

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 Weissella 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

†These authors contributed equally to this work

Two supplementary tables and 47 figures are available with the online version of this article 005284 © 2022 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution NonCommercial License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

Author affiliations: 1Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, L8N 325, 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.

acid, CO., 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 BLAST 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 maximumlikelihood 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

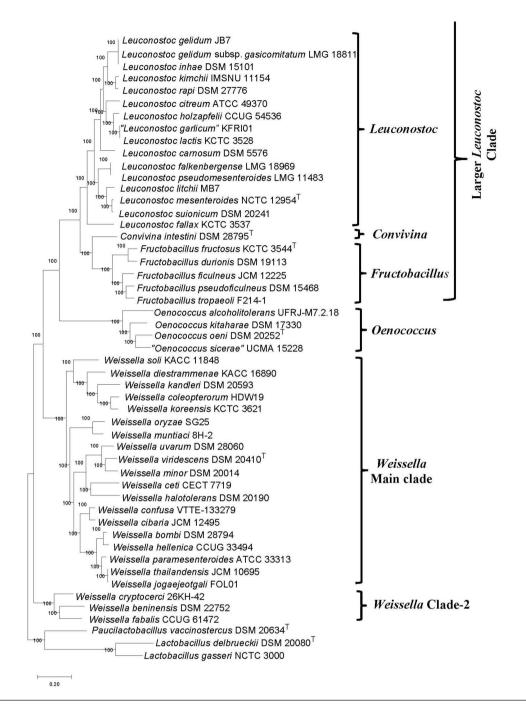


Fig. 1. A bootstrapped maximum-likelihood tree for 47 genomes sequenced *Leuconostocaceae* species based on concatenated sequences for 498 core proteins. The statistical support values for different branches are indicated on the nodes. This tree was rooted by using species from the genus *Lactobacillus*. 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

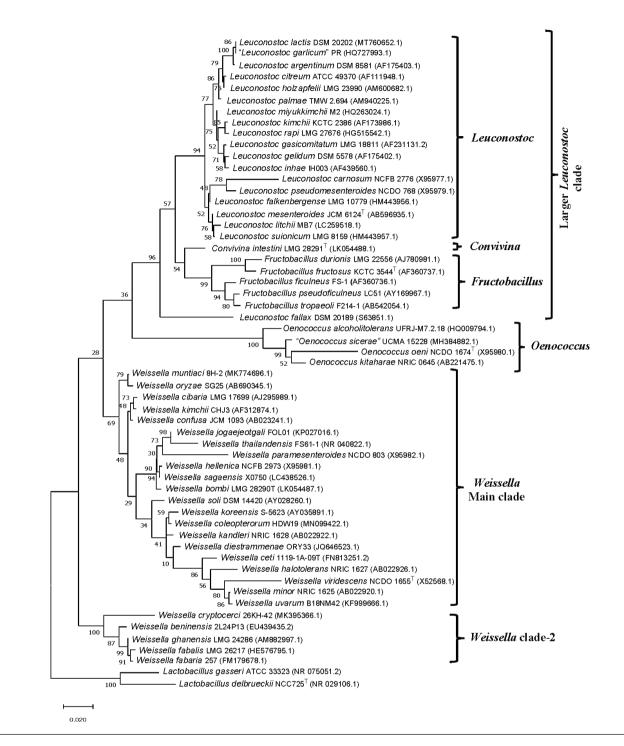


Fig. 2. A maximum-likelihood phylogenetic tree based on 16S rRNA gene sequences for the type strains of all validly published *Leuconostocaceae* species. *Lactobacillus* species *L. delbrueckii* and *L. gasseri* were used to root the tree. The accession numbers of the 16S rRNA gene sequences are given 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

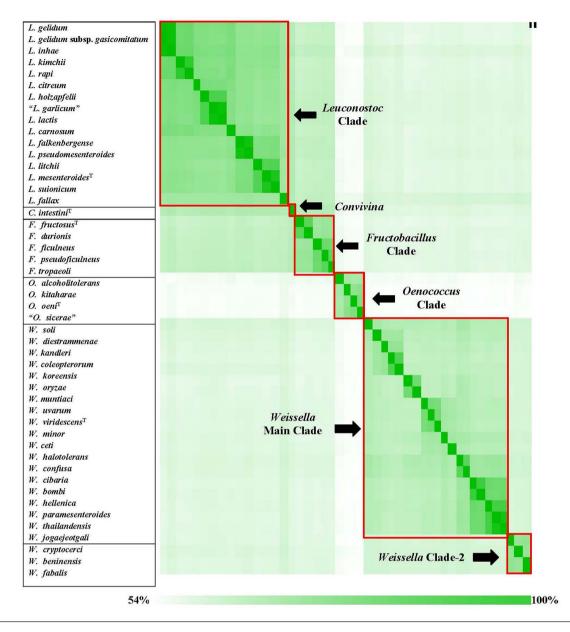


Fig. 3. A matrix indicating the pairwise percentage average amino acid identities of the species from different genera within the family *Leuconostocaceae*. Genome pairs sharing higher amino acid identity are shaded more darkly (green). The regions of the matrix corresponding to different clades have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

