_000185209	Fig. S23	1 aa del	271–310
Phosphoglucosamine mutase	WP_000071132	Fig. S24	1 aa ins	359–399
Glycosyl hydrolase 1 family protein	WP_009671380	Fig. S25	1 aa del	248–283
23S rrna [uracil(1939)-C(5)]-methyltransferase	WP_000046777	Fig. S26	6 aa del	93–132
Co-chaperone HscB	WP_000384406	Fig. S27	1 aa del	97–141
N-acetylmuramoyl-L-alanine amidase	WP_000102887	Fig. S28	1 aa del	85–117
Sulfate ABC transporter ATP-binding protein CysA	AAA23639	Fig. S29	1 aa del	308–346
LPS assembly protein LptD	WP_032172667	Fig. S30	1 aa ins	250–285

as well as sequence alignments for the two other identified CSIs which are distinguishing characteristics of the *Yersinia-Serratia* clade are shown in Figs S53–S55 and their properties are briefly summarized in Table 4.

Clade 5: The *Hafnia-Edwardsiella* clade. The genera *Edwardsiella, Hafnia* and *Obesumbacterium* are minor pathogens of animals and humans (Huys *et al.*, 2010; Janda & Abbott, 2006; Koivula *et al.*, 2006; Janda & Abbott, 1993). An association between the genera *Hafnia* and *Obesumbacterium* has been observed in a number of previous studies (Octavia & Lan, 2014; Paradis *et al.*, 2005; Priest & Barker, 2010), however, the genus *Edwardsiella* shows limited association with the genera *Hafnia* and *Obesumbacterium* in 16S rRNA gene-based phylogenetic trees (Fig. S4). The genera *Edwardsiella, Hafnia* and *Obesumbacterium* form a distinct phylogenetic grouping, referred to as the *Hafnia-Edwardsiella* clade, in our genome-, ribosomal protein- and MLSA-based phylogenetic trees (Fig. 1a–c). We have also identified four CSIs which are shared by *Edwardsiella, Hafnia* and *Obesumbacterium*. An example of one CSI that is uniquely shared by the members of the *Hafnia-Edwardsiella* clade is shown in Fig. 6(b). This CSI consists of a 4-aa insertion in the two-component system response regulator protein GIrR, which is uniquely found in homologues from *Edwardsiella, Hafnia* and *Obesumbacterium*. More detailed information for this CSI and the sequence alignments for the three other CSIs which are distinguishing characteristics of the *Hafnia-Edwardsiella* clade are shown in Figs S56–S59 and their properties are briefly summarized in Table 4.

Clade 6: The *Proteus-Xenorhabdus* clade. The genera *Xenorhabdus* and *Photorhabdus* are a closely related group of symbiotic bacteria associated with nematode hosts with which they have synergistic entomopathogenic effects against insects (Forst *et al.*, 1997; Nielsen-LeRoux *et al.*, 2012). Previous research has suggested that the closest evolutionary neighbours of *Xenorhabdus* and *Photorhabdus* are the genera *Arsenophonus, Proteus* and *Providencia* (Boemare & Akhurst, 2006; Tailliez *et al.*, 2010; Trowbridge *et al.*, 2006). However, *Xenorhabdus, Photorhabdus, Arsenophonus, Proteus* and *Providencia* do not form a monophyletic clade in 16S rRNA gene-based phylogenetic trees (Fig. 1d). In our genome-, ribosomal protein- and MLSA-based phylogenetic trees (Fig. 1a–c), the genera *Arsenophonus, Moellerella, Morganella, Photorhabdus, Proteus, Providencia* and *Xenorhabdus* form a distinct, monophyletic grouping, referred to as the *Proteus-Xenorhabdus* clade. We have identified seven CSIs which are uniquely shared characteristics of the members of the *Proteus-Xenorhabdus* clade. One of these CSIs, a 1-aa deletion in the protein dihydrolipoamide succinyltransferase, in homologues from the *Proteus-Xenorhabdus* clade, is shown in Fig. 7(a). More detailed information for this CSI as well as the sequence information for thesix other identified CSIs which are distinguishing characteristics of the *Proteus-Xenorhabdus* clade are shown in Figs S60–S66 and their properties are briefly summarized in Table 4. These CSIs provide independent evidence in support of the inference from core genome-, ribosomal protein- and MLSA-based phylogenetic trees, that the members of the *Proteus-Xenorhabdus* clade form a monophyletic clade derived from a unique common ancestor.

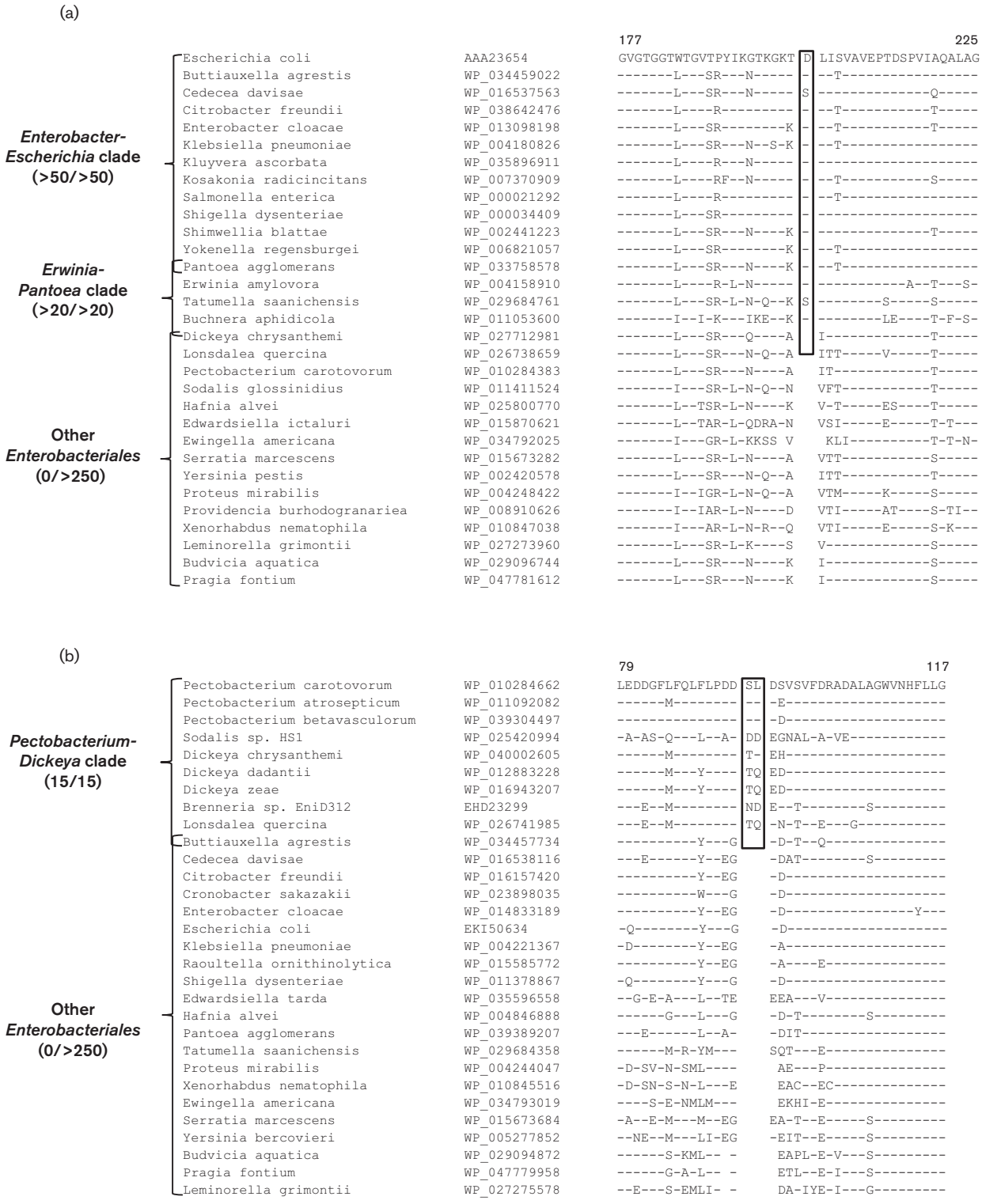

Fig. 5. Partial sequence alignments of (a) the protein cysteine synthase A containing a single amino acid insert (boxed) that is shared exclusively by members of both the *Enterobacter-Escherichia* clade and the *Erwinia-Pantoea* clade, and (b) a hypothetical protein containing a two amino acid insert (boxed) exclusive to members within the *Dickeya-Pectobacterium* clade. Additional CSIs specific for these clades are presented in Table 3 and Figs S43–S48 for CSIs shared by both the *Enterobacter-Escherichia* and *Erwinia-Pantoea* clades, Table 4 and Figs S49–S52 for the *Dickeya-Pectobacterium* clade.

Clade 7: The *Budvicia* clade. The members of the genera *Budvicia, Leminorella* and *Pragia* are characterized by their H_2S-positive phenotypes and have long been thought to be related (Janda, 2006; Paradis *et al.*, 2005; Schindler *et al.*, 1991; Spröer *et al.*, 1999). A grouping of these three genera, referred to as the *Budvicia* clade, is observed in our genome-ribosomal protein- and MLSA-based phylogenetic trees (Fig. 1a–c). A previously reported CSI in the *atpD* gene also supports a specific relationship of the genera *Budvicia, Leminorella* and *Pragia* (Paradis *et al.*, 2005). Here, we have identified nine additional CSIs which are shared by these three genera. One example of a CSI shared by the genera *Budvicia, Leminorella* and *Pragia* is shown in Fig. 7(b). The CSI consists of a 4-aa insertion in the protein bifunctional protein-disulfide isomerise/oxidoreductase DsbC in homologues from *Budvicia, Leminorella* and *Pragia* which is absent in homologues from all other species. Detailed information for this signature is shown in Fig. S67. Sequence alignments for the eight additional CSIs which are also distinguishing characteristics of the *Budvicia* clade are shown in Figs S68–S75 and their properties are briefly summarized in Table 4.

DISCUSSION

Understanding the phylogeny and interrelationships of the genera within the order '*Enterobacteriales*' has proven difficult using the 16S rRNA gene and other single-gene based approaches (Dauga, 2002; Francino *et al.*, 2006; Fukushima *et al.*, 2002; Hauben *et al.*, 1998; Naum *et al.*, 2008; Pham *et al.*, 2007; Roggenkamp, 2007; Spröer *et al.*, 1999; Tailliez *et al.*, 2010). The advent of ubiquitous genome sequencing technology now presents us with a wealth of genomic sequence data from a broad range of organisms, spanning a majority of the diversity within the order '*Enterobacteriales*' (http://www.ncbi.nlm.nih.gov/genome), from which novel and reliable inferences regarding the evolutionary relationships of the genera within the order '*Enterobacteriales*' can be drawn. The analyses of the members of the order '*Enterobacteriales*' presented here, consisting of phylogenetic reconstructions based on 1548 core proteins, 53 ribosomal proteins and four MLSA proteins (Fig. 1a–c), analyses of overall genome similarity (Fig. 2), and the identification of shared distinguishing molecular characteristics (Fig. 8, Tables 1–4), represent the first comprehensive, genome-scale taxonomic analysis of the entirety of the order '*Enterobacteriales*'.

