INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

TAXONOMIC DESCRIPTION

Bello *et al.*, *Int. J. Syst. Evol. Microbiol.* 2022;72:005284

DOI 10.1099/ijsem.0.005284

MICROBIOLOGY SOCIETY

OPEN ACCESS

Phylogenomic and comparative genomic analyses of *Leuconostocaceae* species: identification of molecular signatures specific for the genera *Leuconostoc*, *Fructobacillus* and *Oenococcus* and proposal for a novel genus *Periweissella* gen. nov.

Sarah Bello†, Bashudev Rudra† and Radhey S. Gupta*

Abstract

The genera *Convivina*, *Fructobacillus*, *Leuconostoc*, *Oenococcus* and *Weissella*, which formed the family *Leuconostocaceae*, have recently been merged within the family *Lactobacillaceae*. Using genome sequences for 47 of the 52 named species from these genera, we report here comprehensive phylogenomic and comparative analyses on protein sequences from these species using multiple approaches. In a phylogenomic tree based on concatenated sequences of 498 core proteins from these five genera, and in a 16S rRNA gene tree, members of the genera *Fructobacillus*, *Leuconostoc* and *Oenococcus* formed distinct strongly supported clades. In contrast, *Weissella* species grouped into two distinct unrelated clades designated as the '*Weissella* main clade' and '*Weissella* clade 2'. The presence of these clades is also seen in a matrix of pairwise average amino acid identity based on core protein sequences. In parallel, comparative genomic studies on protein sequences from *Leuconostocaceae* genomes have identified 46 conserved signature indels (CSIs) in diverse proteins that are unique characteristics of the different observed species clades. Of these identified CSIs, five, five and 13 CSIs are uniquely present in members of the genera *Fructobacillus*, *Leuconostoc* and *Oenococcus*, respectively. We also report here six and five CSIs that are exclusively present in the species from the *Weissella* main clade and *Weissella* clade 2, respectively, providing independent evidence supporting their distinctness from each other. The remaining 12 identified CSIs are commonly shared by some or all of the species from the genera *Convivina*, *Fructobacillus* and *Leuconostoc*, clarifying their interrelationships. The identified CSIs provide novel and reliable means for the identification/circumscription of members of the genera *Fructobacillus*, *Leuconostoc* and *Oenococcus* as well as the two *Weissella* species clades in molecular terms. Based on the strong phylogenetic and molecular evidence presented here, we propose that the genus *Weissella* be limited to only the species from the *Weissell*a main clade, whereas the species forming *Weissella* clade 2 should be transferred to a new genus *Periweissella* gen. nov.

DATA SUMMARY

Supplementary data for this manuscript can be found at https://doi.org/10.6084/m9.figshare.18866273.v1 [1].

INTRODUCTION

The family *Leuconostocaceae* [2, 3], comprises Gram-positive, non-spore-forming, anaerobic or aerotolerant bacteria, which are usually found in nutrient-rich environments such as milk, meat, vegetable products, roots, foods and fermented products [3–5]. Similar to the other lactic acid bacteria, the major end products of their heterofermentative carbohydrate metabolism include lactic

Author affiliations: [1]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada.
***Correspondence:** Radhey S. Gupta, gupta@mcmaster.ca
Keywords: *Fructobacillus*, *Oenococcus* and *Weissella*; description of the genus *Periweissella*; phylogenomic and comparative genomic analyses; conserved signature indels (CSIs); molecular signatures specific for *Leuconostoc*.
Abbreviations: aa, amino acid; AAI, average amino acid identity; CSI, conserved signature indel; CSI, Conserved signature indel (insert or deletion); MSA, multiple sequence alignment; SH, Shimodaira–Hasegawa.
†These authors contributed equally to this work
Two supplementary tables and 47 figures are available with the online version of this article

acid, CO_2, ethanol and/or acetate [3, 4, 6]. Until recently, the family *Leuconostocaceae* consisted of five genera, namely *Convivina* [7], *Fructobacillus* [8], *Leuconostoc* [4, 9, 10], *Oenococcus* [11] and *Weissella* [5, 12], which have now been merged within the family *Lactobacillaceae* [13]. However, as the name *Leuconostocaceae* remains a valid name, for the sake of convenience, in the present study we will be referring to this group of species by this family name. Most of the genera within *Leuconostocaceae* have originated from the taxonomic reassignments of species from the genus *Leuconostoc* [2]. Earlier studies based on phylogenetic analysis of 16S rRNA and 23S rRNA gene sequences [12, 14] led to the transfer of a number of *Leuconostoc* species into two novel genera *viz*. *Weissella* [12] and *Oenococcus* [11]. In 2008, Endo and Okada [8], based on their analysis of the genus *Leuconostoc* using 16S rRNA, 16S–23S rRNA gene intergenic spacer region, and *rpoC* and *recA* genes, transferred four additional *Leuconostoc* species into a new genus *Fructobacillus* [8]. Members of the genus *Fructobacillus* also differ from other *Leuconostoc* species in terms of their morphology, preference for growth in presence of fructose, and several genomic characteristics [15, 16]. Later, Praet *et al.* [7], based on their analysis of 16S rRNA gene sequences and the G+C content, proposed the creation of the genus *Convivina*, which branches in between the genera *Leuconostoc* and *Fructobacillus* in the 16S rRNA gene tree. However, in contrast to the other members of the family *Lactobacillaceae*, whose evolutionary relationships have been extensively studied based on genome sequences [17–20], our current understanding of the evolutionary relationships and classification of species within the family *Leuconostocaceae* is based primarily on analysis of the 16S rRNA gene and a limited number of other gene sequences [4, 5, 16, 21–23], and it requires further investigation. In the published 16S rRNA gene trees, the genera *Leuconostoc*, *Fructobacillus* and *Convivina* form a strongly supported clade [7, 8]. Additionally, in most of the phylogenetic studies of *Weissella* species [24–30], several species from this genus (*viz*. *Weissella beninensis*, *Weissella fabalis*, *Weissella fabaria*, *Weissella ghanensis* and *Weissella cryptocerci*) branch distinctly from the main clade of *Weissella* species containing the type species (*Weissella viridescens*) of this genus [12]. These studies suggest that the *Weissella* species are likely to comprise two phylogenetically distant clades, but this inference needs to be confirmed and supported by other more reliable means. Thus, it is important to carry out detailed phylogenomic and comparative genomic studies of members of the family *Leuconostocaceae* to reliably discern their evolutionary relationships.

The family *Leuconostocaceae* presently contains 49 validly published and three non-validly published species [31]. In the past few years, as a result of several major genomic-sequence projects [32–34], genome sequences for 47 of the *Leuconostocaceae* species have become available in the NCBI database (www.ncbi.nlm.nih.gov/genome/). These genomes provide a comprehensive resource for undertaking detailed studies to clarify the evolutionary relationships among *Leuconostocaceae* species. Using these genome sequences, we have reconstructed a highly resolved phylogenetic tree based on concatenated sequences of 498 core proteins for this family. The sequence alignments of the core proteins were also utilized to determine the pairwise average amino acid identity (AAI) [35] for different members of this family. In addition, we have performed comparative genomic analyses to identify molecular signatures in the form of conserved signature indels (CSIs) in protein sequences, which are specific for different main clades within the family *Leuconostocaceae*. Molecular markers such as the CSIs, which are uniquely shared by a given group of organisms, provide strong evidence independent of phylogenetic analyses of the monophyly and genetic cohesiveness of different observed clades. Furthermore, the CSIs specific for a given clade provide reliable means for the circumscription of these clades in molecular terms [36–40]. The results of these analyses have identified 46 CSIs in diverse proteins that are specific for different strongly supported clades within the family *Leuconostocaceae*, including those that are specific for the genera *Leuconostoc*, *Fructobacillus* and *Oenococcus*. In addition, the results presented here provide strong evidence that the genus *Weissella* is polyphyletic and these species form two distinct clades. The members from these two clades can be reliably distinguished from each other based on their branching in phylogenetic trees and multiple identified CSIs that are exclusively shared by the species from these two clades. Based on the compelling evidence obtained from these studies, we are proposing a division of the genus *Weissella* into an emended genus *Weissella* and a novel genus *Periweissella* gen. nov.

METHODS

Reconstruction of phylogenetic trees

Genome sequences for 47 *Leuconostocaceae* species, whose annotated protein sequences were available in the NCBI genome database, were downloaded. In addition, the genomes of three *Lactobacillaceae* species (*Paucilactobacillus vaccinostercus*, *Lactobacillus delbrueckii* and *Lactobacillus gasseri*) were included in our dataset for rooting the tree. Using these genome sequences, a rooted phylogenomic tree was reconstructed based on concatenated sequences of all core proteins from the family *Leuconostocaceae* by methods detailed in our earlier work [38, 41, 42]. Briefly, the CD-HIT program was used to identify protein families where the proteins were present in at least 80% of the genomes in the dataset and shared at least 50% of sequence length and identity [43]. The Clustal Omega program [44] was then used to generate multiple sequence alignments (MSAs) of the proteins. These MSAs were converted into profile hidden Markov models [45], which were then used to search for other members of the protein families in the input genomes. The alignments obtained were trimmed using TrimAl program [46] to remove poorly aligned sections and to create a core proteins alignment. The final alignment used for phylogenetic analysis was based on 498 proteins and it contained 163109 aligned positions. A maximum-likelihood tree based on this sequence alignment was initially reconstructed with FastTree 2 [47], based on the Whelan and Goldman model [48], and it was then optimized using RAxML based on the Le and Gascuel model [49]. RAxML was also used to calculate Shimodaira–Hasegawa (SH)-like statistical support values for each

node. The resultant phylogenetic tree was drawn using MEGA X [50]. The sequence alignment of the 498 core proteins was also used to determine the pairwise average amino acid sequence identity (AAI) between the type species of different genera within the family *Leuconostocaceae* [51].

A 16S rRNA gene tree was also reconstructed based on sequences of all *Leuconostocaceae* species/type strains obtained from the SILVA ribosomal RNA database [52], and the NCBI genome database (www.ncbi.nlm.nih.gov). *Lactobacillus* species *L. delbrueckii* and *L. gasseri* were included in the dataset for rooting purpose. The sequences were aligned using the MUSCLE program in MEGA X [50]. The non-conserved regions as well as regions with gaps were removed, leaving 1269 positions in the final aligned dataset. A maximum-likelihood phylogenetic tree based on this dataset was created using MEGA X [50], employing the Tamura–Nei model [53], based on 100 bootstrap replicates.

Identification of CSIs

The identification of CSIs was carried out as described in detail in earlier work [37, 54]. Briefly, BLASTp searches using the NCBI non-redundant database were carried out on all proteins from the genomes of *Leuconostoc mesenteroides*, *Oenococcus oeni* and *W. viridescens*. Based on these BLAST searches, protein sequences were obtained for 10–15 divergent *Leuconostocaceae* species and 8–10 species from other bacterial taxa. The multiple sequence alignments of various proteins were created using ClustalX 2.1. These alignments as well as the alignment for various protein families obtained from the CD-HIT program were examined for insertions or deletions of fixed length that were present in conserved regions [i.e. flanked on both sides by at least 4–5 conserved amino acids (aa) in the neighbouring 40–50 aa] and specifically shared by species from different main clades of *Leuconostocaceae* species in the core genome tree. The query sequences of interest containing the identified conserved indels and its flanking 30–50 aa (generally beginning and ending with a stretch of completely conserved amino acid residues) were reBLASTed using the NCBI nr (non-redundant) database and the top 500 hits were examined. Based on these BLASTp searches, conserved indels which were specifically shared by all or most of the species from different main clades of *Leuconostocaceae* were identified and further formatted using the SIG_CREATE and SIG_STYLE programs (available from http://gleans.net/) [37]. Due to space constraints, sequence information is presented in the main figures for only a limited number of species. However, unless otherwise stated, the CSIs described here are exclusively shared by the indicated groups of *Leuconostocaceae* and absent in all other bacterial homologues in the top 500 BLASTp hits examined. More detailed information for different CSIs is provided in the supplementary data [55].