(a)			205	242
(4)	Leuconostoc carnosum	WP 135197391	205 DEQEVYAGYYDYSESIKSLLNNK	243 FR ILAVNRGEKDGVL
	Leuconostoc citreum	WP 048700160	-PGQAFPT-TQ	
	Leuconostoc falkenbergense	WP_188356507	KI-EQFDQKE	IDV
	Leuconostoc fallax	WP_010007576	QI-QQFTVPH-KQ-G	
	"Leuconostoc garlicum" Leuconostoc gelidum	WP_077282923 WP 060391542	-PGQAFPT-TQ EQN-PT-RD	
	Leuconostoc holzapfelii	WP 168677018	-PGQAPT-NA-Q	
I au con ortoo	Leuconostoc inhae	WP 220734074	I-GQND	
Leuconostoc	Leuconostoc kimchii	WP_013102348	AI-EQPL-TIIE	
(17/17)	Leuconostoc lactis	WP_216802510	-PGQAFPT-TQ	
	Leuconostoc litchii Leuconostoc mesenteroides ^T	WP_148604053 WP 129531556	KEQFN-TV KEQFT-TV	
	Leuconostoc miyukkimchii	WP 220739328		
	Leuconostoc palmae	WP 220742024	-TNGQG-Q-TQ	
	Leuconostoc pseudomesenteroides	HCU41910	KI-EQFDQKE	
	Leuconostoc rapi	WP_204769179	AI-EQPL-TIIE	
	Leuconostoc suionicum (Weissella viridescens ^T	WP_072613798	KEQFT-T -P-KELF-TPLPQVHQ	
	Weissella ceti	WP_191136251 WP_213409686	-PI-KLF-MP-TK-A-HQ	EIHE VA
	Weissella coleopterorum	WP 166011275	-PKDLF-QKV-QVQ-HQ	MA
Species from other	Weissella jogaejeotgali	WP_083603405	G-FELFDVPV-KVHQ	ISE-I-
-	Weissella koreensis	WP_190275617	-PKSILF-QKQIQ-HQ	A-I-
Leuconostocaceae	Weissella muntiaci	WP_148623856	-PFRLF-QTNVA-HQ	IE
genera	Weissella oryzae Weissella paramesenteroides	WP_045476986 WP_140837586	-P-AIFQLF-QTNVA-HQ G-FELFDSLV-KIQ-HQ	VIEA IE-I-
	Weissella soli	WP 147153127	-Q-RQLF-QQRVR-HQ	VLEN
	Weissella thailandensis	WP 115471723	G-FELFDVPV-KVHQ	ISE-I-
	Companilactobacillus nuruki	WP_217350201	QETFQQTL-K-PPYR	IEKI-
	Lentilactobacillus farraginis	WP_056983607	AAT-QQF-Q-VREVPSYR	E-I-
Other Bacteria	Limosilactobacillus portuensis Levilactobacillus brevis ^T	WP_191913256 ARN90649	LKT-EQFTSPV-D-SSYQ	VIE-I- VIREKI-
(0/500)	Limosilactobacillus frumenti	WP 057748300	QTEFTQ-L-QVPPYR LKT-EQFTSPVNE-SSYQ	IE
	Companilactobacillus musae	WP 176580211	QTFEQAQIPPYR	VIREKI-
		_		
(b)				
	-		323	362
	Fructobacillus tropaeoli	WP_203617595	TQTLAMANFFDAAVNAGADSK SAAK	
Fructobacillus	Fructobacillus durionis	WP_091501315	SNT-KS A	
(05/05)	Fructobacillus fructosus ^T Fructobacillus ficulneus	WP_010017521 GAP00591	SNT-KE A STIASD- A	
(05/05)	Fructobacillus pseudoficulneus	WP 059375808	G G	
	Fructobacillus sp. EFB-N1	WP 047974214		
Fructobacillus	Fructobacillus sp. M1-10	WP_213819472	A A	
unnamed sp.	Fructobacillus sp. M1-21	WP_213792376	SNT-AN P SNT-AN P	
	Fructobacillus sp. \$1-1 Convivina intestini ^T	WP_213821901 WP 089938240	D-Y-QT-QEP-	-A
	Oenococcus kitaharae	WP 007744614	EQT-ANP-	-A
	Oenococcus oeni ^T	WP 071443329	EY-QV-DDP-	-T
	Leuconostoc citreum	WP_040190527	D-YTISA-	-AFM-
	Leuconostoc falkenbergense	WP_188356875	P-	-AM-
	Leuconostoc fallax Leuconostoc holzapfelii	WP_010008020 WP 168675916	MED-Y-QTS-GEP- A-	М-
	Leuconostoc inhae	WP 220735059	SD-YT-TA-	-AFM-
	Leuconostoc lactis	WP_195330200	YT-AVA-	-АМ-
	Leuconostoc litchii	WP_148606638	SD-YT-AP-	-тМ-
~	Leuconostoc mesenteroides ^T	WP_036092096	SD-YT-SSP-	-AM-
Species from other	Leuconostoc palmae Leuconostoc suionicum	WP_220741753 WP 193324991	A- SD-YT-SSP-	-AM- -AM-
Leuconostocaceae	Weissella beninensis	WP 205143122	E-SDT-KEA-	-TMM-
genera	Weissella coleopterorum	WP_166011108	DQT-ELA-	-AF
	Weissella confusa	WP_195226634	P-	-SF
	Weissella fabalis Naissella issessionterli	WP_168722316 WP_075269373	E-SDT-AA-	-TMM-
	Weissella jogaejeotgali Weissella kandleri	WP_075269373 WP_057754749	P- T-AENP-	-TMF -ASMF
	Weissella koreensis	WP 013989742	DQT-EFA-	-AF
	Weissella minor	WP_057787281	E-SDT-SENA-	-AM
	Weissella soli	WP_147152302	ET-TEA-	-AF
	Weissella thailandensis	WP_115471469	ET-SFP-	-TF
	Weissella viridescens ^T	WP_124943404 WP_035455156	E-ST-ANP- -N-KE-SDHNP-	-AM
	$Agrilactobacillus composti^{T}$	WP_035455156	-N-KE-SDHNP-	Q-S-W-M-E
	Agrilactobacillus composti ^T Escherichia coli ^T	WP_035455156 MQJ50554	-N-KE-SDHNP- -L-KE-SDESTIEHV-	Q-S-W-M-E LTS-W-M-GE
Other Bacteria	$Agrilactobacillus composti^{T}$	WP_035455156	-N-KE-SDHNP-	Q-S-W-M-E
	Agrilactobacillus composti ^T Escherichia coli ^T Furfurilactobacillus rossiae ^T	WP_035455156 MQJ50554 WP_161002915	-N-KE-SDHNP- -L-KE-SDESTIEHV- KE-SDEAQA-	Q-S-W-M-E LTS-W-M-GE LTSQ
Other Bacteria (0/500)	Agrilactobacillus composti ^T Escherichia coli ^T Furfurilactobacillus rossiae ^T Limosilactobacillus fermentum ^T Paucilactobacillus kaifaensis Pediococcus damnosus ^T	WP 035455156 MQJ50554 WP 161002915 WP 021349526 WP 137597285 WP 062904134	-N-KE-SDHNP- -L-KE-SDESTIEHV- KE-SDEAQA- KEDEEKD-G-A-	Q-S-W-M-E LTS-W-M-GE LTSQ KMNS
	Agrilactobacillus composti ^T Escherichia coli ^T Furfurilactobacillus rossiae ^T Limosilactobacillus fermentum ^T Paucilactobacillus kaifaensis	WP_035455156 MQJ50554 WP_161002915 WP_021349526 WP_137597285	-N-KE-SDESTIEHP- -L-KE-SDESTIEHV- KE-SDEAQA- KE-D-EE-KD-G-A- KE-SD-EEAKA-	Q-S-W-M-E LTS-W-M-G-E LTSQ KMNS MTMS