The phylogenetic trees produced in this study, utilizing 1548 core proteins, 53 ribosomal proteins and four MLSA proteins from 179 representative genomes from the order '*Enterobacteriales*', consistently support the existence of the seven main groups of genera within the order. Additionally, an independently created genome-based phylogenetic tree produced by the curators of the PATRIC database (Wattam *et al.*, 2014) utilizing over 1000 genome sequences from

Table 3. Summary of conserved signature indels specific for the members of the *Erwinia-Pantoea* clade or the grouping of both the *Enterobacter-Escherichia* and *Erwinia-Pantoea* clades

Protein name	GenBank accession number	Figure number	Indel size	Indel position	Specificity
Glutamate – cysteine ligase	WP_031594175	Fig. 4b Fig. S31	1 aa ins	273–313	*Erwinia-Pantoea* clade
DNA gyrase subunit B	WP_003849642	Fig. S32	2 aa del	597–635	
LPS assembly protein LptD	WP_050499087	Fig. S33	2 aa del	582–622	
Thiol:disulfide interchange protein DsbA precursor	WP_039387151	Fig. S34	1 aa ins	116–155	
Two-component sensor histidine kinase	WP_010670989	Fig. S35	1 aa ins	117–159	
RNA helicase	WP_004155135	Fig. S36	1 aa del	220–254	
Hypothetical protein	WP_022625284	Fig. S37	1 aa ins	137–174	
tRNA pseudouridine(13) synthase TruD	WP_003849102	Fig. S38	1 aa ins	191–232	
Glycine/betaine ABC transporter ATP-binding protein	WP_033778604	Fig. S39	1 aa del	286–331	
Transcriptional regulator	WP_004171762	Fig. S40	3 aa del	59–98	
Superoxide dismutase	WP_004161110	Fig. S41	1 aa del	30–64	
Stationary phase inducible protein CsiE	WP_022624119	Fig. S42	3 aa del	144–185	
Cysteine synthase A	AAA23654	Fig. 5a; Fig. S43	1 aa ins	177–225	Both the *Enterobacter-Escherichia* and *Erwinia-Pantoea* clades
2-oxo-3-deoxygalactonate kinase	WP_024224844	Fig. S44	4 aa del	77–122	
Hypothetical protein	WP_021513077	Fig. S45	1 aa del	77–127	
Ribonucleotide reductase stimulatory protein	WP_000080939	Fig. S46	1 aa del	13–50	
Membrane protein	WP_000589790	Fig. S47	1 aa ins	104–146	
Outer membrane protein assembly factor BamC	WP_000968394	Fig. S48	1 aa del	107–146	

Table 4. Summary of conserved signature indels specific for the members of the *Pectobacterium-Dickeya* clade, the *Yersinia-Serratia* clade, the *Hafnia-Edwardsiella* clade, the *Proteus-Xenorhabdus* clade and the *Budvicia* clade

Protein name	GenBank accession number	Figure number	Indel size	Indel position	Specificity
Hypothetical protein	WP_011411736	Fig. 5b; Fig. S49	2 aa ins	79–117	*Pectobacterium-Dickeya* clade
Transcriptional activator RhaS	WP_010285287	Fig. S50	1 aa ins	150–195	
Two-component sensor histidine kinase protein	WP_011092924	Fig. S51	1 aa ins	408–438	
Flagellar motor protein MotB	WP_011093267	Fig. S52	1 aa ins	234–261	
TetR family transcriptional regulator	CNI31513	Fig. 6a; Fig. S53	1 aa ins	43–89	*Yersinia-Serratia* clade
TetR family transcriptional regulator	CNI31513	Fig. S54	1 aa ins	82–123	
Hypothetical protein	WP_055781853	Fig. S55	7 aa ins	123–159	
Two-component system response regulator GIrR	WP_025800188	Fig. 6b; Fig. S56	1 aa ins	104–149	*Hafnia-Edwardsiella* clade
Glucose-1-phosphate adenylyltransferase	WP_025799356	Fig. S57	2 aa ins	252–286	
Transcriptional activator NhaR	WP_004089142	Fig. S58	2 aa ins	241–272	
Hybrid sensor histidine kinase/response regulator	WP_004847184	Fig. S59	4 aa del	134–168	
Dihydrolipoamide succinyltransferase	WP_006660450	Fig. 7a; Fig. S60	1 aa del	67–101	*Proteus-Xenorhabdus* clade
Xaa-Pro dipeptidase	WP_004246104	Fig. S61	1 aa ins	101–137	
Bifunctional UDP-sugar hydrolase (5′-nucleotidase)	WP_036895513	Fig. S62	2 aa ins	246–287	
Transcription repair coupling factor	WP_060556858	Fig. S63	1 aa del	273–305	
Phosphate acetyltransferase	WP_004248391	Fig. S64	1 aa del	27–60	
Histidine–tRNA ligase	KLU18800	Fig. S65	1 aa ins	308–345	
N-acetylmuramoyl-L-alanine amidase	WP_00449634	Fig. S66	1 aa del	316–374	
Bifunctional protein-disulfide isomerise/ oxidoreductase DsbC	WP_047781864	Fig. 7b; Fig. S67	4 aa ins	71–109	*Budvicia* clade
Hypothetical protein	WP_047781711	Fig. S68	3 aa ins	1281–1314	
Hypothetical protein	WP_047781711	Fig. S69	2 aa ins	1588–1620	
Hypothetical protein	WP_047779510	Fig. S70	2 aa ins	112–156	
Bifunctional protein-disulfide isomerise/ oxidoreductase DsbC	WP_047781864	Fig. S71	1 aa ins	21–52	
Transcriptional regulator	WP_047779627	Fig. S72	1 aa ins	42–79	
L-methionine/branched chain amino acid transporter	WP_047781898	Fig. S73	1 aa ins	284–320	
Hypothetical protein	WP_047779644	Fig. S74	10 aa ins	570–623	
D-alanine—D-alanine ligase	WP_047780169	Fig. S75	3 aa del	96–137	

members of the order '*Enterobacteriales*' exhibits highly similar inter-genus level branching to the phylogenetic trees produced in this work and supports the same groupings. The seven main groupings of genera were also supported by a measure of genomic similarity known as Percentage of Conserved Proteins (Qin *et al.*, 2014) or Alignment Fraction (Varghese *et al.*, 2015) (Fig. 2) which is based on the shared gene/protein families present in the genomes. Conversely, phylogenetic trees produced based on the 16S rRNA gene sequence (Fig. 1d) exhibit limited ability to resolve the clades identified in the genome-, ribosomal protein- and MLSA-based phylogenetic trees (Hauben *et al.*, 1998; Naum *et al.*, 2008; Octavia & Lan, 2014). Additionally, the branching of the genera and species within the order '*Enterobacteriales*' in 16S rRNA gene-based phylogenies shows considerable stochasticity depending on the algorithms used and the organisms analysed (Naum *et al.*, 2008; Octavia & Lan, 2014). Overall, the results obtained here substantiate previous suggestions that the 16S rRNA gene possesses limited utility in accurate phylogenetic reconstruction of inter-genus level relationships within the order '*Enterobacteriales*' (Naum *et al.*, 2008, 2011; Octavia & Lan, 2014).

The CSIs identified in this work provide a novel means of elucidating the common evolutionary ancestry of different groups within the order '*Enterobacteriales*' independently of phylogenetic trees. The most parsimonious explanation of the unique presence of multiple CSIs in a related group of organisms is the existence of a unique shared ancestor in which the genetic changes leading to these CSIs occurred which were

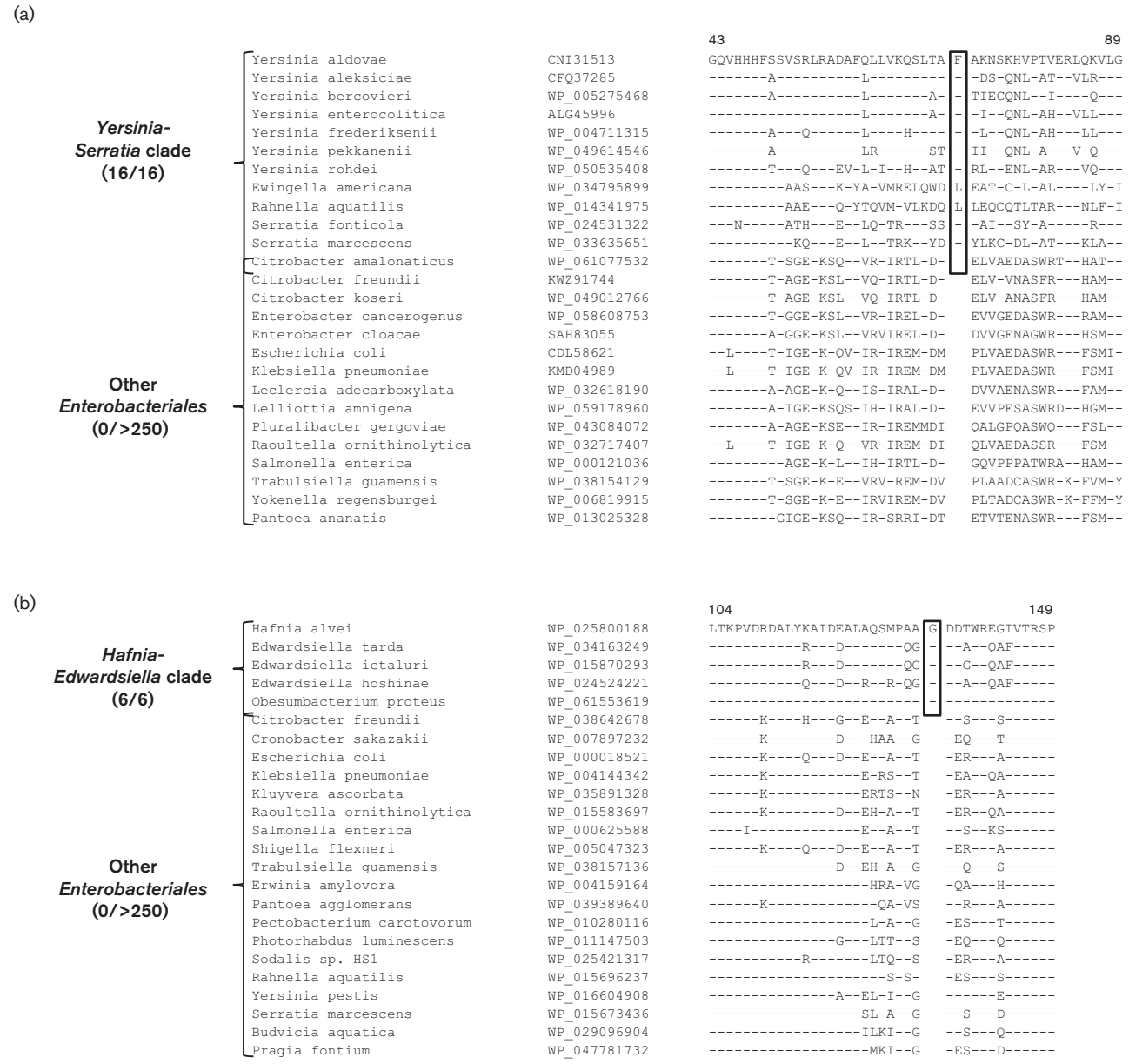

Fig. 6. Partial sequence alignments of (a) the protein TetR family transcriptional regulator containing a single amino acid insert (boxed) that is exclusively found in all members within the *Yersinia-Serratia* clade, and (b) the protein two-component system response regulator GlrR containing a single amino acid insert (boxed) exclusive to members within the *Hafnia-Edwardsiella* clade. Additional CSIs specific for these clades are presented in Figs S53–S55 for the *Yersinia-Serratia* clade and Table 3 and Figs S56–S59 for the *Hafnia-Edwardsiella* clade, and their characteristics are briefly described in Table 4.