RESULTS

Phylogenetic analysis of the species from the family *Leuconostocaceae*

To elucidate the evolutionary relationships among members of the family *Leuconostocaceae*, we have reconstructed a maximum-likelihood phylogenomic tree based on the genomes of 47 *Leuconostocaceae* species whose sequences were available in the NCBI database. The accession numbers and some other characteristics of the genomes that were utilized for this tree reconstruction are provided in Table S1 (available in the online version of this article) [55]. The resulting tree, which is based on concatenated sequences for 498 proteins that are commonly shared by the species from *Leuconostocaceae* genomes, is shown in Fig. 1. This tree, which will be referred to as the core genome tree, was rooted using the sequences for representative *Lactobacillaceae* species (see Methods). As seen from Fig. 1, all of the nodes in this core genome tree are supported by 100% SH-like statistical support values (similar to the bootstrap scores), which indicates that the evolutionary relationships amongst *Leuconostocaceae* species as seen here are reliable. The tree shown in Fig. 1 provides several important insights into the evolutionary relationships among members of the family *Leuconostocaceae*. First, the tree shows that species from the genera *Leuconostoc*, *Fructobacillus* and *Oenococcus* form strongly supported clades. For the genus *Fructobacillus*, in addition to the genomes for named species, genome sequences are also available for a number of unnamed species. Information for these *Fructobacillus* species is included in a phylogenetic tree presented in Fig. S1 [55] and these species also grouped reliably with the other members of the genus *Fructobacillus*. Second, the species *Convivina intestini*, which is the sole species in the genus *Convivina*, branches distinctly in between the genera *Fructobacillus* and *Leuconostoc*. Third, the tree also shows that the species from the genera *Leuconostoc*, *Fructobacillus* and *Convivina* form a strongly supported clade, which is separated from the neighbouring genus *Oenococcus* by a long branch. We will be referring to this clade comprising the genera *Leuconostoc*, *Fructobacillus* and *Convivina* as the 'larger *Leuconostoc* clade'. Lastly, a fourth important aspect of this tree is that it shows that species from the genus *Weissella* do not form a monophyletic lineage, but they are separated into two distinct clades/lineages. Of the two *Weissella* species clades, the larger clade contains the species *W. viridescens*, which is the type species of this genus. Hence, we have designated this clade as the '*Weissella* main clade'. The second *Weissella* clade comprises three genome-sequenced species (*W. cryptocerci*, *W. beninensis*, *W. fabalis*) and this clade forms an outgroup of the remainder of the *Weissella* species as well as other *Leuconostocaceae* genera. We will be referring to this smaller clade as the '*Weissella* clade 2'.

In addition to the core genome tree, we have also reconstructed a phylogenetic tree based on 16S rRNA gene sequences for the type strains of all species from the family *Leuconostocaceae* (Fig. 2). The overall evolutionary relationships among the *Leuconostocaceae* species in the 16S rRNA gene tree are very similar to that seen in the core genome tree (Fig. 1). The species from the genera

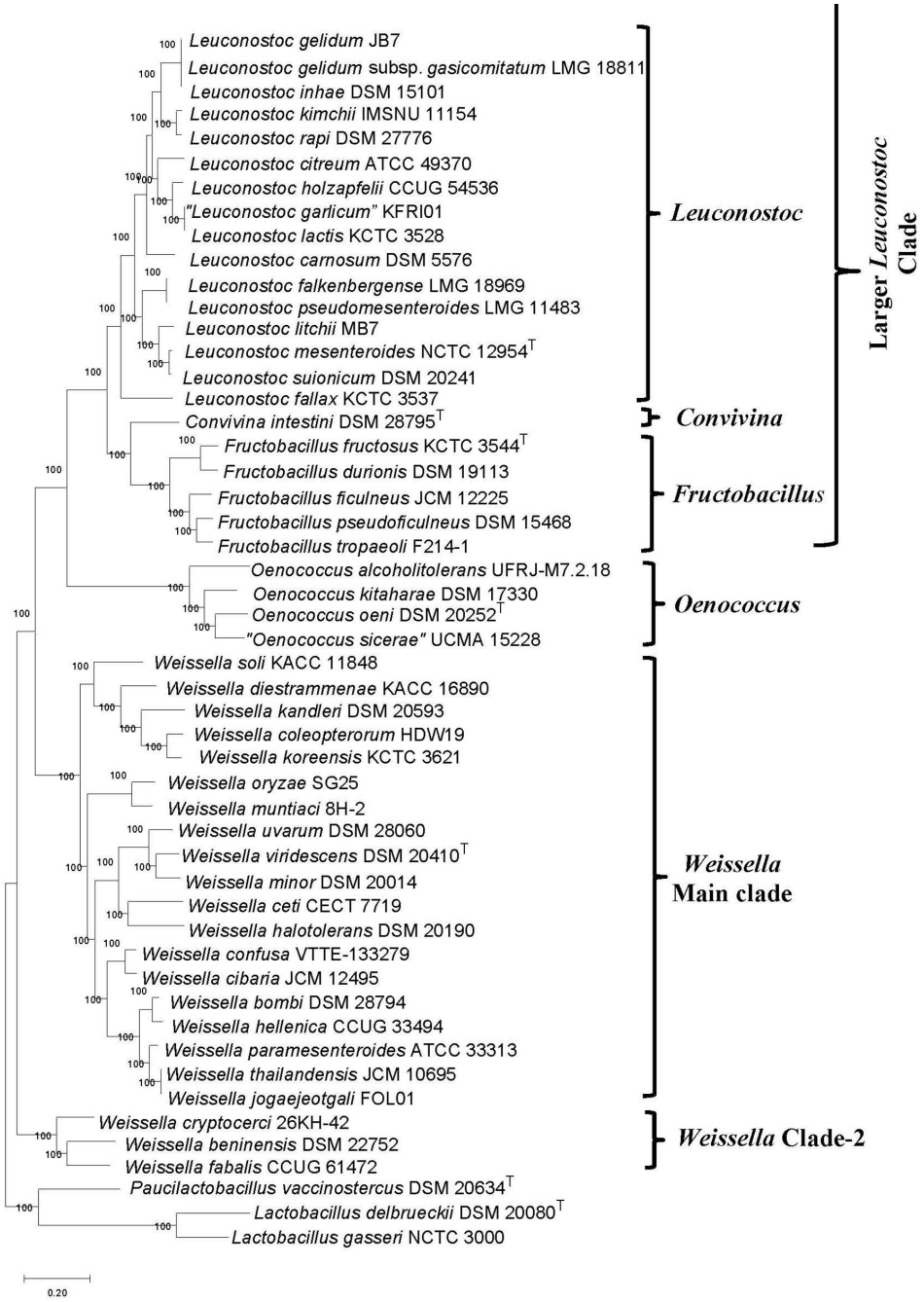

Fig. 1. A bootstrapped maximum-likelihood tree for 47 genomes sequenced *Leuconostocaceae* species based on concatenated sequences for 498 core proteins. The statistical support values for different branches are indicated on the nodes. This tree was rooted by using species from the genus *Lactobacillus*. Non-validly published species are shown within quotation (" ") marks. Different main species clades observed in the tree are identified by the names of the genera or other designated clade names.

Fructobacillus and *Oenococcus* formed monophyletic clades. All species from the genus *Leuconostoc*, except *Leuconostoc fallax* which branched more deeply, also formed a well-supported clade. Similar branching of *L. fallax* in 16S rRNA gene tree has also been observed in an earlier study [8]. Furthermore, a close relationship of the species from the genera *Leuconostoc*, *Fructobacillus* and *Convivina* is also observed in this tree. In addition, species from the genus *Weissella* also formed two distinct clades, which were separated by long branches. Of these two *Weissella* species clades, one clade consisting of three genome sequenced species, namely *W. beninensis*, *W. cryptocerci* and *W. fabalis*, and two other species (*viz. W. fabaria* and *W. ghanensis*) formed a sister lineage of the remainder of the *Weissella* species and other *Leuconostocaceae* genera. As the tree based on core genome proteins is more

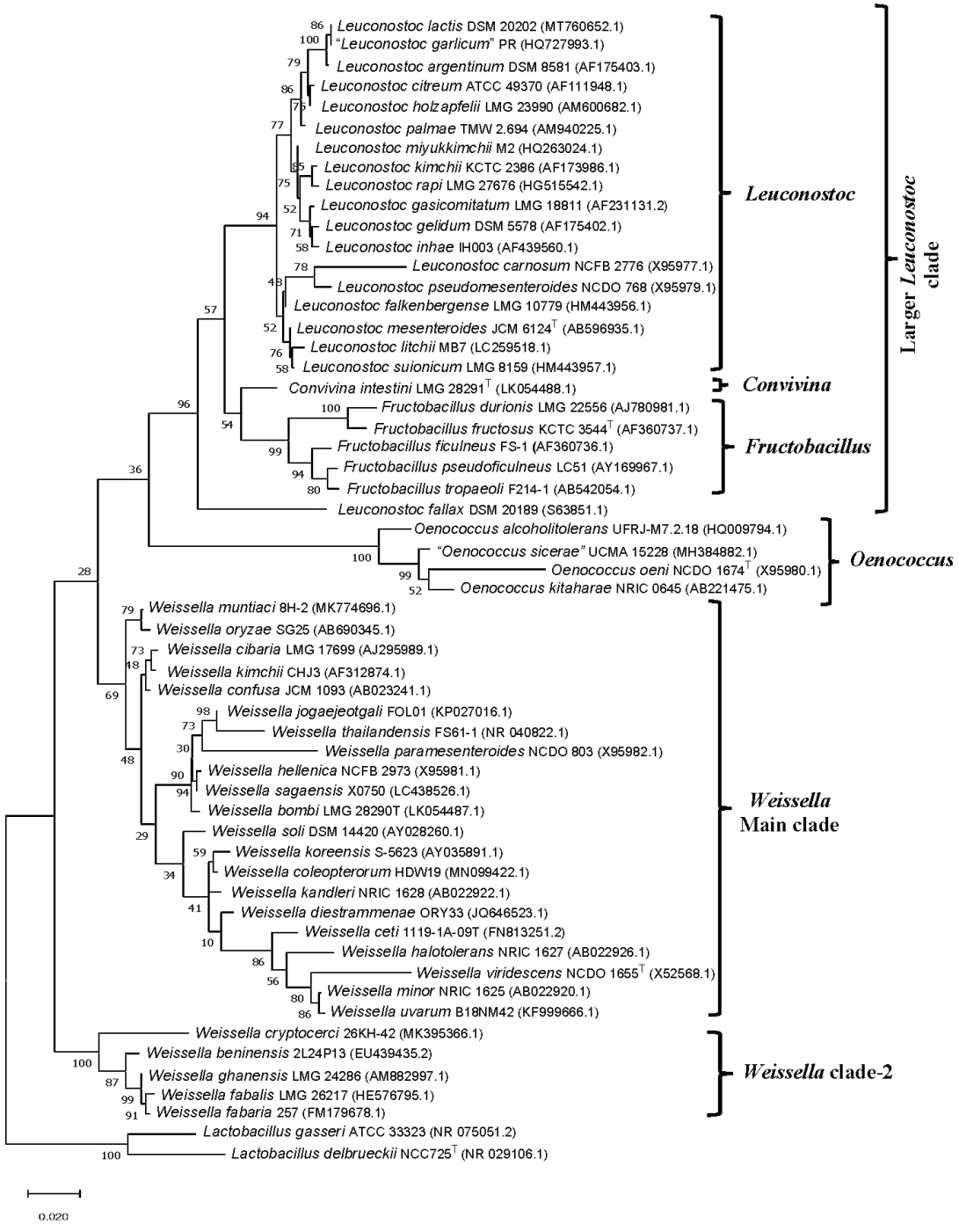

Fig. 2. A maximum-likelihood phylogenetic tree based on 16S rRNA gene sequences for the type strains of all validly published *Leuconostocaceae* species. *Lactobacillus* species *L. delbrueckii* and *L. gasseri* were used to root the tree. The accession numbers of the 16S rRNA gene sequences are given within bracket after each species in the tree. Different main clades within the tree are marked with the names of the genera or other given names.

reliable and provides higher resolution than the 16S rRNA gene tree, we have generally relied on it for most of the phylogenetic inferences derived in this study.