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.

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. \$16		
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,	
RluA family pseudouridine synthase	WP_089937871	3 aa Ins	234-376	Fig. S19	Convivina, Fructobacillus	
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

	(a)			146	183
		Fructobacillus tropaeoli	WP_203618360	SGIDAATVSAVEPVWF	HQQQIDTFQTNLEATLVLADTG
		Fructobacillus durionis	WP_091501401	<u>E</u>	QDKVCSE-D-D
		Fructobacillus ficulneus	GAP00175	SDQ-I	EHKQG-TAK
		Fructobacillus fructosus ^T	WP_187753374	SE ST	KDRINTE-T-D
Fru	ictobacillus and	Fructobacillus pseudoficulneus Fructobacillus sp. EFB-N1	WP_059377114 WP 047974964	ST	-HH-TEP-A
	Convivina	Fructobacillus sp. M1-10	WP 213819411	ss	ENKAE-S-N
	Convertina	Fructobacillus sp. M1-21	WP_213792434	SE	QDHRCSE-S-DI
		Fructobacillus sp. M2-14	WP_213809375	LE	LHKK-EE-S
		Fructobacillus sp. S1-1	WP_213822272		QD-RCSE-S-D
		Convivina intestini ^T Leuconostoc carnosum	WP_089937892 WP 014974482	ASKD-I-Y	ENKT-EMS V KNEEMTP-SML-TGI
		Leuconostoc citreum	WP 004899967		V KNDLPTP-KMS-TGV
		Leuconostoc fallax	WP 010007621		V KNKIMQS-DMH-SGI
		Leuconostoc gelidum	WP_220745791		V KH-ELTP-TMS-TGV
		Leuconostoc inhae	WP_220734391		V KHLTP-KMS-TGV
		Leuconostoc lactis Leuconostoc litchii	WP_195218690 WP 148603876		V KNETLTA-KMT-TGV I KNKTMEK-DMSG
		Leuconostoc mesenteroides ^T	WP 050883859		I KNKTMEK-EMSG
		Leuconostoc miyukkimchii	WP 220739952		I KHLTP-TMS-TGV
		Leuconostoc palmae	WP 220742163		I KNIEMTP-KMS-TGV
Spe	ecies from other	Leuconostoc rapi	WP_204769102		I KHK-LAP-TMS-TGV
Leu	uconostocaceae	Leuconostoc suionicum	WP_211636580		I KNETMEK-EMAG
		Weissella bombi	SCB74220 WP_166010833		I KGK-PEKINMS-DII I KHEVTQP-SLNGVI
	genera	Weissella coleopterorum Weissella cryptocerci	WP_133362375		I KNKELTP-KIDIPGYI
		Weissella kandleri	KRN75766		I KNELTQP-ELNFI
		Weissella koreensis	WP_006845278		I KHKITQS-SLQGVI
		Weissella minor	KRN77063		I KK-T-E-LEMDGY-II
		Weissella muntiaci	WP_148621833		V KG-KPMPISMS-DII
		Weissella oryzae Weissella soli	WP_027698501 WP 070229879		M KGNLPEPIAMD-DII V KGS-PE-IPMDI
		Weissella uvarum	WP 205145410		I KNET-E-L-MGY-II
		Weissella viridescens ^T	WP 124942993	LCASDV	I KKETTQ-LEMTGY-II
		Dellaglioa algida ^T	SOB49463		I KHKEAYSVPL-IV
	Other Bacteria	Enterococcus faecalis ^T	WP 154213282		T KG-PF-Y-PL-ID-F-IV
		Secundilactobacillus oryzae	WP 034527713		V -N-GSTELNLSGAI
	(0/500)	Apilactobacillus ozensis	WP_056966347		I -GKKNEQIPF-INGY-IIS-
		Loigolactobacillus jiayinensis	WP_125551842	MT-STA-I	I KGHASYNLPISGNIS-
	(b)	•		84	127
		Fructobacillus durionis Fructobacillus ficulneus	WP_091502306 GA099778		DRES GELVKIHKAKLRGEVSYGM VQDT-NE-NI
	Fructobacillus .	Fructobacillus fructosus ^T	WP 010692107		K
	Test in the second se	Fructobacillus pseudoficulneus	WP 059378161		/QTT-KE-N
	(05/05)	Fructobacillus tropaeoli	WP_059393967		/Dт-ке-Qт
	Convivina	€Convivina intestini ^T	WP_089938433	QA	QE-KAVE-F
Larger Leuconostoc Clade		Leuconostoc carnosum	WP_014974763		NH MV-QA-NT-Q
ST 9		Leuconostoc citreum	WP_004907393		-HNT DDIIE-KQ-TT-F
Larger uconosi Clade		Leuconostoc falkenbergense Leuconostoc fallax	WP_188356213 WP 010007350		NHDTLSAVSQM
ar		Leuconostoc gelidum	11_010007000		
C II			WP 013231835		-NHIELSQVIE-N -NH-TNVLKPVD-N
Le		Leuconostoc inhae	WP_013231835 CUW17996	VKA	-NHIELSQVIE-N -NH-TNVLKPVD-N -NHETNVL-PVD-N
	Leuconostor	Leuconostoc inhae Leuconostoc kimchii	CUW17996 WP_013103236	VKA VKA ITA	NH-T NVLKPVD-N NHET NVL-PVD-N
	Leuconostoc •	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis	CUW17996 WP_013103236 WP_068850717	VKA VKA ITA VVKA	NH-T NVLKPVD-N NHET NVL-PVD-N KHDT ITLE-VTQD N -K-MTLTAM
	Leuconostoc •	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis Leuconostoc litchii	CUW17996 WP_013103236 WP_068850717 WP_148605255	VKA ITA VVKA VVKA	-NH-T NVLKPVD-N NHET NVL-PVD-N -KHDT ILE-VTQD NH-N IELKAVS