then inherited by the descendent species. Thus, CSIs which are restricted to well-defined groups of organisms can be treated synapomorphic traits and used as independent support of monophyletic phylogenetic relationships (Gupta, 2014; Jones, 2012; Rokas & Holland, 2000). Here we describe 71 CSIs which are distinctive characteristics of the order '*Enterobacteriales*' and its main constituent clades. Five of the identified CSIs are shared by the entire order '*Enterobacteriales*', 21 CSIs are shared by the *Enterobacter-Escherichia* clade, 12 CSIs are shared by the *Erwinia-Pantoea* clade, four CSIs are shared by the *Pectobacterium-Dickeya* clade, three CSIs are shared by the *Yersinia-Serratia* clade, four CSIs are shared by the *Hafnia-Edwardsiella* clade, seven CSIs are shared by the *Proteus-Xenorhabdus* clade, andnine CSIs are shared by the *Budvicia* clade. Each of these CSIs provide independent support for the branching and the groupings of

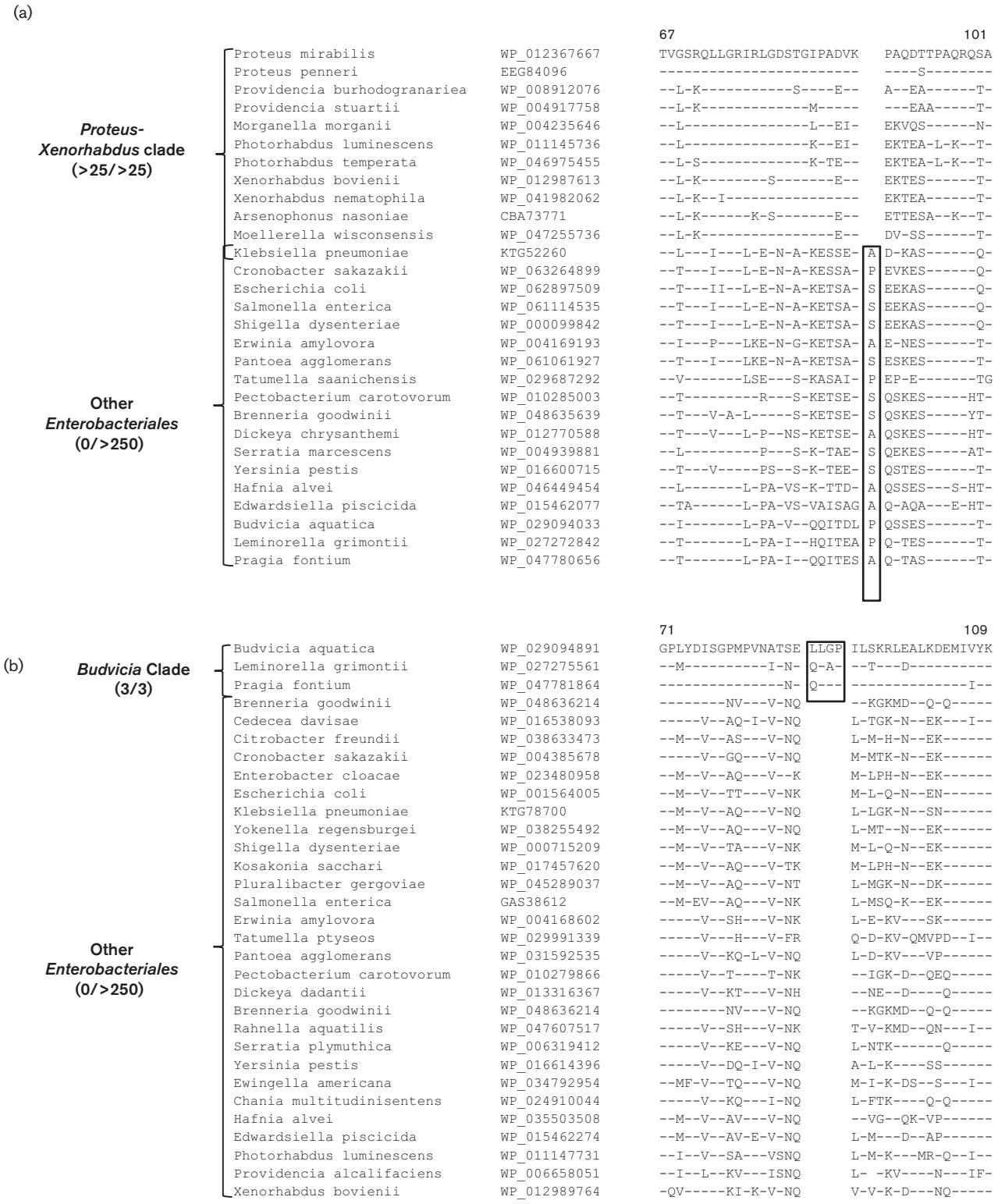

Fig. 7. Partial sequence alignments of (a) the protein Xaa-Pro dipeptidase containing single amino acid deletion (boxed) that is exclusively found in all members within the *Proteus-Xenorhabdus* clade, and (b) the protein bifunctional protein-disulfide isomerise/oxidoreductase DsbC containing a four amino acid insert (boxed) exclusive to members within the *Budvicia* clade. Additional CSIs specific for these clades are presented in Figs S60–S66 for the *Proteus-Xenorhabdus* clade and Table 3 and Figs S67–S75 for the *Budvicia* clade, and their characteristics are briefly described in Table 4.

genera seen in the genome-, ribosomal protein- and MLSA-based phylogenetic trees produced in this work. Additionally, it is now possible to differentiate these groups of genera from each other and all other bacteria on the basis of the presence or absence of these unique CSIs either *in silico* or utilizing PCR-based assays (Ahmod *et al.*, 2011; Wong *et al.*, 2014).

The single constituent family within the order '*Enterobacteriales*' contains over 60 genera and 250 species, making the family *Enterobacteriaceae* one of the most taxonomically diverse bacterial families (www.namesforlife.com; Parte, 2014). The analyses completed in this study, including phylogenetic reconstructions based on 1548 core proteins, 53 ribosomal proteins and four multilocus sequence analysis (MLSA) proteins, analysis of overall genome similarity, and the identification of shared CSIs, strongly support the existence of at least seven main groups within the order '*Enterobacteriales*'. A division of the family *Enterobacteriaceae* into additional family-level taxa would provide a more coherent taxonomic framework for the order '*Enterobacteriales*' that more accurately reflects the interrelationships of the various groups of genera within the order. Additionally, a new taxonomic framework for the order '*Enterobacteriales*' would guide future taxonomic revisions and play a significant role in reducing the prevalence of polyphyletic genera within the order (Brady *et al.*, 2013; Brenner & Farmer III, 2005; Octavia & Lan, 2014). Thus, on the basis of the phylogenetic analyses and utilizing numerous identified conserved molecular characteristics described here, we propose a division of the order '*Enterobacteriales*' into seven families: an emended family *Enterobacteriaceae* (the *Enterobacter-Escherichia* clade), *Erwiniaceae* fam. nov. (the *Erwinia-Pantoea* clade), *Pectobacteriaceae* fam. nov. (the *Pectobacterium-Dickeya* clade), *Yersiniaceae* fam. nov. (the *Yersinia-Serratia* clade), *Hafniaceae* fam. nov. (the *Hafnia-Edwardsiella* clade), *Morganellaceae* fam. nov. (the *Proteus-Xenorhabdus* clade), and *Budviciaceae* fam. nov. (the *Budvicia* clade). Genera which are not sequenced (viz. *Biostraticola, Cosenzaea, Enterobacillus, Gibbsiella, Pseudocitrobacter, Rosenbergiella, Saccharobacter* and *Samsonia*) are placed into one of the families based on 16S rRNA gene sequence identity (Table S5). The branching affinity of the genera *Buchnera* and *Wigglesworthia* within the order '*Enterobacteriales*' has been difficult to resolve in past studies (Herbeck *et al.*, 2005; Husník *et al.*, 2011; Lerat *et al.*, 2003; Williams *et al.*, 2010). Here, we have observed that the genera *Buchnera* and *Wigglesworthia* branch between the *Enterobacter-Escherichia* and the *Erwinia-Pantoea* clades in both the genome- and ribosomal protein-based phylogenetic trees. Furthermore, the genus *Buchnera* shares five CSIs with either the *Erwinia-Pantoea* clade or both the *Enterobacter-Escherichia* and the *Erwinia-Pantoea* clades, while the genus *Wigglesworthia* shares a single CSI with both the *Enterobacter-Escherichia* and the *Erwinia-Pantoea* clades. These findings provide strong suggestive evidence of a specific relationship between the genus *Buchnera* and the *Erwinia-Pantoea* clade and evidence for an association between the genus *Wigglesworthia* and both the *Enterobacter-Escherichia* and the *Erwinia-Pantoea* clades. Thus, at present, the genera *Buchnera* and *Wigglesworthia* will be assigned to the *Erwinia-Pantoea* clade (*Erwiniaceae* fam. nov.). The genus *Plesiomonas* is difficult to place in any of the described families based on phylogeny, CSIs, and 16S rRNA gene sequence identity. Additionally, the homologues of the CSI-containing proteins, specific for all '*Enterobacteriales*', which were found in the genus *Plesiomonas* did not contain the CSIs shared by all other members of the order '*Enterobacteriales*'. Further, the genus *Plesiomonas* was found to consistently branch either earlier than all other members of the '*Enterobacteriales*' or with greater affinity to other orders within the class *Gammaproteobacteria* in phylogenetic trees. These results suggest that the genus *Plesiomonas* has limited association with other members of the order '*Enterobacteriales*' and it may not belong in the order at all. Thus, the genus *Plesiomonas* will not assigned to any family within the order '*Enterobacteriales*', at present, and will be considered family *incertae sedis*. A summary of the taxonomic revisions proposed here is available in Fig. 8 and descriptions of the new and emended taxa are provided below.

Nomenclature of the order '*Enterobacteriales*'

The name of the order '*Enterobacteriales*' has never been validly published in accordance to the rules of the *International Code of Nomenclature of Bacteria* (Lapage *et al.*, 1992). The latest edition of *Bergey's Manual of Systematic Bacteriology* lists the type genus of the order '*Enterobacteriales*' as *Escherichia*, which is the same as the type genus of the family *Enterobacteriaceae* (Imhoff, 2005). However, the name *Enterobacteriaceae* predates the *International Code of Nomenclature of Bacteria* and its original derivation is uncertain (Judicial Commission of the International Committee on Systematic Bacteriology, 1981). The name *Enterobacteriaceae* was validated by the Judicial Commission of the International Committee on Systematic Bacteriology with the type genus *Escherichia* for historical reasons, despite this nomenclature not being in accordance to the rules of the *International Code of Nomenclature of Bacteria* (Brenner, 1983; Wayne, 1982). Thus, an order with the type genus *Escherichia* should be named '*Escherichiales*', not '*Enterobacteriales*', according to the rules of the *International Code of Nomenclature of Bacteria* (Lapage *et al.*, 1992). Furthermore, an order with the type genus *Enterobacter* should be named '*Enterobacterales*' not '*Enterobacteriales*'. To limit the confusion regarding the nomenclature of the '*Enterobacteriales*' which could arise if the name '*Escherichiales*' were to be used to describe the order, we have chosen to utilize the name *Enterobacterales* ord. nov. with the type genus *Enterobacter* to describe the order containing the family *Enterobacteriaceae*.

Description of the order *Enterobacterales* ord. nov.

Enterobacterales (En.te.ro.bac.te.ra'les. N.L. n. *Enterobacter* the type genus of the order; *-ales* ending to denote an order; N.L. fem. pl. n. *Enterobacterales* the order whose nomenclatural type is the genus *Enterobacter*).