The sequence alignment of the core genome proteins from members of the family *Leuconostocaceae* was also used to calculate pairwise AAI, which provides a measure of the overall genetic relatedness among different species [51]. The matrix depicting the pairwise AAI information for members of the family *Leuconostocaceae* is presented in Fig. 3. Detailed information regarding pairwise AAI is provided in Table S2. In Fig. 3, the genome pairs exhibiting higher sequence similarities are shown by a darker

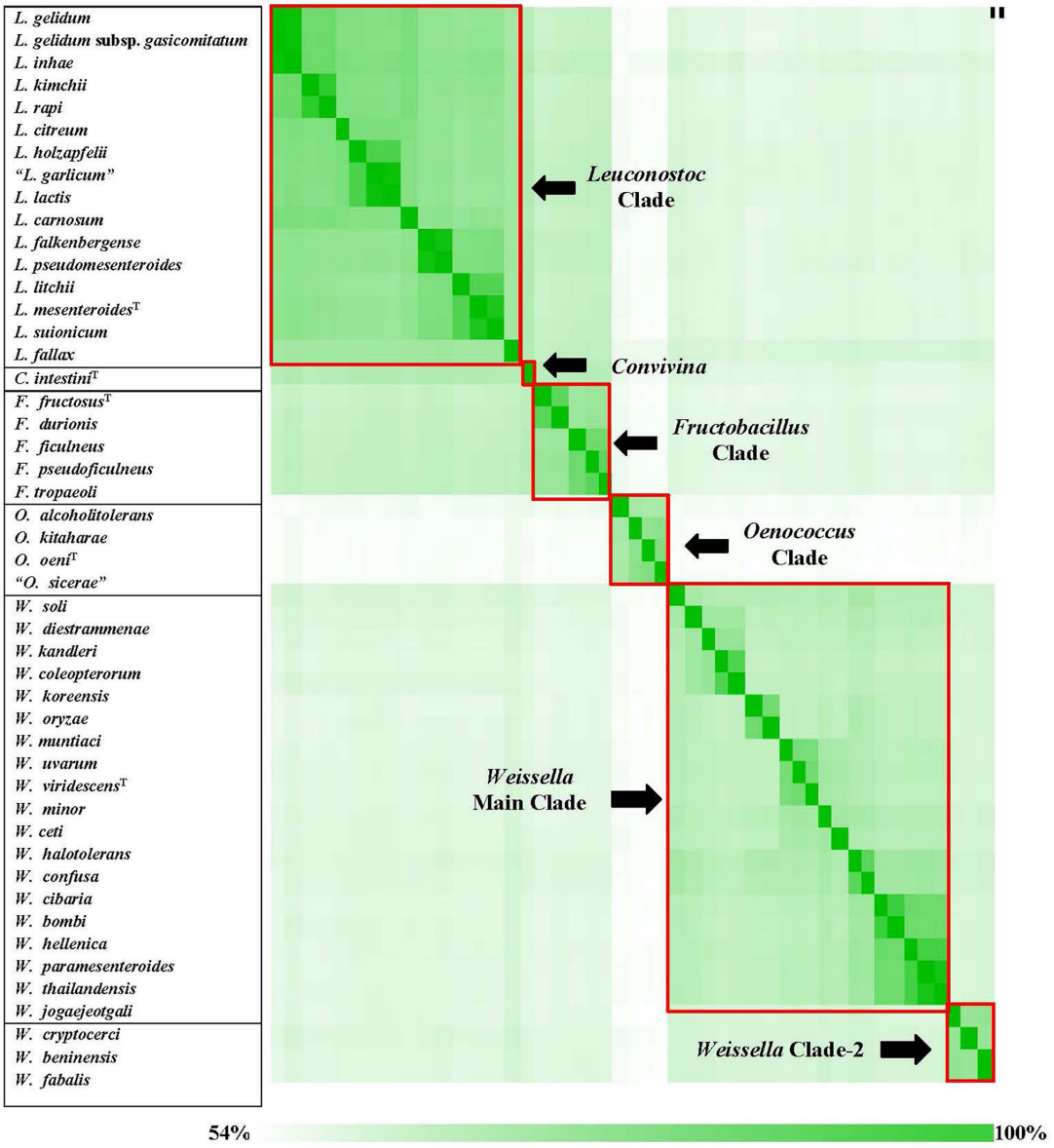

Fig. 3. A matrix indicating the pairwise percentage average amino acid identities of the species from different genera within the family *Leuconostocaceae*. Genome pairs sharing higher amino acid identity are shaded more darkly (green). The regions of the matrix corresponding to different clades have been marked and labelled.

shade of green. As seen from Fig. 3, based on the AAI similarity data, species from various genera within the family *Leuconostocaceae* exhibit higher intra-genus AAI values in comparison to the inter-group AAI values (see Table S2) [55]. Based on the AAI values, a closer relationship is also observed between members of the genera *Leuconostoc*, *Convivina* and *Fructobacillus*. The intra-group AAI value for this larger *Leuconostoc* clade is 0.73 in comparison to the inter-generic AAI values, which are in the range of 0.64–0.69. Based on AAI analysis, species from the two *Weissella* clades are also more closely related to each other than to the other *Leuconostocaceae* genera. However, AAI values are not a reliable tool for the demarcation of genera as there is no established threshold for distinction between adjacent bacterial taxa [35, 56].

Identification of molecular markers specific for different clades within the family *Leuconostocaceae*

The results of our phylogenomic studies and AAI analysis indicate that the family *Leuconostocaceae* comprises a number of distinct clades including some novel species groupings. However, based upon the branching of the species in phylogenetic trees, it is often difficult to reliably delimit the boundaries of different clades [39]. Hence, we have also conducted detailed comparative studies on protein sequences from *Leuconostocaceae* genomes to identify molecular markers in the forms of CSIs, which are

uniquely shared by members of different observed clades. The CSIs in gene/protein sequences, which are specifically shared by the members of a given clade, constitute synapomorphic characteristics and they provide important class of molecular markers for evolutionary and taxonomic studies [38, 41, 42, 57]. Our analyses of protein sequences from *Leuconostocaceae* genomes have identified 46 CSIs specific for different clades within this family and provide important means for their demarcation in molecular terms. The group-specificities and characteristics of the identified CSIs are described below.

CSIs specific for the genera *Leuconostoc* and *Fructobacillus* and for the larger *Leuconostoc* clade

Leuconostoc and *Fructobacillus* are two of the main genera within the family *Leuconostocaceae*. The work that we have carried out has identified five CSIs each in different proteins that are specifically shared by the members of each of these two genera. In Fig. 4(a), we present an example of a CSI consisting of a 2 aa insert in an RNA-binding transcriptional accessory protein, which is uniquely shared by all species from the genus *Leuconostoc*, but not found in any other bacteria within the top 500 BLASTp hits. Interestingly, the homologs of this protein were not found in *Fructobacillus*, *Convivina* and *Oenococcus* species. More detailed information for this CSI and the sequence information for four other CSIs, which are also shown to be specific for the genus *Leuconostoc*, is provided in Figs S2–S6 and some of their characteristics summarized in Table 1. In Fig. 4(b), we show the partial sequence alignment of the protein Asp-tRNA (Asn)/Glu tRNA (Gln) amidotransferase subunit (GatB). The 4 aa insert highlighted in this figure is specific for all members of the genus *Fructobacillus* including a number of unnamed *Fructobacillus* species. Besides the *Fructobacillus* species, this insert is again not found in any other *Leuconostocaceae* species or other bacteria within the top BLASTp 500 hits. In addition to the CSI shown in Fig. 4(b), our analysis has identified four additional CSIs in other proteins, which are also specific for members of the genus *Fructobacillus*. Detailed sequence information for all of the *Fructobacillus*-specific CSIs is provided in Figs S7–S11 [55] and some of their characteristics are summarized in Table 1.

In our core genome tree as well as in the 16S rRNA gene tree (Figs. 1 and 2), the species *C. intestini* forms outgroup of the genus *Fructobacillus.* A close relationship of *C. intestini* to *Fructobacillus* has also been observed in earlier studies [7]. Our analysis has identified two CSIs that are commonly and exclusively shared by the members of these two genera. Sequence information for one of these CSIs, consisting of a 1 aa deletion in the protein mevalonate kinase is presented in Fig. 5(a). As can be seen, the identified CSI is specifically shared by *C. intestini* and all of the *Fructobacillus* species including unnamed members of this genus. More detailed information for this CSI, as well as sequence information for another CSI showing similar specificity found in the protein D-alanyl-lipoteichoic acid biosynthesis protein (DltB), is provided in Figs S12 and S13 and some of their characteristics summarized in Table 1.

In our core genome tree, members of the *Leuconostoc*, *Fructobacillus* and *Convivina* form a strongly supported clade, which we have designated as the 'Larger *Leuconostoc* clade'. A close relationship of the species from these three genera is also evident from the results of the AAI matrix (Fig. 3). Furthermore, a close relationship of the species from these three genera is independently strongly supported by our identification of 10 CSIs in different proteins, which are specifically shared by the members of these three genera. One example of a CSI specific for these three genera is presented in Fig. 5(b), where a 5 aa insert in the protein phenylalanine tRNA ligase beta subunit is exclusively present in all members of these three genera, but it is not found in any other *Leuconostocaceae* genera or other bacteria. More detailed information for this CSI and sequence information for the other nine CSIs, which are also specific for the larger *Leuconostoc* clade, is provided in Figs S14–S23 [55] and some of their characteristics are summarized in Table 1. The results from these CSIs provide strong evidence that the members of these three genera (*viz*. *Leuconostoc*, *Fructobacillus* and *Convivina*) shared a common ancestor exclusive of all other bacteria and they provide reliable means to demarcate the species from this clade in molecular terms.