	Leuconostoc •	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis Leuconostoc litchii Leuconostoc mesenteroides ^T	CUW17996 WP_013103236 WP_068850717 WP_148605255 WP_011679295	VKA VKA V	NHET
	Leuconostoc «	Leuconstoc inhae Leuconstoc kimchii Leuconstoc lactis Leuconstoc litchii Leuconstoc mesenteroides ^T Leuconstoc miyukkimchii	CUW17996 WP_013103236 WP_068850717 WP_148605255 WP_011679295 WP_220741133		NH T NVLKPVD-N NHET NVL-PVD-N KHDT ITLE-VTQD NH K-MTLTAM NH +- TELKAVS NH IELNAVF NH IELKP
	Leuconostoc «	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis Leuconostoc litchii Leuconostoc mesenteroides ^T	CUW17996 WP_013103236 WP_068850717 WP_148605255 WP_011679295	VKA VKA VVKA WKA MKA FKA	NHET
	Leuconostoc •	Leuconstoc inhae Leuconstoc kimchii Leuconstoc lactis Leuconstoc litchii Leuconstoc misenteroides ^T Leuconstoc miyukkimchii Leuconstoc palmae Leuconstoc rapi Leuconstoc suionicum	CUW17996 WP_013103236 WP_068850717 WP_148605255 WP_011679295 WP_220741133 WP_22074291 WP_204770076 WP_072613143	VKA 	NH-T NVLKPVD-N NHET NVL-PVD-N KKDT TILE-VTQD NN-N TLELKAVS NH-N IELKAVF LKAPT IELKAVF LDAAT IELKY-EKN DAAT -K-ELKV-NQE-F -KHDT
	Leuconostoc •	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis Leuconostoc litchii Leuconostoc mesenteroides ^T Leuconostoc miyukkimchii Leuconostoc palmae Leuconostoc rapi Leuconostoc suionicum Genococcus kitaharae	CUW17996 WP_013103236 WP_06850717 WP_148605255 WP_011679295 WP_220741133 WP_220742291 WP_204770076 WP_0772613143 WP_007746618		NH-T WVLKPVD-N NHET NVL-PVD-N KHET NVL-PVQD N K-MTLTAM NH-N IELKAVF NH-N IELKAVF ND-N IELKAVF ND-N IELKAVF ND-N IELKAVF ND-N IELKAVF ND-N IELKAV N-N IELKVQE
	Leuconostoc •	Leuconostoc inhae Leuconostoc lactis Leuconostoc lactis Leuconostoc litchii Leuconostoc miyukkimchii Leuconostoc miyukkimchii Leuconostoc rapi Leuconostoc rapi Leuconostoc suionicum benococcus kitaharae Oenococcus oeni ^T	$\begin{array}{c} CU\overline{w}17996 \\ WP & 013103236 \\ WP & 068850717 \\ WP & 148605255 \\ WP & 011679295 \\ WP & 220741133 \\ WP & 220742291 \\ WP & 204770076 \\ WP & 072613143 \\ WP & 07746618 \\ WP & 171284862 \end{array}$		NH-T RVLKPVD-N NHET NUL-PVD-N KHDT ITLE-VTQD NH-N IELKAVF NH-N IELKAVF NH-N IELKAVF NAT IELKAVF NAT IELKAVF NAT IELKAVQE-F KHDT TLE-VTQD N-N-N TLE-VT-QD N-N-N TLE-VT-QD PNGKEATT-D-QI-Q PNGK
	Leuconostoc •	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis Leuconostoc litchii Leuconostoc miyukkimchii Leuconostoc palmae Leuconostoc rapi Leuconostoc suionicum Denococcus sitaharae Oenococcus sicerae	Cum17996 WP_013103236 WP_068850717 WP_148605255 WP_011679295 WP_220741133 WP_220742291 WP_220742291 WP_072613143 WP_0776618 WP_171284862 WP_1712848664		NH-T WVLKPVD-N NHET NUL-PVD-N KHDT ITLE-VTQD N IELKAVS NH-N IELKAVF LKAPT IELKAVF LDAAT TLELVAVF NN IELKAVQE-F NN IELKAVQE-F NN IELNAVQE-F NN IELNAVQE-F NN IELNAV NN IELNAV NN IELNAV NN IELNAV
	Leuconostoc •	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis Leuconostoc litchii Leuconostoc misenteroides ^T Leuconostoc miyukkimchii Leuconostoc rapi Leuconostoc rapi Leuconostoc suionicum Denococcus kitaharae Oenococcus soir ^T Oenococcus sicerae Weissella beninensis	CUW17996 WP_068850717 WP_148605255 WP_011679295 WP_22074291 WP_22074291 WP_22074291 WP_072613143 WP_007746618 WP_171284862 WP_128686804 WP_205143646		NH-T WVLKPVD-N NHET NVL-PVD-N NHET NVL-PVQD N-N- ITLE-VTQD N-N- IELKAVF N-N IELKAVF N-N IELKAVF N-N IELKAVF N-N IELKAVF N-N IELKAVF N-N IELKAV N-N IELKAV N-N
	Leuconostoc •	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis Leuconostoc litchii Leuconostoc miyukkimchii Leuconostoc palmae Leuconostoc rapi Leuconostoc suionicum Denococcus sitaharae Oenococcus sicerae	$\begin{array}{c} C \overline{w} 17996 \\ w P 013103236 \\ w P 068850117 \\ w P 148605255 \\ w P 011679295 \\ w P 2020741133 \\ w P 220742133 \\ w P 2020741133 \\ w P 200741618 \\ w P 202074618 \\ w P 171284862 \\ w P 171284862 \\ w P 128666804 \\ w P 205143646 \\ w P 205143646 \\ w P 205143646 \\ \end{array}$		NH-T RVLKPVD-N NHET RVL-PVD-N NHET RVL-PVD-N N-TITLE-VTQD NH-N