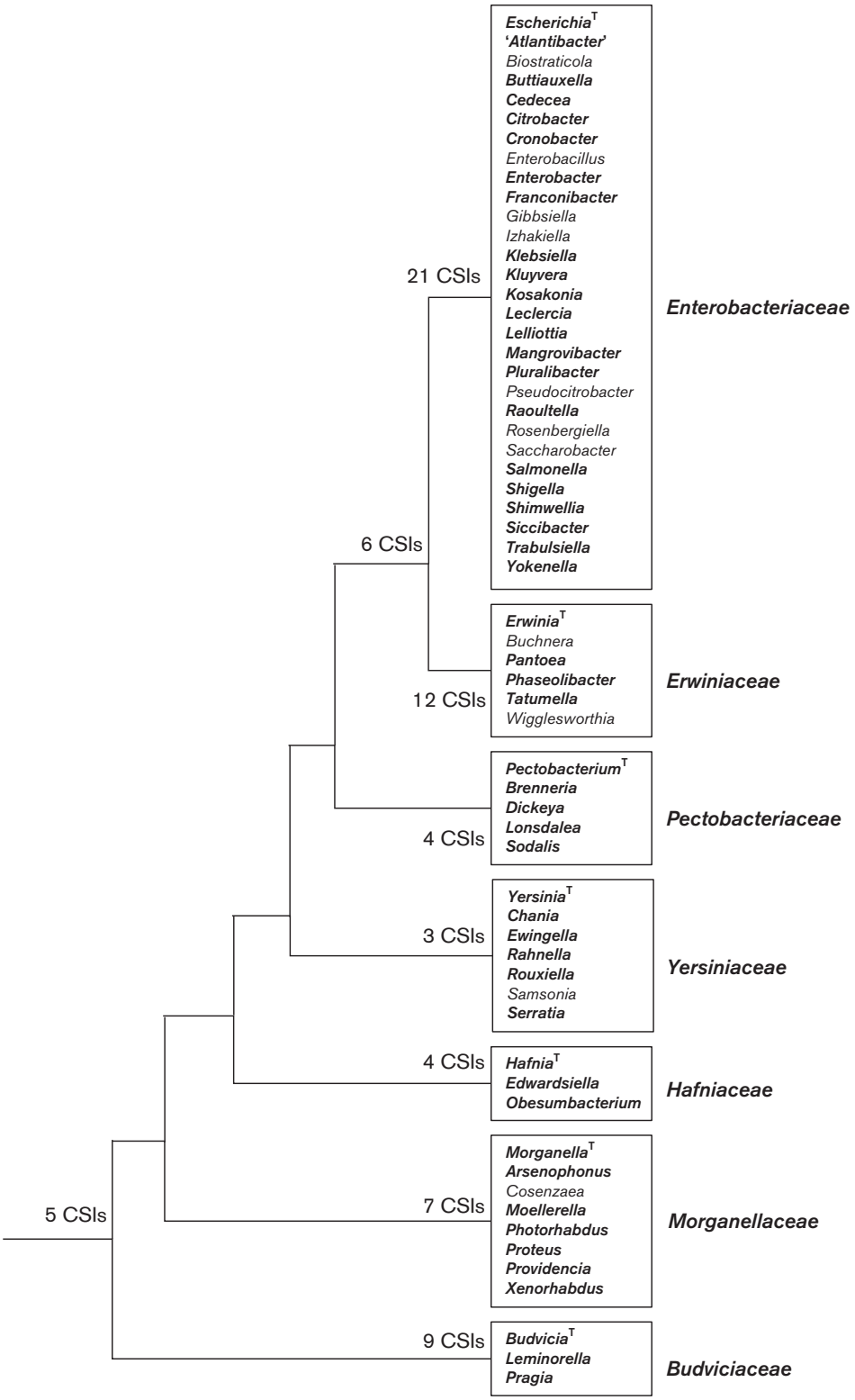

Fig. 8. A summary diagram depicting the distribution of identified CSIs within the order '*Enterobacteriales*' (synonym: *Enterobacterales* ord. nov.) and the proposed families described in this study. Genera which have had their genomes analysed in this study are indicated in bold type. The superscript letter T beside a genus indicates that it is the type genus of the family.

The *Enterobacterales* are an order of Gram-negative, non-spore forming, rod-shaped, facultative anaerobes. The order contains the type genus *Enterobacter* (Rahn, 1937) as well as the families *Enterobacteriaceae* (Rahn, 1937), *Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov. and *Budviciaceae* fam. nov. The description of the order is the same as that of the family *Enterobacteriaceae* given by Brenner & Farmer III (2005) with the following modifications: the members of the order *Enterobacterales* can be distinguished from all other bacteria by the five conserved signature indels in the proteins peptide ABC transporter permease, elongation factor P-like protein YeiP, L-arabinose isomerase, pyrophosphatase, and a hypothetical protein (Table 1).

The type genus is *Enterobacter*.

Emended description of the family *Enterobacteriaceae* (Approved Lists 1980)

The family *Enterobacteriaceae* contains the type genus *Escherichia* (Castellani & Chambers, 1919; Lapage *et al.*, 1992) and the genera '*Atlantibacter*' (Hata *et al.*, 2016), *Biostraticola* (Verbarg *et al.*, 2008), *Buttiauxella* (Ferragut *et al.*, 1981), *Cedecea* (Grimont *et al.*, 1981), *Citrobacter* (Werkman & Gillen, 1932), *Cronobacter* (Iversen *et al.*, 2008), *Enterobacillus* (Patil *et al.*, 2015), *Enterobacter* (Rahn, 1937), *Franconibacter* (Stephan *et al.*, 2014), *Gibbsiella* (Brady *et al.*, 2010a), *Izhakiella* (Aizenberg-Gershtein *et al.*, 2016), *Klebsiella* (Drancourt *et al.*, 2001), *Kluyvera* (Farmer *et al.*, 1981), *Kosakonia* (Brady *et al.*, 2013), *Leclercia* (Tamura *et al.*, 1986), *Lelliottia* (Brady *et al.*, 2013), *Mangrovibacter* (Rameshkumar *et al.*, 2010), *Pluralibacter* (Brady *et al.*, 2013), *Pseudocitrobacter* (Kämpfer *et al.*, 2014), *Raoultella* (Drancourt *et al.*, 2001), *Rosenbergiella* (Halpern *et al.*, 2013a), *Saccharobacter* (Yaping *et al.*, 1990), *Salmonella* (Lignieres, 1900), *Shigella* (Castellani & Chambers, 1919), *Shimwellia* (Priest & Barker, 2010), *Siccibacter* (Stephan *et al.*, 2014), *Trabulsiella* (McWhorter *et al.*, 1991), and *Yokenella* (Kosako *et al.*, 1984). All genera belonging to this group are catalase-positive and oxidase-negative. Members of the family *Enterobacteriaceae* form a distinct monophyletic cluster in genome- and multi-gene-based phylogenetic trees and can be distinguished from all other members of the order *Enterobacterales* by 21 conserved signature indels in the proteins NADH:ubiquinone oxidoreductase (subunit M), twitching motility protein PilT, 2,3-dihydroxybenzoate-AMP ligase, ATP/GTP-binding protein, multifunctional fatty acid oxidation complex (subunit alpha), S-formylglutathione hydrolase, aspartate-semialdehyde dehydrogenase, epimerase, membrane protein, formate dehydrogenylase (subunit 7), glutathione S-transferase, major facilitator superfamily transporter, phosphoglucosamine mutase, glycosyl hydrolase 1 family protein, 23S rrna [uracil(1 939)-C(5)]-methyltransferase, co-chaperone HscB, *N*-acetylmuramoyl-L-alanine amidase, sulfate ABC transporter ATP-binding protein CysA, and LPS assembly protein LptD (Table 2).

Description of *Erwiniaceae* fam. nov.

Erwiniaceae (Er.wi.ni.a.ce′ae. N.L. fem. n. *Erwinia* type genus of the family; *-aceae* ending to denote a family; N.L. fem. pl. n. *Erwiniaceae* the family whose nomenclatural type is the genus *Erwinia*).

The family *Erwiniaceae* contains the type genus *Erwinia* (Hauben *et al.*, 1998) and the genera *Buchnera* (Munson *et al.*, 1991), *Pantoea* (Brady *et al.*, 2010b), *Phaseolibacter* (Halpern *et al.*, 2013b), *Tatumella* (Hollis *et al.*, 1981) and *Wigglesworthia* (Aksoy, 1995). These bacteria are catalase-positive, oxidase-negative, and do not produce indole or hydrogen disulfide. Most species are positive for Voges-Proskauer, with the exception of *Erwinia toletana, Erwinia. typographi* and some strains of *Erwinia oleae.* Members of the family *Erwiniaceae* form a distinct monophyletic cluster in genome- and multi-gene-based phylogenetic trees and can be distinguished fromall other bacteria by 12 conserved signature indels in the proteins glutamate–cysteine ligase, DNA gyrase (subunit B), LPS assembly protein LptD, Thiol: disulfide interchange protein DsbA precursor, two-component sensor histidine kinase, RNA helicase, tRNA pseudouridine(13) synthase TruD, glycine/betaine ABC transporter ATP-binding protein, superoxide dismutase, and stationary phase inducible protein CsiE (Table 3).

The type genus is *Erwinia*.

Description of *Pectobacteriaceae* fam. nov.

Pectobacteriaceae (Pec.to.bac.te.ri.a.ce′ae N.L. neut. n. *Pectobacterium* type genus of the family; *-aceae* ending to denote a family; N.L. fem. pl. n. *Pectobacteriaceae* the family whose nomenclatural type is the genus *Pectobacterium*).

The family *Pectobacteriaceae* contains the type genus *Pectobacterium* (Hauben *et al.*, 1998) and the genera *Brenneria* (Brady *et al.*, 2014a), *Dickeya* (Samson *et al.*, 2005), *Lonsdalea* (Brady *et al.*, 2012), and *Sodalis* (Dale & Maudlin, 1999). Members of the family produce acid from *N*-acetylglucosamine and are negative for arginine dihydrolase, orthinine decarboxylase and lysine decarboxylase. These bacteria are catalase-positive, oxidase-negative, and do not produce hydrogen disulfide. Members of the family *Pectobacteriaceae* form a distinct monophyletic cluster in genome- and multi-gene-based phylogenetic trees and can be distinguished from all other bacteria by four conserved signature indels in the proteins transcriptional activator RhaS, flagellar motor protein MotB, a two-component sensor histidine kinase protein and a hypothetical protein (Table 4).

The type genus is *Pectobacterium*.

Description of *Yersiniaceae* fam. nov.

Yersiniaceae (Yer.si.ni.a.ce′ae. N.L. fem. n. *Yersinia* type genus of the family; *-aceae* ending to denote a family; N.L. fem. pl. n. *Yersiniaceae* the family whose nomenclatural type is the genus *Yersinia*).

The family *Yersiniaceae* contains the type genus *Yersinia* (Van Loghem, 1944) and the genera *Chania* (Ee *et al.*, 2016), *Ewingella* (Grimont *et al.*, 1983), *Rahnella* (Izard *et al.*, 1978), *Rouxiella* (Le Fleche-Mateos *et al.*, 2015), *Samsonia* (Sutra *et al.*, 2001), and *Serratia* (Bizio, 1823). These bacteria are motile, catalase-positive, and do not produce hydrogen disulfide. Members of the family *Yersiniaceae* form a distinct monophyletic cluster in genome -and multi-gene-based phylogenetic trees and can be distinguished from all other bacteria by three conserved signature indels in the protein TetR family transcriptional regulator and a hypothetical protein (Table 4).

The type genus is *Yersinia*.

Description of *Hafniaceae* fam. nov.