CSIs specific for the genus *Oenococcus* and for the two clades of *Weissella* species

In our core genome tree, members of the genus *Oenococcus* form a strongly supported clade, which is separated from all other *Leuconostocaceae* genera by a long branch (Fig. 1). The distinctness of the genus *Oenococcus* from all other *Leuconostocaceae* genera is also strongly supported by 13 identified CSIs, which are exclusively shared by the members of this genus. Sequence information for one of these CSIs is shown in Fig. 6(a), where a 4 aa insertion in the protein DNA-directed RNA polymerase beta subunit is exclusively present in all *Oenococcus* species but not found in the protein homologs from any other bacteria in the top 500 BLASTp hits. Detailed sequence information for this CSI as well as 12 others, which are also specific for the genus *Oenococcus*, is presented in Figs S24–S36 [55] and some of their characteristics are summarized in Table 2.

The *Weissella* species form two distinct clades in the core genome tree as well as in the tree based on 16S rRNA gene sequences. Our comparative genomic analyses have identified multiple CSIs that are specific for the members of these two clades reliably distinguishing them from each other as well as other bacteria. Of these CSIs, six are exclusively found in the species from the '*Weissella* main clade'. One example of such a CSI is presented in Fig. 6B, where an 8 aa insert in the protein phospho-*N*-acetylmuramoyl-pentapeptide-transferases is exclusively present in all species from the *Weissella* main clade but not found in the protein homologs from other *Weissella* species or any other bacteria in the top 500 BLASTp hits. Detailed sequence information for this CSI and five other CSIs specific for the *Weissella* main clade is presented in Figs S37–S42 [55] and some of their characteristics are summarized in Table 2. The *Weissella* clade 2 comprises three genome-sequenced species (*viz. W.*

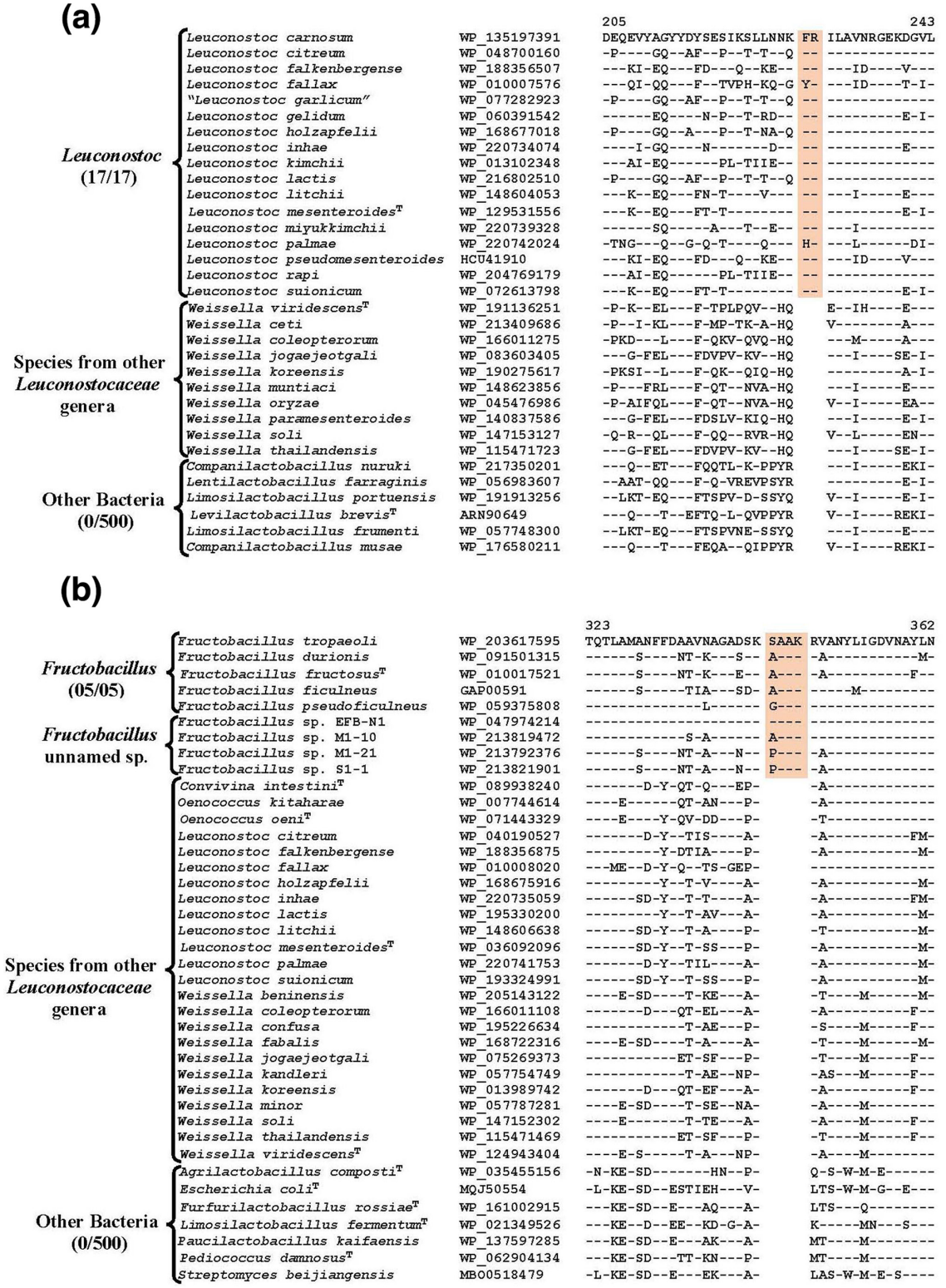

Fig. 4. (a) Partial sequence alignment of the RNA-binding transcriptional accessory protein showing a 2 aa insertion (boxed) that is exclusively shared by all species from the genus *Leuconostoc*. Detailed sequence information for this CSI as well as four other CSIs specific for the genus *Leuconostoc* are presented in Figs S2–S6 and some of their characteristics are summarized in Table 1. (b) Excerpts from the sequence alignment of the protein Asp-tRNA(Asn)/Glu tRNA(Gln) amidotransferase subunit (GatB) showing four aa insertion in a conserved region that is specific for all species from the genus *Fructobacillus*. Detailed sequence information for this CSI as well as four other CSIs specific for the genus *Fructobacillus* are presented in Figs S7–S11 [55] and some of their characteristics are summarized in Table 1.

Table 1. Conserved signature indels specific for different clades of the family *Leuconostocaceae*

Protein name	Accession no	Indel size	Indel position	Figure no	Specificity
RNA-binding transcriptional accessory protein	WP_135197391	2 aa Ins	205–243	Figs 4(a) and S2	
BMP family protein*	WP_150280547	1 aa Ins	51–98	Fig. S3	
BMP family protein*	WP_150280547	1 aa Ins	90–122	Fig. S4	*Leuconostoc*
Universal stress protein	SPJ43140	2 aa Del	5–41	Fig. S5	
Copper resistance protein	WP_150259299	2 aa Del	37–74	Fig. S6	
Asp-tRNA(Asn)/Glu tRNA(Gln) amidotransferase subunit GatB	WP_203617595	4 aa Ins	323–362	Figs 4(b) and S7	
Xanthine phosphoribosyltransferase	WP_059378047	1 aa Ins	136–182	Fig. S8	
ABC transporter ATP-binding protein/ permease	WP_059376430	1 aa Del	326–360	Fig. S9	*Fructobacillus*
NCS2 family nucleobase:cation symporter	WP_187753602	3 aa Ins	205–245	Fig. S10	
Ribonuclease J	WP_059375728	1 aa Ins	186–215	Fig. S11	
Mevalonate kinase	WP_203618360	1 aa Del	146–183	Figs 5(a) and S12	
D-Alanyl-lipoteichoic acid biosynthesis protein DltB†	WP_061992753	1 aa Ins	36–71	Fig. S13	*Fructobacillus* and *Convivina*
Phenylalanine tRNA ligase subunit beta	WP_091502306	5 aa Ins	84–127	Figs 5(b) and S14	
Valine tRNA ligase	WP_010386363	2 aa Del	650–689	Fig. S15	
Phenylalanine tRNA ligase subunit alpha	WP_010692108	4aa Ins	110–148	Fig. S16	
Diphosphomevalonate decarboxylase protein	WP_011680078	1 aa Ins	210–249	Fig. S17	
Single-stranded-DNA-specific exonuclease RecJ	WP_059377347	1 aa Ins	254–295	Fig. S18	Larger *Leuconostoc* clade (*Leuconostoc*, *Convivina*, *Fructobacillus*)
RluA family pseudouridine synthase	WP_089937871	3 aa Ins	234–376	Fig. S19	
ATP-dependent Clp protease ATP-binding subunit ClpX	WP_089938723	1 aa Ins	13–42	Fig. S20	
PolC-type DNA polymerase III	WP_089939457	1 aa Ins	926–957	Fig. S21	
Transcription-repair coupling factor	WP_091502582	2 aa Ins	531–576	Fig. S22	
Chromosome segregation protein SMC	WP_148606465	2 aa Del	565–606	Fig. S23	

*Not shared by *Leuconostoc fallax*.
†Also shared by two *Lactobacillaceae* species.

cryptocerci, *W. beninensis*, *W. fabalis*) and five of the identified CSIs are exclusively shared by these three species. Sequence information for one of the CSIs specific for *Weissella* clade 2 is presented in Fig. 6C. In the example shown, a 3 aa deletion in the protein DEAD/DEAH box helicase is exclusively present in all three members of *Weissella* clade 2, but not found in the protein homologs from any other bacteria in the top 500 BLASTP hits. Detailed sequence information for this CSI and four other CSIs that are also specific for *Weissella* clade 2 is presented in Figs S43–S47 [55] and some of their characteristics are summarized in Table 2.

DISCUSSION

Members of the family *Leuconostocaceae* belong to a group of bacteria commonly referred to as lactic acid bacteria. Most of these bacteria generally produce lactic acid as a byproduct of sugar degradation [3]. Because of this trait, these bacteria have found widespread usage in food manufacturing for the purpose of various fermentation processes/products [3]. These