Spec		Leuconstoc inhae Leuconstoc kimchii Leuconstoc lactis Leuconstoc litchii Leuconstoc miyukkimchii Leuconstoc miyukkimchii Leuconstoc rapi Leuconstoc rapi Leuconstoc suionicum Denococcus kitaharae Oenococcus sierae Weissella beninensis Weisella bombi	CUW17996 WP_068850717 WP_148605255 WP_011679295 WP_22074291 WP_22074291 WP_22074291 WP_072613143 WP_007746618 WP_171284862 WP_128686804 WP_205143646		NH-T WVLKPVD-N NHET NVL-PVD-N NHET NVL-PVQD N-N- ITLE-VTQD N-N- IELKAVF N-N IELKAVF N-N IELKAVF N-N IELKAVF N-N IELKAVF N-N IELKAVF N-N IELKAV N-N IELKAV N-N
	cies from other	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis Leuconostoc litchii Leuconostoc miyukkimchii Leuconostoc miyukkimchii Leuconostoc mi Leuconostoc suionicum Dencoccus kitaharae Oenococcus sicerae Weissella bombi Weissella bombi Weissella ceti Weissella charia Weissella charia	$\begin{array}{c} C \overline{w} 17 996 \\ w P 01310 3236 \\ w P 06885 0117 \\ w P 14860 5255 \\ w P 0116 79295 \\ w P 2014 133 \\ w P 2207 42133 \\ w P 207 42291 \\ w P 207 42134 \\ w P 077 4618 \\ w P 077 4618 \\ w P 1128686804 \\ w P 128686804 \\ w P 2051 436 46 \\ w P 2054 436 46 \\ w P 2094 96184 \\ w P 1064 4703 \\ w P 2054 1064 \\ w P 2054 1064 \\ w P 2054 1064 \\ w 1064 4703 \\ w P 2054 1064 \\ w 1064 4703 \\ w P 2054 1064 \\ w P 2054 1064 \\ w 1064 4703 \\ w P 2054 1064 4703 \\ w 206 1064 106$		NH-T WULKPVD-N NHET WULPVD-N N-N-S NHOT
	cies from other <i>conostocaceae</i>	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis Leuconostoc litchii Leuconostoc miyukkimchii Leuconostoc miyukkimchii Leuconostoc raji Leuconostoc suionicum Dencoccus kitaharae Oenococcus sietae Weissella bombi Weissella ceti Weissella cibarta Weissella cibarta Weissella muntiaci	$\begin{array}{c} C \overline{W} I 7996 \\ W P 013103236 \\ W P 068850717 \\ W P 148605255 \\ W P 01167295 \\ W P 220741133 \\ W P 220742291 \\ W P 220742231 \\ W P 20477076 \\ W P 0774618 \\ W P 07746618 \\ W P 07746618 \\ W P 12868044 \\ W P 12868044 \\ W P 2051436462 \\ W P 2051436462 \\ W P 205436604 \\ W P 205436404 \\ W P 205436404 \\ W P 20543642496 \\ W P 205436184 \\ W P 106444703 \\ W P 022791401 \\ W P 248623152 \\ \end{array}$		NH-T WVLKPVD-N NHET WVLFPVD-N NHET WVLFPVD-N KHDT ITLE-VTQD NN IELKAVF NN IELKAVF NN IELKAVF NN IELKAVF NDAT KELK-NQE-F KHDT IELNAV NGKELKV-NQE-F
	cies from other	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis Leuconostoc litchii Leuconostoc miyukkimchii Leuconostoc miyukkimchii Leuconostoc miyukkimchii Leuconostoc suionicum Denococcus kitaharae Oenococcus sietaae Weissella beninensis Weissella beninensis Weissella cibaria Weissella cibaria Weissella cibaria Weissella halotolerans Weissella mutiaci	$\begin{array}{c} C \overline{w} 17 996 \\ w P 01310 3236 \\ w P 068850117 \\ w P 148605255 \\ w P 01167 9295 \\ w P 20207 42133 \\ w P 2207 42291 \\ w P 20207 42291 \\ w P 20207 42691 \\ w P 20207 42691 \\ w P 0726131 43 \\ w P 0077 46618 \\ w P 128666804 \\ w P 128666804 \\ w P 128666804 \\ w P 205143646 \\ w P 092462496 \\ w P 009496184 \\ w P 002462496 \\ w P 009496184 \\ w P 022791401 \\ w P 148623152 \\ w P 20279386 \end{array}$		NH-T RVULRPVD-N NHET NULPVD-N NHET ITLE-VTQD N ITLE-VTQD NH-N IELKAVS NH-N IELKAVS NN IELKAVS DAT IELKAVQE-F KHDT IELKAVQI P-NELKV-NQD PNGKEATT-D-QI-Q AGSKEATT-D-QI-Q ANGKVT-N -DNKR-GQ-N ADGK-G-GQ-N -GGQGV-N
	cies from other <i>conostocaceae</i>	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis Leuconostoc litchii Leuconostoc miyukkimchii Leuconostoc miyukkimchii Leuconostoc raji Leuconostoc suionicum Dencoccus kitaharae Oenococcus sietae Weissella bombi Weissella ceti Weissella cibarta Weissella cibarta Weissella muntiaci	$\begin{array}{c} C \overline{W} I 7996 \\ W P 013103236 \\ W P 068850717 \\ W P 148605255 \\ W P 01167295 \\ W P 220741133 \\ W P 220742291 \\ W P 220742231 \\ W P 20477076 \\ W P 0774618 \\ W P 07746618 \\ W P 07746618 \\ W P 12868044 \\ W P 12868044 \\ W P 2051436462 \\ W P 2051436462 \\ W P 205436604 \\ W P 205436404 \\ W P 205436404 \\ W P 20543642496 \\ W P 205436184 \\ W P 106444703 \\ W P 022791401 \\ W P 248623152 \\ \end{array}$		NH-T WVLKPVD-N NHET WVLFPVD-N NHET WVLFPVD-N KHDT ITLE-VTQD NN IELKAVF NN IELKAVF NN IELKAVF NN IELKAVF NDAT KELK-NQE-F KHDT IELNAV NGKELKV-NQE-F