Hafniaceae (Haf.ni.a.ce′ae. N.L. fem. n. *Hafnia* type genus of the family; *-aceae* ending to denote a family; N.L. fem. pl. n. *Hafniaceae* the family whose nomenclatural type is the genus *Hafnia*).

The family *Hafniaceae* contains the type genus *Hafnia* (Møller, 1954) and the genera *Edwardsiella* (Ewing *et al.*, 1965) and *Obesumbacterium* (Shimwell, 1963). Members are catalase-positive, oxidase-negative, and negative for lysine decarboxylase. These bacteria are also able to grow on MacConkey media, and are capable of reducing nitrate. Members of the family *Hafniaceae* form a distinct monophyletic cluster in genome- and multi-gene-based phylogenetic trees and can be distinguished from all other bacteria by four conserved signature indels in the proteins two-component system response regulator GIrR, glucose-1-phosphate adenylyltransferase, transcriptional activator NhaR, and the hybrid sensor histidine kinase/response regulator (Table 4).

The type genus is *Hafnia*.

Description of *Morganellaceae* fam. nov.

Morganellaceae (Mor.ga.nel.la.ce′ae. N.L. fem. n. *Morganella* the type genus of the family; *-aceae* ending to denote a family; N.L. fem. pl. n. *Morganellaceae* the family whose nomenclatural type is the genus *Morganella*).

The family *Morganellaceae* contains the type genus *Morganella* (Fulton, 1943) and the genera *Arsenophonus* (Gherna *et al.*, 1991), *Cosenzaea* (Giammanco *et al.*, 2011), *Moellerella* (Hickman-Brenner *et al.*, 1984), *Photorhabdus* (Boemare *et al.*, 1993), *Proteus* (Hauser, 1885), *Providencia* (Ewing, 1962) and *Xenorhabdus* (Thomas & Poinar Jr, 1979). These bacteria are oxidase-negative, and negative for arginine decarboxylase and Voges–Proskauer. Members of the family *Morganellaceae* form a distinct monophyletic cluster in genome- and multi-gene-based phylogenetic trees and can be distinguished from all other bacteria by seven conserved signature indels in the proteins dihydrolipoamide succinyltransferase, Xaa-Pro dipeptidase, bifunctional UDP-sugar hydrolase (5′-nucleotidase), transcriptional repair coupling factor, phosphate acetyltransferase, histidine–tRNA ligase, and *N*-acetylmuramoyl-L-alanine amidase (Table 4).

The type genus is *Morganella*.

Description of *Budviciaceae* fam. nov.

Budviciaceae (Bud.vi.ci.a.ce′ae. L. fem. n. *Budvicia* type genus of the family; *-aceae* ending to denote a family; N.L. fem. pl. n. *Budviciaceae* the family whose nomenclatural type is the genus *Budvicia*).

The family *Budviciaceae* contains the type genus *Budvicia* (Lang *et al.*, 2013) and the genera *Leminorella* (Hickman-Brenner *et al.*, 1985) and *Pragia* (Aldová *et al.*, 1988). Members are catalase-positive, oxidase-negative, and negative for indole, arginine dihydrolase, orthinine decarboxylase, and lysine decarboxylase. These bacteria are capable of producing hydrogen disulfide and reducing nitrate, but are incapable of growing on KCN media. Members of the family *Budviciaceae* form a distinct monophyletic cluster in genome- and multi-gene-based phylogenetic trees and can be distinguished from all other bacteria by nine conserved signature indels in the proteins bifunctional protein-disulfide isomerise/oxidoreductase DsbC, L-methionine/branched chain amino acid transporter, D-alanine-D-alanine ligase, and hypothetical proteins (Table 4).

The type genus is *Budvicia*.

ACKNOWLEDGEMENTS

This work was supported by the research grant no. 249924 from the Natural Science and Engineering Research Council of Canada awarded to R. S. G. We thank Sanam Gurm and Bo 'Jim' Zeng (McMaster University) for their technical assistance in identifying some of the CSIs described in this work. We also thank Dr Brian J. Tindall and Dr Aharon Oren for their expert taxonomic advice regarding the nomenclature of the order *Enterobacterales*.

REFERENCES

Ahmod, N. Z., Gupta, R. S. & Shah, H. N. (2011). Identification of a *Bacillus anthracis* specific indel in the *yeaC* gene and development of a rapid pyrosequencing assay for distinguishing *B. anthracis* from the *B. cereus* group. *J Microbiol Methods* **87**, 278–285.

Aizenberg-Gershtein, Y., Halpern, M., Samuni-Blank, M. & Laviad, S. (2016). *Izhakiella capsodis* gen. nov., sp. nov., in the family *Enterobacteriaceae*, isolated from the mirid bug *Capsodes infuscatus*. *Int J Syst Evol Microbiol* **66**, 1364–1370.

Aksoy, S. (1995). *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. *Int J Syst Evol Microbiol* **45**, 848–851.

Aldová, E., Hausner, O., Brenner, D. J., Kocmoud, Z., Schindler, J., Potužníková, B. & Petráš, P. (1988). *Pragia fontium* gen. nov., sp. nov. of the family *Enterobacteriaceae*, isolated from water. *Int J Syst Evol Microbiol* **38**, 183–189.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Auch, A. F., von Jan, M., Klenk, H. P. & Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* **2**, 117–134.

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics* **21**, 163–193.

Bizio, B. (1823). Lettera di bartolomeo bizio al chiarissimo canonico angelo Bbllani sopra il fenomeno della polenta porporina. *Biblioteca Italiana O Sia Giornale Di Letteratura, Scienze E Arti* **30**, 275–295.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Rode, C. K., Rode, C. K. & other authors (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462.

Boemare, N. & Akhurst, R. (2006). The genera *Photorhabdus Xenorhabdus*. In *The Prokaryotes*, pp. 451–494. New york, NY: Springer.

Boemare, N. E., Akhurst, R. J. & Mourant, R. G. (1993). DNA relatedness between *Xenorhabdus* spp. (*Enterobacteriaceae*), symbiotic bacteria of entomopathogenic nematodes, and a proposal to transfer *Xenorhabdus luminescens* to a new genus, *Photorhabdus* gen. nov. *Int J Syst Bacteriol* **43**, 249–255.

Bonn, W. G. & van der Zwet, T. (2000). Distribution and economic importance of fire blight. In *Fire Blight: the Disease and its Causative Agent, Erwinia Amylovora*, pp. 37–53. Edited by J. L. Vanneste. Wallingford, UK: CABI.

Brady, C., Cleenwerck, I., Venter, S., Vancanneyt, M., Swings, J. & Coutinho, T. (2008). Phylogeny and identification of *Pantoea* species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). *Syst Appl Microbiol* **31**, 447–460.

Brady, C., Denman, S., Kirk, S., Venter, S., Rodríguez-Palenzuela, P. & Coutinho, T. (2010a). Description of *Gibbsiella quercinecans* gen. nov., sp. nov., associated with Acute Oak Decline. *Syst Appl Microbiol* **33**, 444–450.

Brady, C. L., Cleenwerck, I., Venter, S. N., Engelbeen, K., De Vos, P. & Coutinho, T. A. (2010b). Emended description of the genus *Pantoea*, description of four species from human clinical samples, *Pantoea septica* sp. nov., *Pantoea eucrina* sp. nov., *Pantoea brenneri* sp. nov. and *Pantoea conspicua* sp. nov., and transfer of Pectobacterium cypripedii (Hori 1911) Brenner, *et al.* 1973 emend. Hauben, *et al.* 1998 to the genus as P*antoea cypripedii* comb. nov. *Int J Syst Evol Microbiol* **60**, 2430–2440.

Brady, C. L., Cleenwerck, I., Denman, S., Venter, S. N., Rodríguez-Palenzuela, P., Coutinho, T. A. & De Vos, P. (2012). Proposal to reclassify *Brenneria quercina* (Hildebrand and Schroth 1967) Hauben, *et al.* 1999 into a new genus, *Lonsdalea* gen. nov., as *Lonsdalea quercina* comb. nov., descriptions of *Lonsdalea quercina* subsp. quercina comb. nov., *Lonsdalea quercina* subsp. *berica* subsp. nov. and *Lonsdalea quercina* subsp. *britannica* subsp. nov., emendation of the description of the genus *Lonsdalea quercina* subsp. *britannica* subsp. nov., emendation of the description of the genus *Brenneria*, reclassification of *Dickeya dieffenbachiae* as *Dickeya dadantii* subsp. *dieffenbachiae* comb. nov., and emendation of the description of *Dickeya dadantii*. *Int J Syst Evol Microbiol* **62**, 1592–1602.

Brady, C., Cleenwerck, I., Venter, S., Coutinho, T. & De Vos, P. (2013). Taxonomic evaluation of the genus *Enterobacter* based on multilocus sequence analysis (MLSA): proposal to reclassify *E. nimipressuralis* and *E. amnigenus* into *Lelliottia* gen. nov. as *Lelliottia nimipressuralis*comb. nov. and *Lelliottia amnigena* comb. nov., respectively, *E. gergoviae* and *E. pyrinus* into *Pluralibacter* gen. nov. as *Pluralibacter gergoviae* comb. nov. and *Pluralibacter pyrinus* comb. nov., respectively, *E. cowanii, E. radicincitans, E. oryzae and E. arachidis into Kosakonia gen. nov.* as *Kosakonia cowanii* comb. nov., *Kosakonia radicincitans* comb. nov., *Kosakonia oryzae* comb. nov. and *Kosakonia arachidis* comb. nov., respectively, and *E. turicensis, E. helveticus* and *E. pulveris* into *Cronobacter* as *Cronobacter zurichensis nom.* nov., *Cronobacter helveticus* comb. nov. and *Cronobacter pulveris* comb. nov., respectively, and emended description of the genera *Enterobacter* and *Cronobacter*. *Syst Appl Microbiol* **36**, 309–319.

Brady, C., Hunter, G., Kirk, S., Arnold, D. & Denman, S. (2014a). Description of *Brenneria roseae* sp. nov. and two subspecies, *Brenneria roseae* subspecies *roseae* ssp. nov and *Brenneria roseae* subspecies *americana* ssp. nov. isolated from symptomatic oak. *Syst Appl Microbiol* **37**, 396–401.

Brady, C., Hunter, G., Kirk, S., Arnold, D. & Denman, S. (2014b). *Rahnella victoriana* sp. nov., *Rahnella bruchi sp. nov., Rahnella woolbedingensis* sp. nov., classification of *Rahnella* genomospecies 2 and 3 as *Rahnella variigena* sp. nov. and *Rahnella inusitata* sp. nov., respectively and emended description of the genus *Rahnella*. *Syst Appl Microbiol* **37**, 545–552.

Brenner, D. J. (1983). Opposition to the proposal to replace the family name *Enterobacteriaceae*†. *Int J Syst Evol Microbiol* **33**, 892–895.

Brenner, D. J. & Farmer III, J. J. (2005). Family I. *Enterobacteriaceae*. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, vol. 2, pp. 587–607. Edited by D. J. Brenner, N. R. Krieg, J. T. Staley, G. M. Garrity, D. R. Boone, P. Vos, M. Goodfellow, F. A. Rainey & K.-H. Schleifer. New York, NY: Springer.

Brzuszkiewicz, E., Waschkowitz, T., Wiezer, A. & Daniel, R. (2012). Complete genome sequence of the B12-producing *Shimwellia blattae* strain DSM 4481, isolated from a cockroach. *J Bacteriol* **194**, 4436.

Campbell, C., Adeolu, M. & Gupta, R. S. (2015). Genome-based taxonomic framework for the class *Negativicutes*: division of the class *Negativicutes* into the orders *Selenomonadales emend., Acidaminococcales* ord. nov. and *Veillonellales* ord. nov. *Int J Syst Evol Microbiol* **65**, 3203–3215.