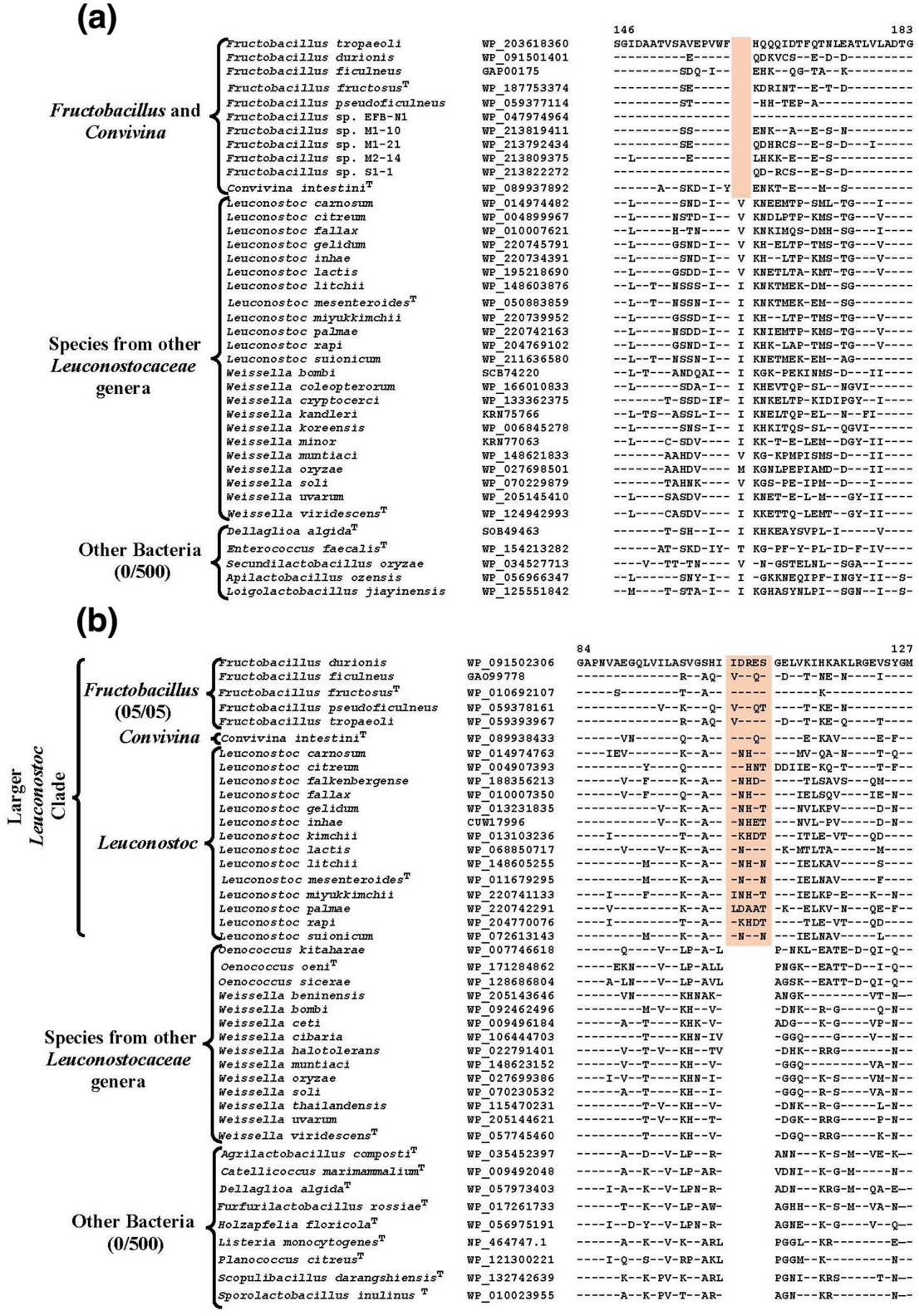

Fig. 5. (a) A partial sequence alignment of the protein mevalonate kinase showing a 1 aa deletion (boxed) that is exclusively shared by all species from the genera *Fructobacillus* and *Convivina*. Sequence information for one more CSI specific for these two genera is presented in Fig. S13 (Table 1). (b) Excerpts from the sequence alignment of the protein phenylalanine tRNA ligase subunit beta showing a 5 aa insertion in a conserved region that is specifically present in all species from the genera *Leuconostoc*, *Convivina* and *Fructobacillus*. Sequence information for nine other CSIs showing similar specificities are presented in Figs S15–S23 [55] and some of their characteristics are summarized in Table 1.

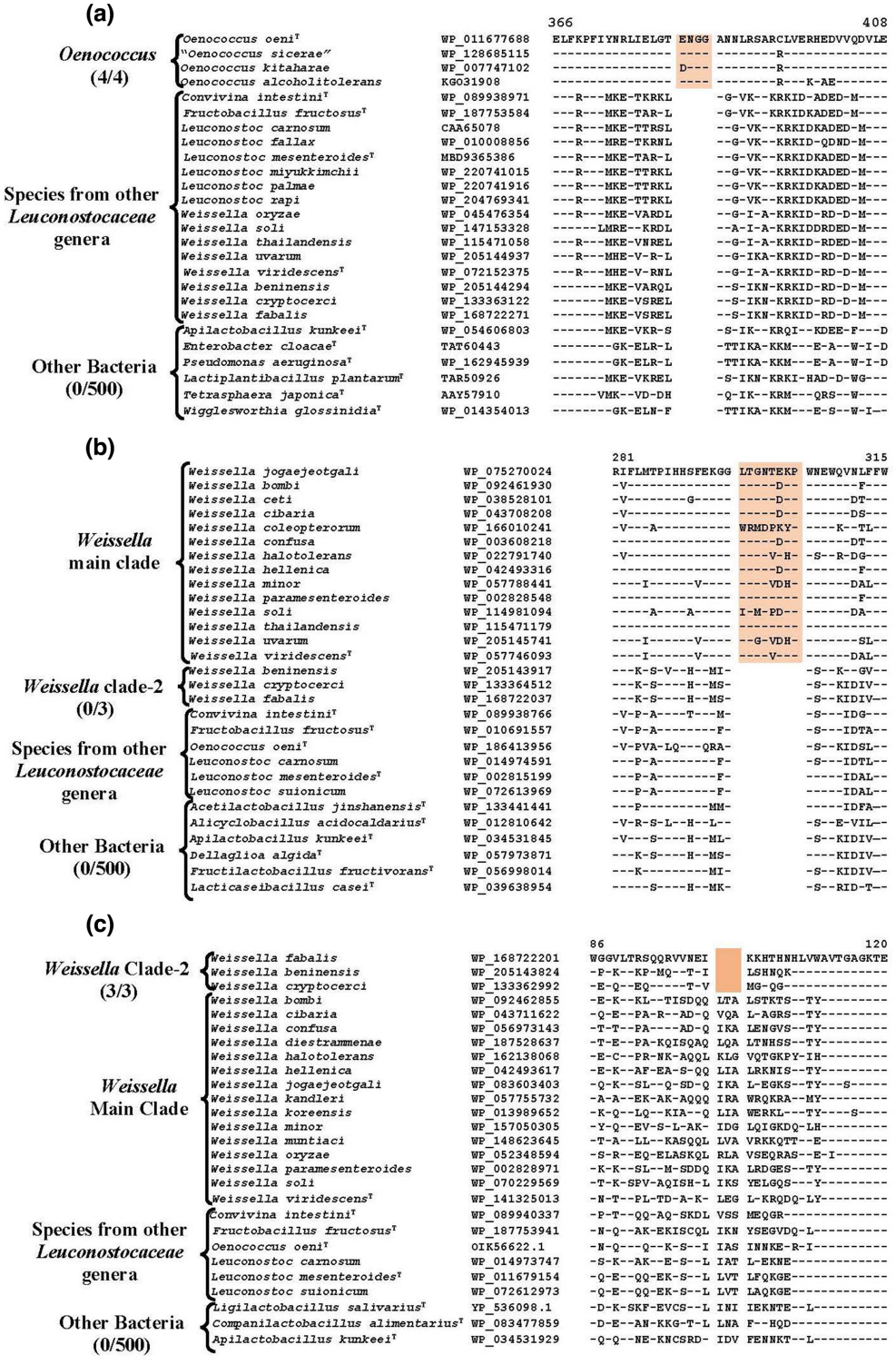

Fig. 6. (a) Partial sequence alignment of the protein DNA-directed-RNA polymerase subunit beta showing a 4 aa insertion (boxed) that is exclusively present in all species from the genus *Oenococcus*. Sequence information for 12 other CSIs specific for the genus *Oenococcus* are presented in Figs S25–S36 and some of their characteristics are summarized in Table 2. (b) Excerpts from the sequence alignment of the protein phospho-*N*-acetylmuramoyl-pentapeptide-transferases showing eight aa insertion in a conserved region that is exclusively shared by all species from the *Weissella* main clade. Sequence information for five other CSIs specific for this clade are presented in Figs S38–S42 [55] and some of their characteristics are summarized in Table 2. (c) Partial sequence alignment of the protein DEAD/DEAH box helicase showing a 3 aa deletion in a conserved region that is specifically present in all species from the *Weissella* clade 2. Sequence information for four other CSIs showing similar specificity is presented in Figs S44–S47 and some of their characteristics are summarized in Table 2.

Table 2. Conserved signature indels specific for the genus *Oenococcus* and for the two *Weissella* species clades

Protein name	Accession no	Indel size	Indel position	Figure no	Specificity
DNA-directed-RNA polymerase subunit beta	WP_011677688	4 aa Ins	366–408	Figs 6(a) and S24	
Preprotein translocase subunit SecY	WP_071450607	1 aa Ins	31–72	Fig. S25	
Preprotein translocase subunit SecY	WP_096866568	2 aa Ins	358–408	Fig. S26	
50S ribosomal protein L13	WP_096877109	1 aa Ins	62–98	Fig. S27	
Molecular chaperone DnaK	WP_002816776	4 aa Del	244–287	Fig. S28	
Riboflavin kinase	WP_180370943	5 aa Ins	22–83	Fig. S29	
Sua5/YciO/YrdC/YwlC family protein	WP_143795362	1 aa Ins	54–89	Fig. S30	*Oenococcus*
Amino acid permease	WP_143805135	2 aa Ins	24–63	Fig. S31	
RluA family pseudouridine synthase	WP_007746277	1 aa Ins	145–189	Fig. S32	
TatD family hydrolase	WP_180369227	2 aa Ins	48–70	Fig. S33	
YidC/Oxa1 family membrane protein insertase	WP_180369397	1 aa Ins	19–76	Fig. S34	
Class I SAM-dependent RNA methyltransferase	WP_002824330	1 aa Ins	73–111	Fig. S35	
GTPase HflX	WP_071438355	2 aa Ins	67–135	Fig. S36	
Phospho-*N*-acetylmuramoyl-pentapeptide-transferases	WP_075270024	8 aa Ins	281–315	Figs 6(b) and S37	
APC family permease	WP_070229375	1 aa Del	579–609	Fig. S38	
Alanine tRNA ligase*	WP_075269877	1 aa Del	368–401	Fig. S39	*Weissella* main clade
Cytochrome d ubiquinol oxidase subunit II	WP_070230395	1 aa Ins	165–205	Fig. S40	
Response regulator transcription factor	WP_070230808	1 aa Ins	178–211	Fig. S41	
Endonuclease MutS2	WP_115471653	2 aa Ins	614–655	Fig. S42	
DEAD/DEAH box helicase	WP_168722201	3 aa Del	86–120	Figs 6(c) and S43	
Hydroxyethylthiazole kinase	WP_133364496	2 aa Ins	48–93	Fig. S44	
ArgR family transcriptional regulator	WP_133363833	1 aa Ins	3–45	Fig. S45	*Weissella* clade 2
Flp pilus assembly complex ATPase component TadA	WP_168722024	1 aa Ins	136–174	Fig. S46	
Amidophosphoribosyl transferase protein	WP_133362570	1 aa Ins	64–102	Fig. S47	

*Also shared by a few *Bacillales* species.

bacteria are commonly present in human and animal gastrointestinal tracts, plants, dairy products and some beverages and some of them are also of clinical significance [3, 23]. Due to these characteristics, it is important to understand the evolutionary relationships among these bacteria and to identify novel and reliable means for the identification of different groups within these bacteria [3, 4, 13, 15]. In the present work, we have examined the evolutionary relationships among members of the family *Leuconostocaceae* based on phylogenetic and comparative analyses of protein sequences from whole genomes. Although the family *Leuconostocaceae* has recently been merged within the family *Lactobacillaceae* [13], for the sake of convenience, this group is referred to here by the name *Leuconostocaceae*, which remains a valid name under the prokaryotic code [58]. Unlike other members of the family *Lactobacillaceae*, whose evolutionary relationships has been studied in detail based on genomic sequences [13], our current understanding of the evolutionary relationships among members of the family '*Leuconostocaceae*' is mainly based on phylogenetic analysis of 16S rRNA gene and in some cases a few housekeeping genes [11, 21, 22, 59]. Thus, the focus of the present work was to examine the evolutionary relationships among members of the family *Leuconostocaceae* based on genome sequence data.