	cies from other <i>conostocaceae</i>	Leuconostoc inhae Leuconostoc inhae Leuconostoc lactis Leuconostoc litchii Leuconostoc miyukkimchii Leuconostoc miyukkimchii Leuconostoc miyukkimchii Leuconostoc suionicum Denococcus osinicum Oenococcus osini Cenococcus sicerae Weissella bombi Weissella bombi Weissella cibaria Weissella cibaria Weissella cibaria Weissella dalotolerans Weissella muntiaci Weissella ozyzae Weissella soli Weissella thailandensis Weissella uvarum	$\begin{array}{c} C \overline{w} 17 996 \\ w P 01310 3236 \\ w P 0685 0717 \\ w P 0685 0717 \\ w P 14860 5255 \\ w P 0116 7295 \\ w P 2074 1133 \\ w P 22074 2133 \\ w P 22074 2133 \\ w P 2074 133 \\ w P 0774 6618 \\ w P 0726 13143 \\ w P 1074 6618 \\ w P 122866 6004 \\ w P 12286 6604 \\ w P 20514 3646 \\ w P 20514 3646 \\ w P 2054 43646 \\ w P 0994 96184 \\ w P 1064 44703 \\ w P 1064 44703 \\ w P 12672 1461 \\ w P 12672 1461 \\ w P 2076 9386 \\ w P 0030 832 \\ \end{array}$		NH-T WVLRPVD-N NHET WVLPVD-N NHET WVLPVQDN N-N ITLE-VTQD N ITLE-VTQD N ITLELXV N IELXAVF N IELXAVF N NDAT K
	cies from other <i>conostocaceae</i>	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis Leuconostoc lactis Leuconostoc mischii Leuconostoc mischii Leuconostoc miyukkimchii Leuconostoc miyukkimchii Leuconostoc miyukkimchii Leuconostoc suionicum Cenococcus orali Leuconostoc suionicum Cenococcus orali Cenococcus orali Cenococcus sieatae Weissella beninensis Weissella beninensis Weissella cibaria Weissella cibaria Weissella cibaria Weissella halotolerans Weissella muntiaci Weissella oryzae Weissella thailandensis	$\begin{array}{c} CW 17996\\ WP 013103236\\ WP 068850117\\ WP 148605255\\ WP 01167295\\ WP 220741133\\ WP 220742291\\ WP 220742133\\ WP 22074133\\ WP 2074618\\ WP 072613143\\ WP 007746618\\ WP 112868604\\ WP 112868604\\ WP 11286462\\ WP 11284444\\ WP 092462496\\ WP 092462496\\ WP 092462496\\ WP 009496184\\ WP 009496184\\ WP 009496184\\ WP 009496184\\ WP 00949623152\\ WP 027699386\\ WP 070230532\\ WP 0102312\\ \end{array}$		NH-T WULRPVD-N NHET WULPVD-N N-HET WULPVD-N N-N- TILE-VTQD N-N- TILELVAVP N-N-N IELLNAVP N-N-N IELNAVP N-N-N IELNAVP NGK-EXT-D PHGK-EATT-D PHGK PHGK-EATT-D
	cies from other <i>conostocaceae</i>	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis Leuconostoc litchii Leuconostoc miyukkimchii Leuconostoc miyukkimchii Leuconostoc miyukkimchii Leuconostoc suionicum Denococcus osinicum Oenococcus osini Cenococcus sicerae Weissella bombi Weissella bombi Weissella cibaria Weissella cibaria Weissella cibaria Weissella dalotolerans Weissella muntiaci Weissella ozyzae Weissella soli Weissella thailandensis Weissella uvarum	$\begin{array}{c} C \overline{w} 17 996 \\ w P 01310 3236 \\ w P 06885 0117 \\ w P 148605255 \\ w P 011679295 \\ w P 201741133 \\ w P 220742291 \\ w P 202742133 \\ w P 202742133 \\ w P 072613143 \\ w P 072613143 \\ w P 0726618 \\ w P 128686604 \\ w P 128686604 \\ w P 128686604 \\ w P 12864624 \\ w P 205143646 \\ w P 099496184 \\ w P 009496184 \\ w P 009496184 \\ w P 0022791401 \\ w P 148623152 \\ w P 070230532 \\ w P 115470231 \\ w P 2144621 \end{array}$		NH-T RVULRPVD-N NHET NULPVD-N NHET ITLE-VTQD N ITLE-VTQD NH-N IELKAVS NH-N IELKAVS NDAT IELKAVG NAT IELKAVQE-F NAT IELKAVQE-F NAT IELKAVQE-F NAT IELKAVQE-F NAT IELKAVQE-F NH-N IELKAVQI-Q NGKEATT-D-QI-Q ANGKV-VI-N -DNKR-GQ-N ADG