Castellani, A. & Chambers, A. J. (1919). *Manual of Tropical Medicine*, 3rd edn. New York: William, Wood and Co.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**, 540–552.

Chaudhuri, R. R. & Henderson, I. R. (2012). The evolution of the *Escherichia coli* phylogeny. *Infect, Genet Evol* **12**, 214–226.

Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B. & Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283–1287.

Clifford, R. J., Hang, J., Riley, M. C., Onmus-Leone, F., Kuschner, R. A., Lesho, E. P. & Waterman, P. E. (2012). Complete genome sequence of *Providencia stuartii* clinical isolate MRSN 2154. *J Bacteriol* **194**, 3736–3737.

Coutinho, T. A. & Venter, S. N. (2009). *Pantoea ananatis*: an unconventional plant pathogen. *Mol Plant Pathol* **10**, 325–335.

Croxen, M. A. & Finlay, B. B. (2010). Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol* **8**, 26–38.

Dagan, T. & Martin, W. (2006). The tree of one percent. *Genome Biol* **7**, 118.

Dale, C. & Maudlin, I. (1999). *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. *Int J Syst Evol Microbiol* **49**, 267–275.

Dauga, C. (2002). Evolution of the *gyrB* gene and the molecular phylogeny of *Enterobacteriaceae*: a model molecule for molecular systematic studies. *Int J Syst Evol Microbiol* **52**, 531–547.

Drancourt, M., Bollet, C., Carta, A. & Rousselier, P. (2001). Phylogenetic analyses of *Klebsiella* species delineate *Klebsiella* and *Raoultella* gen. nov., with description of *Raoultella ornithinolytica* comb. nov., *Raoultella terrigena* comb. nov. and *Raoultella planticola* comb. nov. *Int J Syst Evol Microbiol* **51**, 925–932.

Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**, e1002195.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461.

Ee, R., Madhaiyan, M., Ji, L., Lim, Y. L., Nor, N. M., Tee, K. K., Chen, J. W. & Yin, W. F. (2016). *Chania multitudinisentens* gen. nov., sp. nov., an N-acyl-homoserine-lactone-producing bacterium in the family *Enterobacteriaceae* isolated from landfill site soil. *Int J Syst Evol Microbiol* **66**, 2297–2304.

Eppinger, M., Worsham, P. L., Nikolich, M. P., Riley, D. R., Sebastian, Y., Mou, S., Achtman, M., Lindler, L. E. & Ravel, J. (2010).

Genome sequence of the deep-rooted *Yersinia pestis* strain Angola reveals new insights into the evolution and pangenome of the plague bacterium. *J Bacteriol* **192**, 1685–1699.

Ewing, W. (1962). The tribe Proteae: its nomenclature and taxonomy. *Int J Syst Evol Microbiol* **12**, 93–102.

Ewing, W., McWhorter, A., Escobar, M. & Lubin, A. (1965). *Edwardsiella*, a new genus of *Enterobacteriaceae* based on a new species, *E. tarda*. *Int J Syst Evol Microbiol* **15**, 33–38.

Farmer, J. J., Fanning, G. R., Huntley-Carter, G. P., Holmes, B., Hickman, F. W., Richard, C. & Brenner, D. J. (1981). *Kluyvera*, a new (redefined) genus in the family *Enterobacteriaceae*: identification of *Kluyvera ascorbata* sp. nov. and *Kluyvera cryocrescens* sp. nov. in clinical specimens. *J Clin Microbiol* **13**, 919–933.

Ferragut, C., Izard, D., Gavini, F., Lefebvre, B. & Leclerc, H. (1981). *Buttiauxella*, a new genus of the family *Enterobacteraceae*. *Zentralblatt Für Bakteriologie Mikrobiologie Und Hygiene: I. Abt. Originale C: Allgemeine, Angewandte Und Ökologische Mikrobiologie* **2**, 33–44.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M. & other authors (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279–D285.

Forst, S., Dowds, B., Boemare, N. & Stackebrandt, E. (1997). *Xenorhabdus* and *Photorhabdus* spp.: bugs that kill bugs. *Annu Rev Microbiol* **51**, 47–72.

Francino, M. P., Santos, S. R. & Ochman, H. (2006). Phylogenetic relationships of bacteria with special reference to endosymbionts and enteric species. In *The Prokaryotes*, pp. 41–59. Springer.

Fukushima, M., Kakinuma, K. & Kawaguchi, R. (2002). Phylogenetic analysis of *Salmonella, Shigella*, and *Escherichia coli* strains on the basis of the *gyrB* gene sequence. *J Clin Microbiol* **40**, 2779–2785.

Fulton, M. (1943). The identity of Bacterium Columbensis Castellani. *J Bacteriol* **46**, 79.

Gao, B., Mohan, R. & Gupta, R. S. (2009). Phylogenomics and protein signatures elucidating the evolutionary relationships among the *Gammaproteobacteria*. *Int J Syst Evol Microbiol* **59**, 234–247.

Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., Stackebrandt, E., Van de Peer, Y., Vandamme, P. & other authors (2005). Opinion: re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**, 733–739.

Gherna, R., Werren, J., Weisburg, W., Cote, R., Woeste, C., Mandelco, L. & Brenner, D. (1991). *Arsenophonus nasoniae* gen. nov., sp. nov., the causative agent of the son-killer trait in the parasitic wasp Nasonia vitripennis. *Int J Syst Bacteriol* **41**, 563–565.

Giammanco, G. M., Grimont, P. A., Grimont, F., Lefevre, M., Giammanco, G. & Pignato, S. (2011). Phylogenetic analysis of the genera *Proteus, Morganella and Providencia* by comparison of *rpoB* gene sequences of type and clinical strains suggests the reclassification of *Proteus myxofaciens* in a new genus, *Cosenzaea* gen. nov., as *Cosenzaea myxofaciens* comb. nov. *Int J Syst Evol Microbiol* **61**, 1638–1644.

Glaeser, S. P. & Kämpfer, P. (2015). Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst Appl Microbiol* **38**, 237–245.

Goodrich-Blair, H. & Clarke, D. J. (2007). Mutualism and pathogenesis in *Xenorhabdus* and *Photorhabdus*: two roads to the same destination. *Mol Microbiol* **64**, 260–268.

Gordienko, E. N., Kazanov, M. D. & Gelfand, M. S. (2013). Evolution of pan-genomes of *Escherichia coli, Shigella* spp., and *Salmonella enterica*. *J Bacteriol* **195**, 2786–2792.

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P. & Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**, 81–91.

Grimont, P. A., Grimont, F., Farmer III, J. & Asbury, M. A. (1981). *Cedecea davisae* gen. nov., sp. nov. and *Cedecea lapagei* sp. nov., new *Enterobacteriaceae* from clinical specimens. *Int J Syst Evol Microbiol* **31**, 317–326.

Grimont, P., Farmer, J., Grimont, F., Asbury, M., Brenner, D. & Deval, C. (1983). *Ewingella americana* gen. nov., sp. nov., a new *Enterobacteriaceae* isolated from clinical specimens. In *Annales De l'Institut Pasteur/ Microbiologie*, pp. 39–52 Paris: Elsevier.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307–321.

Gupta, R. S. (2000). The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol Rev* **24**, 367–402.

Gupta, R. S. (2014). Identification of conserved indels that are useful for classification and evolutionary studies. In *Methods in Microbiology*, vol. 41, pp. 153–182. Oxford, UK: Academic Press.

Gupta, R. S. (2016). Impact of genomics on the understanding of microbial evolution and classification: the importance of Darwin's views on classification. *FEMS Microbiol Rev* **40**, 520–553.

Gupta, R. S., Naushad, S. & Baker, S. (2015a). Phylogenomic analyses and molecular signatures for the class *Halobacteria* and its two major clades: a proposal for division of the class *Halobacteria* into an emended order *Halobacteriales* and two new orders, *Haloferacales* ord. nov. and *Natrialbales* ord. nov., containing the novel families *Haloferacaceae* fam. nov. and *Natrialbaceae* fam. nov. *Int J Syst Evol Microbiol* **65**, 1050–1069.

Gupta, R. S., Naushad, S., Chokshi, C., Griffiths, E. & Adeolu, M. (2015b). A phylogenomic and molecular markers based analysis of the phylum Chlamydiae: proposal to divide the class *Chlamydiia* into two orders, *Chlamydiales* and *Parachlamydiales* ord. nov., and emended description of the class *Chlamydiia*. *Antonie van Leeuwenhoek* **108**, 765–781.

Gupta, R. S., Naushad, S., Fabros, R. & Adeolu, M. (2016). A phylogenomic reappraisal of family-level divisions within the class *Halobacteria*: proposal to divide the order *Halobacteriales* into the families *Halobacteriaceae, Haloarculaceae* fam. nov *and Halococcaceae* fam. nov., and the order *Haloferacales* into the families, *Haloferacaceae* and *Halorubraceae* fam nov,. *Antonie van Leeuwenhoek* **109**, 565–587.

Halpern, M., Fridman, S., Atamna-Ismaeel, N. & Izhaki, I. (2013a). *Rosenbergiella nectarea* gen. nov., sp. nov., in the family *Enterobacteriaceae*, isolated from floral nectar. *Int J Syst Evol Microbiol* **63**, 4259–4265.

Halpern, M., Fridman, S., Aizenberg-Gershtein, Y. & Izhaki, I. (2013b). Transfer of *Pseudomonas flectens* Johnson 1956 to *Phaseolibacter* gen. nov., in the family *Enterobacteriaceae*, as *Phaseolibacter flectens* gen. nov., comb. nov. *Int J Syst Evol Microbiol* **63**, 268–273.

Hata, H., Natori, T., Mizuno, T., Kanazawa, I., Eldesouky, I., Hayashi, M., Miyata, M., Fukunaga, H., Ohji, S. & other authors (2016). Phylogenetics of family *Enterobacteriaceae* and proposal to reclassify *Escherichia hermannii* and *Salmonella subterranea* as *Atlantibacter hermannii* and *Atlantibacter subterranea* gen. nov., comb. nov. *Microbiol Immunol* **60**, 303–311.

Hauben, L., Moore, E. R., Vauterin, L., Steenackers, M., Mergaert, J., Verdonck, L. & Swings, J. (1998). Phylogenetic position of phytopathogens within the *Enterobacteriaceae*. *Syst Appl Microbiol* **21**, 384–397.

Hauser, G. (1885). *Über Fäulnissbacterien und deren Beziehungen zur Septicämie*, vol. 1250. FCW Vogel.

Herbeck, J. T., Degnan, P. H. & Wernegreen, J. J. (2005). Nonhomogeneous model of sequence evolution indicates independent origins of primary endosymbionts within the *enterobacteriales* (gamma-Proteobacteria). *Mol Biol Evol* **22**, 520–532.

Hickman-Brenner, F. W., Huntley-Carter, G. P., Saitoh, Y., Steigerwalt, A. G., Farmer, J. J. & Brenner, D. J. (1984). A new genus and species of *Enterobacteriaceae* found in human stool specimens. *J Clin Microbiol* **19**, 460–463.

Hickman-Brenner, F. W., Vohra, M. P., Huntley-Carter, G. P., Fanning, G. R., Lowery, V. A., Brenner, D. J. & Farmer, J. J. (1985). *Leminorella*, a new genus of *Enterobacteriaceae*: identification of *Leminorella grimontii* sp. nov. and *Leminorella richardii* sp. nov. found in clinical specimens. *J Clin Microbiol* **21**, 234–239.

Hollis, D. G., Hickman, F. W., Fanning, G. R., Farmer, J. J., Weaver, R. E. & Brenner, D. J. (1981). *Tatumella ptyseos* gen. nov., sp. nov., a member of the family *Enterobacteriaceae* found in clinical specimens. *J Clin Microbiol* **14**, 79–88.