The present work reports comprehensive phylogenomic and comparative analyses on the genome sequences for most of the species (47 of the 52 named species) from *Leuconostocaceae* family using a number of different approaches. The approaches

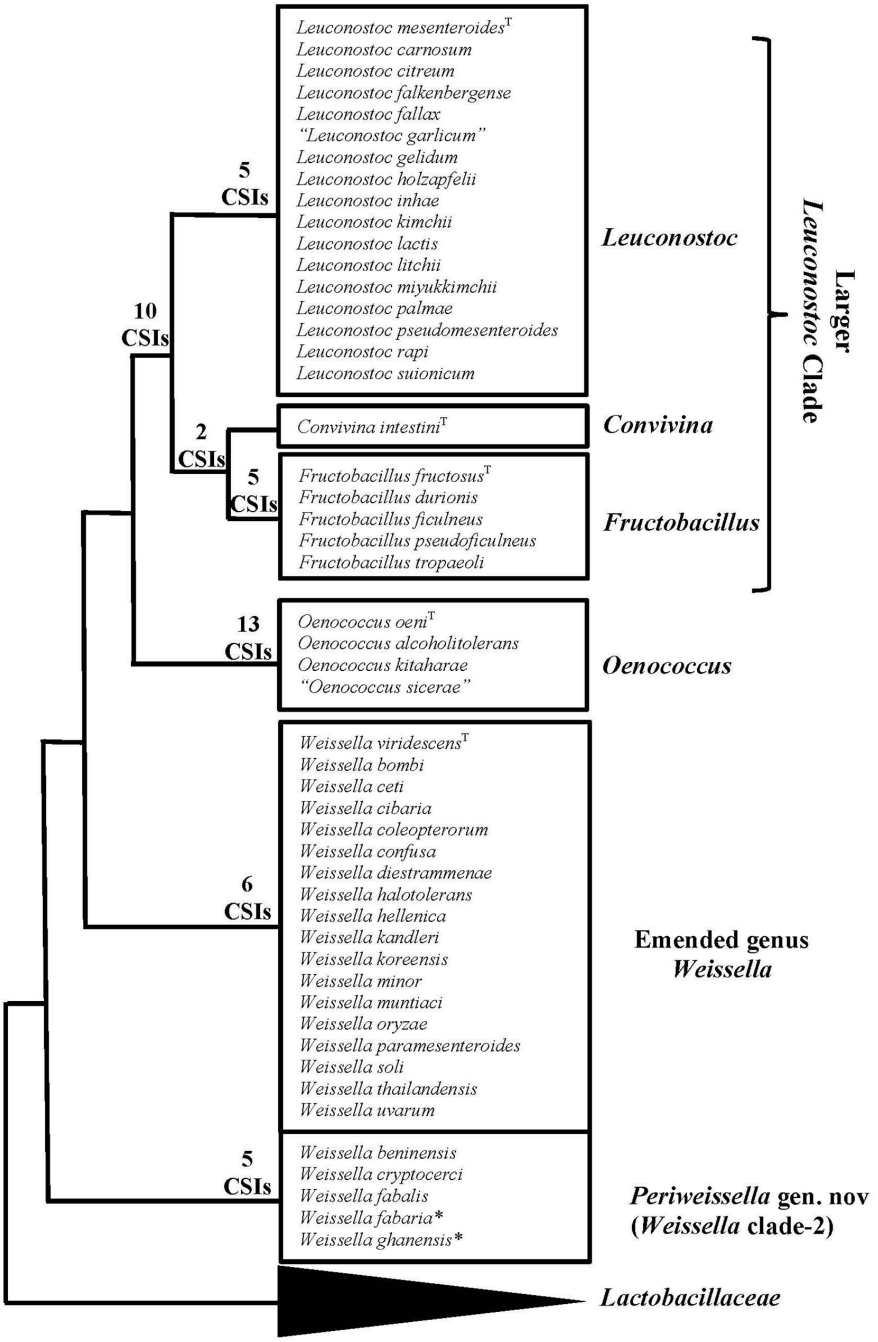

Fig. 7. A conceptual diagram summarizing the results of our phylogenomic and comparative genomic studies on members of the family *Leuconostocaceae*. The numbers of CSIs, which constitute molecular syanpomorphies, that are specifically shared by members of different obsevred clades are shown on the nodes. Members of the *Weisssella* clade 2 are proposed as a novel genus *Periweissella* gen. nov. The * indicates that the placement of these non-genome sequenced species into the genus *Periweisslla* is based on branching in the 16S rRNA gene tree.

used include: (i) reconstruction of a phylogenetic tree based on concatenated sequences of 498 core proteins from their genomes (Fig. 1); (ii) reconstruction of a phylogenetic tree based on 16S rRNA gene sequences for all *Leuconostocaceae* species with validly published names (Fig. 2); (iii) reconstruction of a pairwise AAI matrix for different species based on core genome proteins (Fig. 3), and (iv) detailed analyses of protein sequences from *Leuconostocaceae* species to identify CSIs that are specific for members of different clades. These latter studies have identified 46 novel CSIs that are uniquely shared characteristics of different main clades of *Leuconostocaceae* species observed in core genome tree providing reliable means for their demarcation in molecular terms. The results from all these analyses present a consistent picture concerning the

evolutionary relationships among different *Leuconostocaceae* species/genera. A conceptual diagram summarizing the results of these studies as well as the numbers and clade specificities of different identified CSIs is presented in Fig. 7.

As seen from Fig. 7, the results from different studies support the monophyletic grouping of the species from the genera *Fructobacillus*, *Leuconostoc* and *Oenococcus*. The members of these genera can also be reliably distinguished from each other as well as all other bacteria on the basis of five, five and 13 CSIs identified in the present work that are uniquely shared properties of the members of these genera. The genus *Convivina* contains only a single species, which branches in between the genera *Fructobacillus* and *Leuconostoc* as an outgroup of the genus *Fructobacillus*. A close relationship of *C. intestini* to *Fructobacillus* is also supported by two CSIs that are uniquely shared by these two groups of species. The results presented here also show that members of the genera *Convivina*, *Fructobacillus* and *Leuconostoc* form a strongly supported clade in the core genome tree. A specific grouping of the species from these three genera is also supported by 10 identified CSIs, which are commonly and uniquely shared by the members of these three genera. Furthermore, in an AAI matrix reconstructed based on the core proteins from the family *Leuconostocaceae*, the clade consisting of these three genera has average AAI of value of 0.73, which is comparable to the AAI values seen for some other genera *viz*. *Oenococcus*. Thus, based on their phylogenetic grouping, AAI value, and the sharing of large numbers of CSIs, a case can be made for combining species from all three genera into the genus *Leuconostoc.* However, we do not favour the amalgamation of these three genera, as extensive work on *Fructobacillus* species provide compelling evidence that they differ from *Leuconostoc* species both in terms of their morphology as well as large numbers of biochemical characteristics, including their preference for fructose, need for an electron acceptor for glucose assimilation [3, 8, 15, 16]. In addition, *Fructobacillus* species have smaller genome sizes and lower G+C content in comparison to the *Leuconostoc* species [3, 8, 15, 16]. Furthermore, multiple CSIs identified in the present work, which are exclusively shared by either the *Fructobacillus* or *Leuconostoc* species, also strongly support the distinctness of these two groups of bacteria.

The results presented here also provide compelling evidence that the members of the genus *Weissella* do not constitute a monophyletic grouping but instead comprise two distinct unrelated clades, designated in this work as the '*Weissella* main clade' and '*Weissella* clade 2'. The branching of *Weissella* species into two distantly related clades is also observed in earlier studies on the members of this genus [24–29]. In the present work, we have identified six and five CSIs that exclusively found in either different species from the *Weissella* main clade, or which are specific for the *Weissella* clade 2 species. In contrast, no CSI was identified that is commonly shared by all *Weissella* species. The identified CSIs provide strong independent evidence supporting the distinctness of these two clades. It should be noted that the *Weissella* species are also assigned into different clades in the Genome Taxonomy Database [60], which is now a widely used resource for taxonomic studies. Based upon these results, we are proposing division of the genus *Weissella* into two genera, an emended genus *Weissella* corresponding to the *Weissella* main clade, which contains the type species of this genus *Weissella mesenteroides* [61], and a new genus *Periweissella* gen. nov., harbouring various species from the *Weissella* clade 2.

The CSIs in protein sequences result from rare genetic changes [37, 39, 62]. Hence, the shared presence of these molecular synapomorphies by a given clade of species provides strong evidence, independently of the phylogenetic tree, that the species from that clade shared a common ancestor exclusive of all other bacteria and they are specifically related to each other [37, 39, 62]. Additionally, earlier work on CSIs provides evidence that these molecular markers possess high degree of predictive ability to be found in other unidentified or uncharacterized members of these clades [39, 41, 63]. In the present work, the CSIs specific for the genus *Fructobacillus* are not only commonly shared by all named species from this genus, but also in several unnamed strains/species of *Fructobacillus*, demonstrating the predictive ability of these markers to be present in other novel or uncharacterized members of a given group. In view of these characteristics, the CSIs that are specific for different clades now provide novel and reliable means for the demarcation of different clades of organisms in molecular terms and have proven very useful for evolutionary/taxonomic studies [41, 42, 57, 63]. To incorporate the information for the CSIs that are specific for the genera *Fructobacillus*, *Leuconostoc* and *Oenococcus*, emended descriptions of these taxa are also provided. The descriptions of the emended and novel taxa are given below.

EMENDED DESCRIPTION OF THE GENUS *LEUCONOSTOC* VAN TIEGHEM 1878 (APPROVED LISTS 1980)

Leuconostoc (Leu.co.nos'toc. Gr. masc. adj. *leukos*, clear, light; N.L. neut. n. *Nostoc*, algal generic name; N.L. neut. n. *Leuconostoc*, colourless *Nostoc*).

The description of this genus is partially based on the original description by van Tieghem *et al.* [9] and Bjorkroth *et al.* [4]. Cells are Gram-positive, non-spore forming, non-fructophillic, facultatively anaerobic, heterofermentative, non-motile, catalase- and oxidase negative, ovoid or coccus shaped bacteria. Most species have been isolated from fermented dairy and legumes. They grow within the temperature range of 10–40 °C with optimum growth around 25–30 °C in medium with pH between 6 and 7. Growth requires NaCl concentration (0–6% w/v) with optimal growth achieved at 3 and 4% NaCl for most species. Genome size of the species ranges between 1.6–2.1 Mbp and G+C content ranging between 35.4–44.0 mol%. The majority of known species can utilize D-glucose, D-fructose, D-mannose, lactose to produce lactic acid and CO_2 gas as the end products. Species from this

genus are used in dairy industries to produce aroma. Members of this genus form a monophyletic clade in phylogenetic trees based on concatenated sequences for large datasets of core proteins. In addition, they can be reliably distinguished from all other *Leuconostocaceae* and *Lactobacillaceae* genera by the shared presence of 5 identified CSIs (Table 1) in the following four proteins: BMP family protein, copper resistance protein, RNA binding transcriptional accessory protein and universal stress protein. These CSIs, in most cases, are exclusively shared by either all or most members of this genus.

The type species is *Leuconostoc mesenteroides* (Approved Lists) [10].

EMENDED DESCRIPTION OF THE GENUS *FRUCTOBACILLUS* ENDO AND OKADA 2008

Fructobacillus (Fruc.to.ba.cil'lus. N.L. masc. n. *Fructobacillus*, arbitrarily derived from fructose and *Lactobacillus*, intended to mean fructose-loving lactic acid-producing bacillus).