	cies from other <i>conostocaceae</i>	Leuconostoc inhae Leuconostoc kimchii Leuconostoc lactis Leuconostoc litchii Leuconostoc miyukkimchii Leuconostoc miyukkimchii Leuconostoc palmae Leuconostoc rapi Leuconostoc suionicum Dencoccus osinicum Dencoccus osinicum Dencoccus osinicum Dencoccus sicerae Weissella bombi Weissella bombi Weissella cibaria Weissella cibaria Weissella halotolerans Weissella oyrae Weissella oyrae Weissella luvarum Weissella luvarum Weissella uvarum	$\begin{array}{c} C \overline{w} 17 996 \\ w P 01310 3236 \\ w P 068850117 \\ w P 068850117 \\ w P 148605255 \\ w P 011679295 \\ w P 2014133 \\ w P 220742131 \\ w P 20742131 \\ w P 20742131 \\ w P 07746618 \\ w P 07746618 \\ w P 11286620 \\ w P 12866804 \\ w P 12866804 \\ w P 12866804 \\ w P 205143646 \\ w P 09496184 \\ w P 106444703 \\ w P 009496184 \\ w P 106444703 \\ w P 02791401 \\ w P 148623152 \\ w P 0279386 \\ w P 070230532 \\ w P 115470231 \\ w P 2057454621 \\ w P 057745460 \end{array}$		NH-T WULREVD-N NHET WULPVD-N NHET WULPVQDN N-N ITLE-VTQD N ITLELV N ITLELX N ITLELX N IELXAV N IELX-V NDAT ITLE-VTQD P-NKL-EATE-D-QI-Q
	cies from other <i>conostocaceae</i>	Leuconostoc inhae Leuconostoc linchi Leuconostoc licchi Leuconostoc licchi Leuconostoc nivekimchi Leuconostoc miyukkimchi Leuconostoc miyukkimchi Leuconostoc suionicum Denococcus osini Cenococcus osierae Weissella bombi Weissella bombi Weissella ceti Weissella cholerans Weissella cholerans Weissella muntiaci Weissella soli Weissella soli Weissella soli Weissella halandensis Weissella thailandensis Weissella thidescens ^T Agrilactobacillus composti ^T	$\begin{array}{c} CW 17996\\ WP 013103236\\ WP 068850117\\ WP 148605255\\ WP 01167295\\ WP 220741133\\ WP 220742291\\ WP 220742133\\ WP 220742133\\ WP 207746618\\ WP 072613143\\ WP 007746618\\ WP 112866604\\ WP 112866604\\ WP 10644703\\ WP 092462496\\ WP 092462496\\ WP 092462496\\ WP 009496184\\ WP 009492048\\ WP 00949204\\ WP 0094920\\ WP 009492\\ WP 009492\\ WP 009492\\ WP 009492\\ WP 009$		NH-T WULKPVD-N NHET WULKPVD-N N-HET WULFVQDN N-HET ITLE-VTQD N-HELKAVS IELKAVS N-N IELKAVQE-F N-N IELKAVQE-F N-N IELKAVQE-F N-N IELNAVQE-F N-N IELNAVQE-F NGKEATT-D-QI-QI-Q IELNAV NGKEATT-D-QI-QI-Q IELNAV NGG NGKEATT-D-QI-QI-Q
	cies from other <i>conostocaceae</i>	Leuconostoc inhae Leuconostoc lachis Leuconostoc lachis Leuconostoc litchii Leuconostoc miyukkunchii Leuconostoc miyukkunchii Leuconostoc miyukkunchii Leuconostoc miynikkunchi Leuconostoc miynikkunchi Denococcus kitaharae Oenococcus oeni ^T Oenococcus oeni ^T Oenococcus siezae Weissella beninensis Weissella beninensis Weissella cibaria Weissella cibaria Weissella cibaria Weissella cibaria Weissella halotolerans Weissella halotolerans Weissella balatoi Weissella balatoi Weissella balatoi Weissella balatoise Weissella thailandensis Weissella uvarum Weissella viridescens ^T Agrilactobacillus composti ^T Catellicoccus marimammalium ^T Dellaglioa algida ^T	$\begin{array}{c} C \overline{w} 17 996 \\ w P 01310 3236 \\ w P 068850117 \\ w P 148605255 \\ w 01167 9295 \\ w P 20207 42133 \\ w P 2207 42291 \\ w P 20207 42291 \\ w P 20207 42691 \\ w P 20207 4618 \\ w P 20207 4618 \\ w P 1028 4662 \\ w P 1028 4646 \\ w P 205143646 \\ w P 205143646 \\ w P 205143646 \\ w P 202543646 \\ w P 202543646 \\ w P 202543646 \\ w P 20254352 \\ w P 205143646 \\ w P 202791401 \\ w P 205143646 \\ w P 20231352 \\ w P 205144621 \\ w P 205144621 \\ w P 205144621 \\ w P 205144621 \\ w P 20544623 \\ w P 205452397 \\ w P 203592048 \\ w P 0057973403 \end{array}$		NH-T WVLKPVD-N NHET WVLFPVD-N N-HET WVLFPVD-N N-N TILE-VTQD N-H IELKAVF N-N IELKAVF N-N IELKAVF N-N IELKAVF N-N IELKAV N-N IELKAV NGK-EAT-D-I-Q-P