Husník, F., Chrudimský, T. & Hypša, V. (2011). Multiple origins of endosymbiosis within the *Enterobacteriaceae (γ-Proteobacteria*): convergence of complex phylogenetic approaches. *BMC Biol* **9**, 87.

Huys, G., Cnockaert, M., Abbott, S. L., Janda, J. M. & Vandamme, P. (2010). *Hafnia paralvei* sp. nov., formerly known as *Hafnia alvei* hybridization group 2. *Int J Syst Evol Microbiol* **60**, 1725–1728.

Imhoff, J. F. (2005). Order XIII. 'Enterobacteriales'. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, vol. 2, pp. 587. Edited by D. J. Brenner, N. R. Krieg, J. T. Staley, G. M. Garrity, D. R. Boone, P. Vos, M. Goodfellow, F. A. Rainey & K.-H. Schleifer. New York, NY: Springer.

Iversen, C., Mullane, N., McCardell, B., Tall, B. D., Lehner, A., Fanning, S., Stephan, R. & Joosten, H. (2008). *Cronobacter* gen. nov., a new genus to accommodate the biogroups of *Enterobacter sakazakii*, and proposal of *Cronobacter sakazakii* gen. nov., comb. nov., *Cronobacter malonaticus* sp. nov., *Cronobacter turicensis* sp. nov., *Cronobacter muytjensii* sp. nov., *Cronobacter dublinensis* sp. nov., *Cronobacter* genomospecies 1, and of three subspecies, *Cronobacter dublinensis* subsp. *dublinensis* subsp. nov., *Cronobacter dublinensis* subsp. *lausannensis* subsp. nov. and *Cronobacter dublinensis* subsp. *lactaridi* subsp. nov. *Int J Syst Evol Microbiol* **58**, 1442–1447.

Izard, D., Gavini, F., Trinel, P. & Leclere, H. (1978). *Rahnella aquatilis*, a new member of the *Enterobacteriaceae*. *Ann Microbiol* **130**, 163–177.

Janda, J. (2005). Genus XXVII. *Plesiomonas*. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, vol. 2, pp. 740–744. Edited by D. J. Brenner, N. R. Krieg, G. M. Garrity & J. T. Staley. New York: Springer.

Janda, J. M. (2006). New members of the family *Enterobacteriaceae* . In *The Prokaryotes*, pp. 5–40. New York, NY: Springer.

Janda, J. M. & Abbott, S. L. (1993). Infections associated with the genus *Edwardsiella*: the role of *Edwardsiella tarda* in human disease. *Clin Infect Dis* **17**, 742–748.

Janda, J. M. & Abbott, S. L. (2006). The genus *Hafnia*: from soup to nuts. *Clin Microbiol Rev* **19**–28.

Janda, J. M., Abbott, S. L. & McIver, C. J. (2016). *Plesiomonas shigelloides* Revisited. *Clin Microbiol Rev* **29**, 349–374.

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998). Multiple sequence alignment with CLUSTAL X. *Trends Biochem Sci* **23**, 403.

Jones, A. L. (2012). The future of taxonomy. In *Adv Appl Microbiol*, 1st edn, vol. 80, pp. 23–35. Edited by G. M. Gadd & S. Sariaslani. San Diego: Academic Press Inc.

Judicial Commission of the International Committee on Systematic Bacteriology (1981). Present Standing of the Family Name *Enterobacteriaceae* Rahn 1937. *Int J Syst Bacteriol* **31**, 104.

Kämpfer, P., Glaeser, S. P., Raza, M. W., Abbasi, S. A. & Perry, J. D. (2014). *Pseudocitrobacter* gen. nov., a novel genus of the *Enterobacteriaceae* with two new species *Pseudocitrobacter faecalis* sp. nov., and *Pseudocitrobacter anthropi* sp. nov, isolated from fecal samples from hospitalized patients in Pakistan. *Syst Appl Microbiol* **37**, 17–22.

Koivula, T. T., Juvonen, R., Haikara, A. & Suihko, M. L. (2006). Characterization of the brewery spoilage bacterium *Obesumbacterium proteus* by automated ribotyping and development of PCR methods for its biotype 1. *J Appl Microbiol* **100**, 398–406.

Konstantinidis, K. T. & Tiedje, J. M. (2005). Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* **187**, 6258–6264.

Kosako, Y., Sakazaki, R. & Yoshizaki, E. (1984). *Yokenella regensburgei* gen. nov., sp. nov.: a new genus and species in the family *Enterobacteriaceae*. *Jpn J Med Sci Biol* **37**, 117–124.

Lang, E., Schumann, P., Knapp, B. A., Kumar, R., Spröer, C. & Insam, H. (2013). *Budvicia diplopodorum* sp. nov. and emended description of the genus *Budvicia*. *Int J Syst Evol Microbiol* **63**, 260–267.

Lapage, S. P., Sneath, P. H. A., Lessel, E. F., Skerman, V. B. D., Seeliger, H. P. R. & Clark, W. A. (1992). *International Code of Nomenclature of Bacteria: Bacteriological Code, 1990 Revision*. Washington, DC: ASM Press International Union of Microbiological Societies.

Le Flèche-Matéos, A., Levast, M., Lomprez, F., Arnoux, Y., Andonian, C., Perraud, M., Vincent, V., Ar Gouilh, M., Thiberge, J. M. & other authors (2015). *Rouxiella chamberiensis* gen. nov., sp. nov., a member of the family *Enterobacteriaceae* isolated from parenteral nutrition bags. *Int J Syst Evol Microbiol* **65**, 1812–1818.

Le, S. Q. & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol Biol Evol* **25**, 1307–1320.

Lerat, E., Daubin, V. & Moran, N. A. (2003). From gene trees to organismal phylogeny in prokaryotes: the case of the Proteobacteria. *PLoS Biol* **1**, e19.

Lignieres, J. (1900). Maladies du porc. *Bulletin of the Society for Central Medical Veterinarians* **18**, 389–431.

Livermore, D. M. (2012). Current epidemiology and growing resistance of gram-negative pathogens. *Korean J Intern Med* **27**, 128–142.

Ma, B., Hibbing, M. E., Kim, H. S., Reedy, R. M., Yedidia, I., Breuer, J., Breuer, J., Glasner, J. D., Perna, N. T. & other authors (2007). Host range and molecular phylogenies of the soft rot enterobacterial genera *Pectobacterium* and dickeya. *Phytopathology* **97**, 1150–1163.

McWhorter, A. C., Haddock, R. L., Nocon, F. A., Steigerwalt, A. G., Brenner, D. J., Aleksić, S., Bockemühl, J. & Farmer, J. J. (1991). *Trabulsiella guamensis*, a new genus and species of the family *Enterobacteriaceae* that resembles *Salmonella* subgroups 4 and 5. *J Clin Microbiol* **29**, 1480–1485.

Møller, V. (1954). Distribution of amino acid decarboxylases in *Enterobacteriaceae*. *Acta Pathol Microbiol Scand* **35**, 259.

Morelli, G., Song, Y., Mazzoni, C. J., Eppinger, M., Roumagnac, P., Wagner, D. M., Feldkamp, M., Kusecek, B., Vogler, A. J. & other authors (2010). *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet* **42**, 1140–1143.

Munson, M. A., Baumann, P. & Kinsey, M. G. (1991). *Buchnera* gen. nov. and *Buchnera aphidicola* sp. nov., a taxon consisting of the mycetocyte-associated, primary endosymbionts of aphids. *Int J Syst Evol Microbiol* **41**, 566–568.

Nataro, J. P., Bopp, C. A., Fields, P. I., Kaper, J. B. & Strockbine, N. A. (2011). *Escherichia, Shigella*, and *Salmonella*. In *Manual of Clinical Microbiology*, 10th edn, pp. 603–626. Edited by J. Versalovic, K. C. Carroll, G. Funke, J. H. Jorgensen, M. L. Landry & D. W. Warnock. American Society of Microbiology.

Naum, M., Brown, E. W. & Mason-Gamer, R. J. (2008). Is 16S rDNA a reliable phylogenetic marker to characterize relationships below the family level in the *enterobacteriaceae? J Mol Evol* **66**, 630–642.

Naum, M., Brown, E. W. & Mason-Gamer, R. J. (2011). Is a robust phylogeny of the enterobacterial plant pathogens attainable? *Cladistics* **27**, 80–93.

Naushad, H. S. & Gupta, R. S. (2012). Molecular signatures (conserved indels) in protein sequences that are specific for the order *Pasteurellales* and distinguish two of its main clades. *Antonie van Leeuwenhoek* **101**, 105–124.

Naushad, H. S. & Gupta, R. S. (2013). Phylogenomics and molecular signatures for species from the plant pathogen-containing order xanthomonadales. *PLoS One* **8**, e55216.

Naushad, H. S., Lee, B. & Gupta, R. S. (2014). Conserved signature indels and signature proteins as novel tools for understanding microbial phylogeny and systematics: identification of molecular signatures that are specific for the phytopathogenic genera *Dickeya, Pectobacterium* and *Brenneria*. *Int J Syst Evol Microbiol* **64**, 366–383.

Naushad, S., Adeolu, M., Goel, N., Khadka, B., Al-Dahwi, A. & Gupta, R. S. (2015a). Phylogenomic and molecular demarcation of the core members of the polyphyletic genera *Actinobacillus, Haemophilus Pasteurella*. *Int J Genomics* **2015**, 198560.

Naushad, S., Adeolu, M., Wong, S., Sohail, M., Schellhorn, H. E. & Gupta, R. S. (2015b). A phylogenomic and molecular marker based taxonomic framework for the order *Xanthomonadales*: proposal to transfer the families *Algiphilaceae* and *Solimonadaceae* to the order *Nevskiales* ord. nov. and to create a new family within the order *Xanthomonadales*, the family *Rhodanobacteraceae* fam. nov., containing the genus *Rhodanobacter* and its closest relatives. *Antonie van Leeuwenhoek* **107**, 467–485.

Nielsen-LeRoux, C., Gaudriault, S., Ramarao, N., Lereclus, D. & Givaudan, A. (2012). How the insect pathogen bacteria *Bacillus thuringiensis* and *Xenorhabdus/Photorhabdus* occupy their hosts. *Curr Opin Microbiol* **15**, 220–231.

Octavia, S. & Lan, R. (2014). *The Family Enterobacteriaceae*, pp. 225–286. The Prokaryotes: Gammaproteobacteria.

Paradis, S., Boissinot, M., Paquette, N., Bélanger, S. D., Martel, E. A., Boudreau, D. K., Picard, F. J., Ouellette, M., Roy, P. H. & other authors (2005). Phylogeny of the *Enterobacteriaceae* based on genes encoding elongation factor Tu and F-ATPase beta-subunit. *Int J Syst Evol Microbiol* **55**, 2013–2025.

Parkhill, J., Wren, B. W., Thomson, N. R., Titball, R. W., Holden, M. T., Prentice, M. B., Sebaihia, M., James, K. D., Churcher, C. & other authors (2001). Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523–527.

Parte, A. C. (2014). LPSN – list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res* **42**, D613–D616.

Patil, V. S., Salunkhe, R. C., Patil, R. H., Husseneder, C., Shouche, Y. S. & Ramana, V. V. (2015). *Enterobacillus tribolii* gen. nov., sp. nov., a novel member of the family *Enterobacteriaceae*, isolated from the gut of a red flour beetle, *Tribolium castaneum*. *Antonie van Leeuwenhoek* **107**, 1207–1216.

Perry, R. D. & Fetherston, J. D. (1997). *Yersinia pestis* – etiologic agent of plague. *Clin Microbiol Rev* **10**, 35–66.