The description of this genus is modified from the original description by Endo *et al.* [8]. Cells are facultatively anaerobic, short rod-shaped, non-spore-forming, non-motile bacteria. Catalase activity varies between species. Members are heterofermentative and produce acetic acid, CO_2 and lactic acid from D-glucose supplemented with electron acceptors and D-fructose. They can be differentiated from other genera based on their fructophillic metabolism indicating a preference for D-fructose. Temperature range for growth is 5–40 °C with optimal growth at around 30 °C. The pH range for the growth of these species is between pH 4–8 with an optimum around pH 6.5. These bacteria have been isolated from fructose-rich environments ranging from flowers to fruits. Most species require NaCl (2.5–8.0% w/v). The genome size of the known *Fructobacillus* species ranges between 1.30–1.70 Mbp and their G+C content ranges from 43.90 to 44.70 mol%. Members of this genus form a monophyletic clade in phylogenetic trees based on 16S rRNA gene sequences and concatenated sequences for several large datasets of proteins. In addition, members of this genus can reliably be distinguished from all other *Leuconostocaceae* and *Lactobacillaceae* genera by the five CSIs described in this work (Table 1), found within the following proteins: Asp-tRNA(Asn)/Glu tRNA(Gln) amidotransferase subunit (GatB), xanthine phosphoribosyltransferase, ABC transporter ATP-binding protein/permease, NCS2 family nucleobase:cation symporter and ribonuclease J. These CSIs, in most cases, are exclusively shared by either all or most members of this genus.

The type species is *Fructobacillus fructosus* [8]

EMENDED DESCRIPTION OF THE GENUS *OENOCOCCUS* DICKS *ET AL.* 2015

Oenococcus (Oe.no.coc'cus. Gr. masc. n. *oînos*, wine; N.L. masc. n. *coccus*, berry; from Gr. masc. n. *kokkos*, grain; N.L. masc. n. *Oenococcus*, coccus from wine).

The description of this genus is modified from that given by Dicks *et al.* [11]. Cells are Gram-positive ellipsoidal cocci, usually appear in pairs. Can grow either in anaerobic or aerobic conditions. Obligately heterofermentative, non-motile, non-spore-forming, and oxidase- and catalase-negative. Member species have been isolated from various alcoholic beverages which is attributed to the ability of most species to undergo malolactic fermentation needed for alcohol fermentation. Most species are mesophilic and require NaCl (0–2.5% w/v). Due to their acidophilic nature, they grow at pH values ranging from pH 3.5 to 7.5, with optimal pH between 6.0–6.8. Temperature range for growth is 5–40 °C, with optimum growth at around 25–30 °C. The DNA G+C content ranges from 37.60 to 42.70 mol%. Members of this genus form a monophyletic clade in 16S RNA gene sequences and phylogenetic trees based on concatenated sequences for several large datasets of proteins. In addition, the members of this genus can reliably be distinguished from all other *Leuconostocaceae* and *Lactobacillaceae* species by the 13 CSIs described in this work in the following proteins: DNA-directed-RNA polymerase subunit beta, two different CSIs in preprotein translocase subunit SecY, 50S ribosomal protein L13, molecular chaperone DnaK, riboflavin kinase, Sua5/YciO/YrdC/YwlC family protein, RluA family pseudouridine synthase, amino acid permease, TatD family hydrolase, YidC/Oxa1 family membrane protein insertase, class I SAM-dependent RNA methyltransferase and GTPase HflX. The described CSIs in most cases are exclusively shared by either all or most members of this genus.

The type species is *Oenococcus oeni* [11].

EMENDED DESCRIPTION OF THE GENUS *WEISSELLA* COLLINS *ET AL.* 1994

Weissella (Weiss.el'la. N.L. fem. dim. n. *Weissella*, named after Norbert Weiss, a German microbiologist known for his many research contributions to the taxonomy of the lactic acid bacteria).

The description of this genus is partially based on the original description by Collins *et al.* [12] and emended by Padonou *et al.* [24]. Some other characteristics of this genus are reviewed by Björkroth *et al.* [5]. Cells are Gram-positive, obligately heterofermentative, non-spore-forming, non-motile short rods or cocci. Growth occurs at pH ranging from pH 3 to 8 and at 10–37 °C (optimum growth for most species at 18–25 °C). Genome sizes range between 1.33–2.51 Mbp and their G+C content vary between 35.40–45.40 mol%. Although these species are generally non-pathogenic, some *Weissella* species (*W. ceti*, *W. viridescens*,

W.confusa, *W.cibaria*) are opportunistic bacteria infecting in post-operative patients as well as some animals. Many of these species are unable to hydrolyse arginine and have been isolated from fermenting meat, dairy and vegetables. Members of this genus form a monophyletic clade, distinct from all other bacteria including those from the genus *Periweissella*, in 16S rRNA gene tree and in a phylogenetic tree based on concatenated sequences for large datasets of core genome proteins. In addition, members of this genus can be reliably distinguished from all other *Leuconostocaceae* and *Lactobacillaceae* genera, including *Periweissella*, by the shared presence of six CSIs identified in the present work (Table 2) in the following proteins: phospho-*N*-acetylmuramoyl-pentapeptide-transferases, APC family permease, alanine tRNA ligase, cytochrome d ubiquinol oxidase subunit II, response regulator transcription factor and endonuclease MutS2. These CSIs, in most cases, are exclusively shared by either all or most members of this genus.

The type species is *Weissella viridescens* Collins *et al.* [61]

DESCRIPTION OF *PERIWEISSELLA* GEN. NOV.

Periweissella (Pe.ri.weiss.el'la. Gr. prep. *peri*, about, around or nearby; N.L. fem. dim. n. *Weissella*, a bacterial genus named after Norbert Weiss, a German microbiologist; N.L. fem. dim. n. *Periweissella*, a genus about or nearby *Weissella*)

Cells are Gram-positive, obligately heterofermentative, non-spore-forming, short rods or cocci. Most species within this genus are non-motile apart from *W. beninensis*. Growth occurs in the presence of 0–5% NaCl (w/v) in the pH range from pH 3.9 to 9.0 (optimum, pH 6.0–7.0) across different species. Genome sizes range between 1.80–3.10 Mbp and G+C content varies between 35.40–41.10 mol%. Most of the colonies grow in temperatures ranging from 15 to 37 °C (optimum, 28–30 °C). Several of these species are able to hydrolyse arginine and have been isolated from fermenting cocoa or cassava. Members of this genus form a monophyletic clade, distinct from all other bacteria including those from the genus *Weissella*, in a 16S rRNA gene tree and in a phylogenetic tree based on concatenated sequences for large datasets of core genome proteins. In addition, members of this genus can be reliably distinguished from all other *Leuconostocaceae* and *Lactobacillaceae* genera, including *Weissella*, by the shared presence of five CSIs identified in the present work (Table 2) in the following proteins: amidophosphoribosyl transferase protein, DEAD/DEAH box helicase, ArgR family transcriptional regulator, Flp pilus assembly complex ATPase component (TadA) and hydroxyethylthiazole kinase. Most of these CSIs are exclusively shared by either all or most members of this genus.

The type species is *Periweissella ghanensis*

DESCRIPTION OF *PERIWEISSELLA GHANENSIS* COMB. NOV.

Periweissella ghanensis (gha.nen'sis. N.L. fem. adj. *ghanensis*, pertaining to Ghana).

Basonym: *Weissella ghanensis* De Bruyne *et al.* 2008.

The description of this species is as provided by De Bruyne *et al.* [64] for *Weissella ghanensis.*

Type strain: 215^{T}=DSM 19935^{T}=LMG 24286^{T}

DESCRIPTION OF *PERIWEISSELLA BENINENSIS* COMB. NOV.

Periweissella beninensis (be.nin.en'sis. N.L. fem. adj. *beninensis* pertaining to Benin, where the type strain was isolated).

Basonym: *Weissella beninensis* Padonou *et al.* 2010

The description of this species is as provided by Padonou *et al.* [24] for *Weissella beninensis.*

Type strain: 2L24P13^{T}=DSM 22752^{T}=LMG 25373^{T}

DESCRIPTION OF *PERIWEISSELLA CRYPTOCERCI* COMB. NOV.

Periweissella cryptocerci(cryp.to.cer'ci. N.L. gen. n. *cryptocerci*, of *Cryptocercus*, a genus of insect from which the species was isolated).

Basonym: *Weissella cryptocerci* Heo *et al.* 2019

The description of this species is as provided by Heo *et al.* [27] for *Weissella cryptocerci.*

Type strain: 26KH-42^{T}=KACC 18423^{T}=NBRC 113066^{T}

DESCRIPTION OF *PERIWEISSELLA FABALIS* COMB. NOV.

Periweissella fabalis (fa.ba'lis. L. fem. adj. *fabalis*, of or belonging to beans).

Basonym: *Weissella fabalis* Snauwaert *et al.* 2013

The description of this species is as provided by Snauwaert *et al.* [26] for *Weissella fabalis.*

Type strain: CCUG 61472^T=DSM 28407^T=LMG 26217^T=M75^T

DESCRIPTION OF *PERIWEISSELLA FABARIA* COMB. NOV.

Periweissella fabaria (fa.ba'ri.a. L. fem. adj. *fabaria*, of or belonging to beans).

Basonym: *Weissella fabaria* De Bruyne *et al.* 2010

The description of this species is as provided by De Bruyne *et al.* [25] for *Weissella fabaria*.

Type strain: 257^T=DSM 21416^T=LMG 24289^T

Funding information
This work was supported by a Discovery Research Grant (RGPIN-2019-06397) and Alliance Research Grant from Natural Science and Engineering Research Council of Canada, and a grant from the Ontario Research Fund.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Bello S, Rudra B, Gupta RS. Phylogenomic and comparative genomic analyses of Leuconostocaceae species: identification of molecular signatures specific for the genera Leuconostoc, Fructobacillus and Oenococcus and proposal for a novel genus Periweissella gen. nov. *Figshare* 2022. DOI: 10.6084/M9.FIGSHARE.18866273.V1.
2. Schleifer K-H. Family V. *Leuconostocaceae* fam. nov. In: Jones D, Krieg NR, Ludwig W and Rainey FA (eds). *Bergey's Manual of Systematic Bacteriology (The Firmicutes)*. New York: Springer; 2009.
3. Nieminen TT, Säde E, Endo A, Johansson P, Björkroth J, *et al*. The family *Leuconostocaceae*. In: Rosenberg E, DeLong EF and Lorey S (eds). *The Prokaryotes: Firmicutes and Tenericutes*. Springer-Verlag; 2014.
4. Holzapfel WH, Bjorkroth JA, Dicks LMT. Genus I. *Leuconostoc* van Tiegehem 1878, 198^{AL} emend. mut. char. (Hucker and Pederson 1930), 66^{AL}. In: De Vos P, Jones D, Krieg NR, Ludwig W and Rainey FA (eds). *Bergey's Manual of Systematic Bacteriology (The Firmicutes)*. New York: Springer; 2009. pp. 624–636.
5. Bjorkroth JA, Dicks LMT, Holzapfel WH. Genus III. *Weissella* Collins, Samelis, Metaxopoulos and Wallbanks 1994, 370VP (effective publication: Collins, Samelis, Metaxopoulos, and Wallbanks 1993, 597). In: De Vos P, Jones D, Krieg NR, Ludwig W and Rainey FA (eds). *Bergey's Manual of Systematic Bacteriology (The Firmicutes)*. New York: Springer; 2009. pp. 643–649.
6. Mokoena MP. Lactic acid bacteria and their bacteriocins: classification, biosynthesis and applications against uropathogens: a mini-review. *Molecules* 2017;22:E1255.
7. Praet J, Meeus I, Cnockaert M, Houf K, Smagghe G, *et al*. Novel lactic acid bacteria isolated from the bumble bee gut: *Convivina intestini* gen. nov., sp. nov., *Lactobacillus bombicola* sp. nov., and *Weissella bombi* sp. nov. *Antonie van Leeuwenhoek* 2015;107:1337–1349.
8. Endo A, Okada S. Reclassification of the genus *Leuconostoc* and proposals of *Fructobacillus fructosus* gen. nov., comb. nov., *Fructobacillus durionis* comb. nov., *Fructobacillus ficulneus* comb. nov. and *Fructobacillus pseudoficulneus* comb. nov. *Int J Syst Evol Microbiol* 2008;58:2195–2205.
9. Van Tieghem P. Sur la gomme de sucrerie (*Leuconostoc mesenteroides*). *Annales des Sciences Naturelles Botanique* 1878;7:180–203.
10. Sneath PHA, McGowan V, Skerman VBD. Approved lists of bacterial names. *Int J Syst Bacteriol* 1980;30:225–420.
11. Dicks LM, Dellaglio F, Collins MD. Proposal to reclassify *Leuconostoc oenos* as *Oenococcus oeni* [corrig.] gen. nov., comb. nov. *Int J Syst Bacteriol* 1995;45:395–397.
12. Collins MD, Samelis J, Metaxopoulos J, Wallbanks S. Taxonomic studies on some leuconostoc-like organisms from fermented sausages: description of a new genus Weissella for the *Leuconostoc paramesenteroides* group of species. *J Appl Bacteriol* 1993;75:595–603.
13. Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, *et al*. A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int J Syst Evol Microbiol* 2020;70:2782–2858.
14. Martinez-Murcia AJ, Collins MD. A phylogenetic analysis of the genus *Leuconostoc* based on reverse transcriptase sequencing of 16 S rRNA. *FEMS Microbiol Lett* 1990;58:73–83.
15. Endo A, Maeno S, Tanizawa Y, Kneifel W, Arita M, *et al*. Fructophilic lactic acid bacteria, a unique group of fructose-fermenting microbes. *Appl Environ Microbiol* 2018;84:19.
16. Endo A, Tanizawa Y, Tanaka N, Maeno S, Kumar H, *et al*. Comparative genomics of *Fructobacillus* spp. and *Leuconostoc* spp. reveals niche-specific evolution of *Fructobacillus* spp. *BMC Genomics* 2015;16:1117.
17. Salvetti E, Harris HMB, Felis GE, O'Toole PW. Comparative genomics of the genus *Lactobacillus* reveals robust phylogroups that provide the basis for reclassification. *Appl Environ Microbiol* 2018;84:17.
18. Sun Z, Harris HMB, McCann A, Guo C, Argimón S, *et al*. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nat Commun* 2015;6:8322.
19. Duar RM, Lin XB, Zheng J, Martino ME, Grenier T, *et al*. Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. *FEMS Microbiol Rev* 2017;41:S27–S48.
20. Zheng J, Ruan L, Sun M, Gänzle M. A genomic view of lactobacilli and pediococci demonstrates that phylogeny matches ecology and physiology. *Appl Environ Microbiol* 2015;81:7233–7243.
21. Chelo IM, Zé-Zé L, Tenreiro R. Congruence of evolutionary relationships inside the *Leuconostoc-Oenococcus-Weissella* clade assessed by phylogenetic analysis of the 16S rRNA gene, dnaA, gyrB, rpoC and dnaK. *Int J Syst Evol Microbiol* 2007;57:276–286.

22. Endo A, Irisawa T, Futagawa-Endo Y, Sonomoto K, Itoh K, *et al*. *Fructobacillus tropaeoli* sp. nov., a fructophilic lactic acid bacterium isolated from a flower. *Int J Syst Evol Microbiol* 2011;61:898–902.
23. Nel S, Davis SB, Endo A, Dicks LMT. Phylogenetic analysis of *Leuconostoc* and *Lactobacillus* species isolated from sugarcane processing streams. *Microbiologyopen* 2020;9:e1065.
24. Padonou SW, Schillinger U, Nielsen DS, Franz CMAP, Hansen M, *et al*. *Weissella beninensis* sp. nov., a motile lactic acid bacterium from submerged cassava fermentations, and emended description of the genus *Weissella*. *Int J Syst Evol Microbiol* 2010;60:2193–2198.
25. De Bruyne K, Camu N, De Vuyst L, Vandamme P. *Weissella fabaria* sp. nov., from a Ghanaian cocoa fermentation. *Int J Syst Evol Microbiol* 2010;60:1999–2005.
26. Snauwaert I, Papalexandratou Z, De Vuyst L, Vandamme P. Characterization of strains of *Weissella fabalis* sp. nov. and *Fructobacillus tropaeoli* from spontaneous cocoa bean fermentations. *Int J Syst Evol Microbiol* 2013;63:1709–1716.
27. Heo J, Hamada M, Cho H, Weon H-Y, Kim J-S, *et al*. *Weissella cryptocerci* sp. nov., isolated from gut of the insect *Cryptocercus kyebangensis*. *Int J Syst Evol Microbiol* 2019;69:2801–2806.
28. Li YQ, Tian WL, Gu CT. *Weissella sagaensis* sp. nov., isolated from traditional Chinese yogurt. *Int J Syst Evol Microbiol* 2020;70:2485–2492.
29. Hyun D-W, Lee J-Y, Sung H, Kim PS, Jeong Y-S, *et al*. *Brevilactibacter coleopterorum* sp. nov., isolated from the intestine of the dark diving beetle *Hydrophilus acuminatus*, and *Weissella coleopterorum* sp. nov., isolated from the intestine of the diving beetle *Cybister lewisianus*. *Int J Syst Evol Microbiol* 2021;71.
30. Oh SJ, Shin N-R, Hyun D-W, Kim PS, Kim JY, *et al*. *Weissella diestrammenae* sp. nov., isolated from the gut of a camel cricket (*Diestrammena coreana*). *Int J Syst Evol Microbiol* 2013;63:2951–2956.
31. Parte AC. LPSN - the list of prokaryotic names with standing in nomenclature. *Int J Syst Evol Microbiol* 2018;68:1825–1829.
32. Mukherjee S, Seshadri R, Varghese NJ, Eloe-Fadrosh EA, Meier-Kolthoff JP, *et al*. 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat Biotechnol* 2017;35:676–683.
33. Whitman WB, Woyke T, Klenk H-P, Zhou Y, Lilburn TG, *et al*. Genomic encyclopedia of bacterial and archaeal type strains, phase III: the genomes of soil and plant-associated and newly described type strains. *Stand Genomic Sci* 2015;10:26.
34. Wu L, McCluskey K, Desmeth P, Liu S, Hideaki S, *et al*. The global catalogue of microorganisms 10K type strain sequencing project: closing the genomic gaps for the validly published prokaryotic and fungi species. *Gigascience* 2018;7.
35. Konstantinidis KT, Tiedje JM. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr Opin Microbiol* 2007;10:504–509.
36. Gupta RS. Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* 1998;62:1435–1491.
37. Gupta RS. Identification of conserved indels that are useful for classification and evolutionary studies. In: Goodfellow M, Sutcliffe IC and Chun J (eds). *Bacterial Taxonomy, Methods in Microbiology*, vol. 41. London: Elsevier; 2014. pp. 153–182.
38. Gupta RS. Impact of genomics on clarifying the evolutionary relationships amongst mycobacteria: identification of molecular signatures specific for the tuberculosis-complex of bacteria with potential applications for novel diagnostics and therapeutics. *High Throughput* 2018;7:E31.
39. Gupta RS. Impact of genomics on the understanding of microbial evolution and classification: the importance of Darwin's views on classification. *FEMS Microbiol Rev* 2016;40:520–553.
40. Gupta RS. Microbial taxonomy: how and why name changes occur and their significance for (clinical) microbiology. *Clin Chem* 2021;68:134–137.
41. Patel S, Gupta RS. A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: Proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. *Int J Syst Evol Microbiol* 2020;70:406–438.
42. Adeolu M, Alnajar S, Naushad S, S Gupta R. 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. *Int J Syst Evol Microbiol* 2016;66:5575–5599.
43. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012;28:3150–3152.
44. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;7:539.
45. Sonnhammer EL, Eddy SR, Birney E, Bateman A, Durbin R. Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res* 1998;26:320–322.
46. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;25:1972–1973.
47. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;5:e9490.
48. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 2001;18:691–699.
49. Le SQ, Lartillot N, Gascuel O. Phylogenetic mixture models for proteins. *Philos Trans R Soc Lond B Biol Sci* 2008;363:3965–3976.
50. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol* 2018;35:1547–1549.
51. Thompson CC, Chimetto L, Edwards RA, Swings J, Stackebrandt E, *et al*. Microbial genomic taxonomy. *BMC Genomics* 2013;14:913.
52. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, *et al*. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res* 2014;42:D643-8.
53. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 1993;10:512–526.
54. Naushad HS, Lee B, Gupta RS. 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* 2014;64:366–383.
55. Bello S, Rudra B, Gupta RS. Supplemental data: phylogenomic and comparative genomic analyses on *Leuconostocaceae* species: identification of molecular signatures specific for the genera *Leuconostoc*, *Fructobacillus* and *Oenococcus* and proposal for a novel genus *Periweissella* gen. nov. *Int J Syst Evol Microbiol* 2022:1–108.
56. Qin Q-L, Xie B-B, Zhang X-Y, Chen X-L, Zhou B-C, *et al*. A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol* 2014;196:2210–2215.
57. Dobritsa AP, Samadpour M. Reclassification of *Burkholderia insecticola* as *Caballeronia insecticola* comb. nov. and reliability of conserved signature indels as molecular synapomorphies. *Int J Syst Evol Microbiol* 2019;69:2057–2063.
58. Parker C, Tindall BJ, Garrity GM. International Code of Nomenclature of Prokaryotes. *Int J Syst Evol Microbiol* 2019;69:S1–S111.
59. Chelo IM, Zé-Zé L, Tenreiro R. Genome diversity in the genera *Fructobacillus*, *Leuconostoc* and *Weissella* determined by physical and genetic mapping. *Microbiology (Reading)* 2010;156:420–430.

60. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 2019:btz848.
61. Collins MD, Samelis J, Metaxopoulous J, Wallbanks S. Validation of the publication of new names and new combinations previously effectively published outside the IJSB: List No. 49. *Int J Syst Bacteriol* 1994;44:370–371.
62. Rokas A, Holland PW. Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol* 2000;15:454–459.
63. Barbour AG, Adeolu M, Gupta RS. Division of the genus Borrelia into two genera (corresponding to Lyme disease and relapsing fever groups) reflects their genetic and phenotypic distinctiveness and will lead to a better understanding of these two groups of microbes (Margos et al. (2016) There is inadequate evidence to support the division of the genus Borrelia. *Int J Syst Evol Microbiol* 2017;67:2058–2067.
64. De Bruyne K, Camu N, Lefebvre K, De Vuyst L, Vandamme P. *Weissella ghanensis* sp. nov., isolated from a Ghanaian cocoa fermentation. *Int J Syst Evol Microbiol* 2008;58:2721–2725.