Pham, H. N., Ohkusu, K., Mishima, N., Noda, M., Monir Shah, M., Sun, X., Hayashi, M. & Ezaki, T. & Shah, M. M. (2007). Phylogeny and species identification of the family *Enterobacteriaceae* based on *dnaJ* sequences. *Diagn Microbiol Infect Dis* **58**, 153–161.

Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N. & Delsuc, F. (2005). Heterotachy and long-branch attraction in phylogenetics. *BMC Evol Biol* **5**.

Price, M. N., Dehal, P. S. & Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490.

Priest, F. G. & Barker, M. (2010). Gram-negative bacteria associated with brewery yeasts: reclassification of biogroup 2 as *Shimwellia pseudoproteus* gen. nov., sp. nov., and transfer of *Escherichia blattae* to *Shimwellia blattae* comb. nov. *Int J Syst Evol Microbiol* **60**, 828–833.

Pritchard, L., Humphris, S., Saddler, G. S., Elphinstone, J. G., Pirhonen, M. & Toth, I. K. (2013). Draft genome sequences of 17 isolates of the plant pathogenic bacterium dickeya. *Genome Announc* **1**, e00978–00913.

Qin, Q. L., Xie, B. B., Zhang, X. Y., Chen, X. L., Zhou, B. C., Zhou, J., Oren, A. & Zhang, Y. Z. (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol* **196**, 2210–2215.

Rahn, O. (1937). New principles for the classification of bacteria. *Zentralbl Bakteriol Parasitenkd Infektionskr Hyg* **96**, 273–286.

Rameshkumar, N., Lang, E. & Nair, S. (2010). *Mangrovibacter plantisponsor* gen. nov., sp. nov., a nitrogen-fixing bacterium isolated from a mangrove-associated wild rice (*Porteresia coarctata* Tateoka). *Int J Syst Evol Microbiol* **60**, 179–186.

Roggenkamp, A. (2007). Phylogenetic analysis of enteric species of the family *Enterobacteriaceae* using the *oriC*-locus. *Syst Appl Microbiol* **30**, 180–188.

Rokas, A. & Holland, P. W. (2000). Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol* **15**, 454–459.

Rokas, A., Williams, B. L., King, N. & Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798–804.

Rosselló-Mora, R. (2006). DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In *Molecular Identification, Systematics, and Population Structure of Prokaryotes*, pp. 23–50. Edited by E. Stackebrandt. Springer.

Ruimy, R., Breittmayer, V., Elbaze, P., Lafay, B., Boussemart, O., Gauthier, M. & Christen, R. (1994). Phylogenetic analysis and assessment of the genera *Vibrio, Photobacterium, Aeromonas*, and *Plesiomonas* deduced from small-subunit rRNA sequences. *Int J Syst Evol Microbiol* **44**, 416–426.

Salerno, A., Delétoile, A., Lefevre, M., Ciznar, I., Krovacek, K., Grimont, P. & Brisse, S. (2007). Recombining population structure of *Plesiomonas shigelloides* (*Enterobacteriaceae*) revealed by multilocus sequence typing. *J Bacteriol* **189**, 7808–7818.

Samson, R., Legendre, J. B., Christen, R., Fischer-Le Saux, M., Achouak, W. & Gardan, L. (2005). Transfer of *Pectobacterium chrysanthemi* (Burkholder et al. 1953) Brenner et al. 1973 and *Brenneria paradisiaca* to the genus *Dickeya* gen. nov. as *Dickeya chrysanthemi* comb. nov. and *Dickeya paradisiaca* comb. nov. and delineation of four novel species, *Dickeya dadantii* sp. nov., *Dickeya dianthicola* sp. nov., *Dickeya dieffenbachiae* sp. nov. and *Dickeya zeae* sp. nov. *Int J Syst Evol Microbiol* **55**, 1415–1427.

Samuel, G., Hogbin, J. P., Wang, L. & Reeves, P. R. (2004). Relationships of the *Escherichia coli* O157, O111, and O55 O-antigen gene clusters with those of *Salmonella enterica* and *Citrobacter freundii*, which express identical O antigens. *J Bacteriol* **186**, 6536–6543.

Sawana, A., Adeolu, M. & Gupta, R. S. (2014). Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: proposal for division of this genus into the emended genus *Burkholderia* containing pathogenic organisms and a new genus *Paraburkholderia* gen. nov. harboring environmental species. *Front Genet* **5**, 429.

Schindler, J., Potuznikova, B. & Aldová, E. (1991). Classification of strains of *Pragia fontium, Budvicia aquatica* and of *Leminorella* by whole-cell protein pattern. *J Hyg Epidemiol Microbiol Immunol* **36**, 207–216.

Shimwell, J. (1963). *Obesumbacterium* gen. nov. *Brewers' J* **99**, 759–760.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M. & other authors (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.

Snopková, K., Sedlář, K., Bosák, J., Chaloupková, E., Provazník, I. & Šmajs, D. (2015). Complete genome sequence of 24613, an environmental bacterium from the *family Enterobacteriaceae*. *Genome Announc* **3**, e00740–00715.

Spröer, C., Mendrock, U., Swiderski, J., Lang, E. & Stackebrandt, E. (1999). The phylogenetic position of *Serratia, Buttiauxella* and some other genera of the family *Enterobacteriaceae*. *Int J Syst Evol Microbiol* **49**, 1433–1438.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.

Stephan, R., Grim, C. J., Gopinath, G. R., Mammel, M. K., Sathyamoorthy, V., Trach, L. H., Chase, H. R., Fanning, S. & Tall, B. D. (2014). Re-examination of the taxonomic status of *Enterobacter helveticus, Enterobacter pulveris* and *Enterobacter turicensis* as members of the genus

Cronobacter and their reclassification in the genera *Franconibacter* gen. nov. and *Siccibacter* gen. nov. as *Franconibacter helveticus* comb. nov., *Franconibacter pulveris* comb. nov. and *Siccibacter turicensis* comb. nov. respectively. *Int J Syst Evol Microbiol* **64**, 3402–3410.

Sutra, L., Christen, R., Bollet, C., Simoneau, P. & Gardan, L. (2001). *Samsonia erythrinae* gen. nov., sp. nov., isolated from bark necrotic lesions of *Erythrina* sp., and discrimination of plant-pathogenic *Enterobacteriaceae* by phenotypic features. *Int J Syst Evol Microbiol* **51**, 1291–1304.

Tailliez, P., Laroui, C., Ginibre, N., Paule, A., Pagès, S. & Boemare, N. (2010). Phylogeny of *Photorhabdus* and *Xenorhabdus* based on universally conserved protein-coding sequences and implications for the taxonomy of these two genera. Proposal of new taxa: *X. vietnamensis* sp. nov., *P. luminescens* subsp. *caribbeanensis* subsp. nov., *P. luminescens* subsp. *hainanensis* subsp. nov., *P. temperata* subsp. *khanii* subsp. nov., *P. temperata* subsp. *tasmaniensis* subsp. nov., and the reclassification of *P. luminescens* subsp. *thracensis* as *P. temperata* subsp. *thracensis* comb. nov. *Int J Syst Evol Microbiol* **60**, 1921–1937.

Talavera, G. & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**, 564–577.

Tamura, K., Sakazaki, R., Kosako, Y. & Yoshizaki, E. (1986). *Leclercia adecarboxylata* gen. nov., comb. nov., formerly known as *Escherichia adecarboxylata*. *Curr Microbiol* **13**, 179–184.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**, 2725–2729.

Thomas, G. M. & Poinar Jr, G. O. (1979). *Xenorhabdus* gen. nov., a genus of entomopathogenic, nematophilic bacteria of the family *Enterobacteriaceae*. *Int J Syst Evol Microbiol* **29**, 352–360.

Toh, H., Weiss, B. L., Perkin, S. A., Yamashita, A., Oshima, K., Hattori, M. & Aksoy, S. (2006). Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res* **16**, 149–156.

Trowbridge, R. E., Dittmar, K. & Whiting, M. F. (2006). Identification and phylogenetic analysis of - and *Photorhabdus*-type bacteria from adult Hippoboscidae and Streblidae (Hippoboscoidea). *J Invertebr Pathol* **91**, 64–68.

Tyler, H. L. & Triplett, E. W. (2008). Plants as a habitat for beneficial and/or human pathogenic bacteria. *Annu Rev Phytopathol* **46**, 53–73.

UniProt Consortium (2015). UniProt: a hub for protein information. *Nucleic Acids Res* **43**, D204–212.

Van Loghem, J. (1944). The classification of the plague-bacillus. *Antonie van Leeuwenhoek* **10**, 15–16.

Varghese, N. J., Mukherjee, S., Ivanova, N., Konstantinidis, K. T., Mavrommatis, K., Kyrpides, N. C. & Pati, A. (2015). Microbial species delineation using whole genome sequences. *Nucleic Acids Res* **43**, 6761–6771.

Verbarg, S., Frühling, A., Cousin, S., Brambilla, E., Gronow, S., Lünsdorf, H. & Stackebrandt, E. (2008). *Biostraticola tofi* gen. nov., spec. nov., a novel member of the family *Enterobacteriaceae*. *Curr Microbiol* **56**, 603–608.

Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., Gillespie, J. J., Gough, R., Hix, D. & other authors (2014). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* **42**, D581–591.

Wayne, L. G. (1982). Actions of the Judicial Commission of the International Committee on Systematic Bacteriology on requests for opinions published between July 1979 and April 1981. *Int J Syst Evol Microbiol* **32**, 464–465.

Werkman, C. H. & Gillen, G. F. (1932). Bacteria producing trimethylene glycol. *J Bacteriol* **23**, 167.

Whelan, S. & Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* **18**, 691–699.

Williams, K. P., Gillespie, J. J., Sobral, B. W., Nordberg, E. K., Snyder, E. E., Shallom, J. M. & Dickerman, A. W. (2010). Phylogeny of *Gammaproteobacteria*. *J Bacteriol* **192**, 2305–2314.

Wong, S. Y., Paschos, A., Gupta, R. S. & Schellhorn, H. E. (2014). Insertion/deletion-based approach for the detection of *Escherichia coli* O157:H7 in freshwater environments. *Environ Sci Technol* **48**, 11462–11470.

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., Kunin, V., Goodwin, L., Wu, M. & other authors (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462**, 1056–1060.

Yaping, J., Xiaoyang, L. & Jiaqi, Y. (1990). *Saccharobacter fermentatus* gen. nov., sp. nov., a new ethanol-producing bacterium. *Int J Syst Evol Microbiol* **40**, 412–414.

Yarza, P., Richter, M., Peplies, J., Euzeby, J., Amann, R., Schleifer, K. H., Ludwig, W., Glöckner, F. O. & Rosselló-Móra, R. (2008). The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* **31**, 241–250.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W. & other authors (2014). The SILVA and All-species Living Tree Project (LTP) taxonomic frameworks. *Nucleic Acids Res* **42**, D643–648.

Young, J. M. & Park, D. C. (2007). Relationships of plant pathogenic enterobacteria based on partial *atpD, carA*, and *recA* as individual and concatenated nucleotide and peptide sequences. *Syst Appl Microbiol* **30**, 343–354.

Zhang, Y. & Qiu, S. (2015). Examining phylogenetic relationships of *Erwinia* and *Pantoea* species using whole genome sequence data. *Antonie van Leeuwenhoek* **108**, 1037–1046.

Zhang, Y., Fan, Q. & Loria, R. (2016). A re-evaluation of the taxonomy of phytopathogenic genera *Dickeya* and *Pectobacterium* using whole-genome sequencing data. *Syst Appl Microbiol* **39**, 252–259.