Leu	cies from other conostocaceae genera	Leuconostoc inhae Leuconostoc lachis Leuconostoc lachis Leuconostoc litchii Leuconostoc miyukkimchii Leuconostoc miyukkimchii Leuconostoc manae Leuconostoc rapi Leuconostoc rapi Leuconostoc suionicum Denococcus osini Conococcus osini Oenococcus sicerae Weissella bombi Weissella bombi Weissella cibaria Weissella cibaria Weissella halotolerans Weissella halotolerans Weissella halotolerans Weissella longrae Weissella coli Weissella coli Weissella coli Weissella uvarum Weissella uvarum Weissella uvidescens ^T Agrilactobacillus composti ^T Catellicoccus marimammalium ^T Dellaglioa algida ^T Furfurilactobacillus rossiae ^T	$\begin{array}{c} C w_{II} 1996 \\ w_{P} 0 0 131 0 323 6 \\ w_{P} 0 6 0 8 50117 \\ w_{P} 148 6 0 5255 \\ w_{P} 0 0 116 79295 \\ w_{P} 2027 41133 \\ w_{P} 2207 42291 \\ w_{P} 2027 42291 \\ w_{P} 0 7261 3143 \\ w_{P} 0 0774 6618 \\ w_{P} 0 7246612 \\ w_{P} 128 66804 \\ w_{P} 128 66804 \\ w_{P} 2051 43646 \\ w_{P} 0 92462496 \\ w_{P} 0 0949 6184 \\ w_{P} 106444703 \\ w_{P} 0 02495316 \\ w_{P} 0 70230532 \\ w_{P} 0 1769336 \\ w_{P} 0 70230532 \\ w_{P} 115470231 \\ w_{P} 205144621 \\ w_{P} 0 57745460 \\ w_{P} 0 0 9492048 \\ w_{P} 0 0 7973403 \\ w_{P} 0 17261733 \end{array}$		NH-T RVULRPVD-N NHET RVULRPVD-N NHET RVILRPVTD-N NH-T
Leu	cies from other conostocaceae genera	Leuconostoc inhae Leuconostoc linchi Leuconostoc licchi Leuconostoc licchi Leuconostoc niyukkimchi Leuconostoc miyukkimchi Leuconostoc palmae Leuconostoc suionicum Denococcus osini Cenococcus osini Cenococcus sicerae Weissella bombi Weissella bombi Weissella ceti Weissella charia Weissella charia Weissella charia Weissella charia Weissella colerans Weissella colerans Weissella colerans Weissella colerans Weissella colerans Weissella uvarum Weissella uvarum Weissella uvarum Weissella viridescens ^T Agrilactobacillus composti ^T Catellicoccus marimammalium ^T Dellaglioa algida ^T Furfurilactobacillus rossiae ^T Holzapfelia floricola ^T	$\begin{array}{c} C w_{II} 7996 \\ w_{P} 013103236 \\ w_{P} 06850117 \\ w_{P} 148605255 \\ w_{P} 011679295 \\ w_{P} 202741133 \\ w_{P} 220742291 \\ w_{P} 202741343 \\ w_{P} 202741343 \\ w_{P} 007746618 \\ w_{P} 0124662 \\ w_{P} 0124662 \\ w_{P} 0224296 \\ w_{P} 0294262496 \\ w_{P} 009496184 \\ w_{P} 0294262496 \\ w_{P} 009496184 \\ w_{P} 0294262496 \\ w_{P} 02791401 \\ w_{P} 02791401 \\ w_{P} 02791401 \\ w_{P} 0279336 \\ w_{P} 070230532 \\ w_{P} 01544621 \\ w_{P} 035452397 \\ w_{P} 005977403 \\ w_{P} 005973403 \\ w_{P} 017261733 \\ w_{P} 056975191 \end{array}$		NH-T WULKPVD-N NHET WULKPVD-N NHET WULKPVQD-N WULKPVQD-N N-N K-WITLTAM NH-N NDAT K-ELKX-N
Leu	cies from other conostocaceae genera	Leuconostoc inhae Leuconostoc lachis Leuconostoc lachis Leuconostoc lachis Leuconostoc mischii Leuconostoc mischi Leuconostoc mischi Leuconostoc minicum Denococcus suinicum Oenococcus sienae Oenococcus sienae Weissella beninensis Weissella beninensis Weissella cibaria Weissella cibaria Weissella cibaria Weissella hahlotolerans Weissella thailandensis Weissella thailandensis Weissella viridescens ^T Agrilactobacillus composti ^T Catellicoccus marimammalium ^T Dellaglioa algida ^T Furfurilactobacillus rossiae ^T Holzapfella floricola ^T Listeria moncytogenes ^T	$\begin{array}{c} Cu \overline{u}, 1996\\ WP & 013103236\\ WP & 06850117\\ WP & 06850117\\ WP & 1167295\\ WP & 201742291\\ WP & 201742291\\ WP & 201742291\\ WP & 2017423143\\ WP & 20174618\\ WP & 201746618\\ WP & 112866604\\ WP & 10644703\\ WP & 105745460\\ WP & 105745460\\ WP & 057973403\\ WP & 057973403\\ WP & 057973403\\ WP & 017261733\\ WP & 0575191\\ NP & 464747.1\\ \end{array}$		NH-T WULRPVD-N NHET WULRPVD-N N-HET WULRPVD-N N-HET ITLE-VTQD N-H IELKAVF N-N IELKAVF N-N IELKAVF N-N IELKAVQE-F N-N IELKAVQE-F NGK-ELKY-NQE-F NGK-EATT-D-QI-QI
Leu	cies from other conostocaceae genera	Leuconostoc inhae Leuconostoc lachis Leuconostoc lachis Leuconostoc lachis Leuconostoc mischii Leuconostoc mischi Leuconostoc mischi Leuconostoc rapi Leuconostoc rapi Leuconostoc suionicum Denococcus sitaharae Oenococcus sierae Weissella beninensis Weissella beninensis Weissella beninensis Weissella cibaria Weissella cibaria Weissella cibaria Weissella cibaria Weissella cibaria Weissella oryzae Weissella thailandensis Weissella viridescens ^T Agrilactobacillus composti ^T Catellicoccus marimammalium ^T Dellaglioa algida ^T Furfurilactobacillus rossiae ^T Holzapfelia floricola ^T Listeria monocytogenes ^T Planococcus citreus ^T	$\begin{array}{c} CWB_{17996}\\ cWP_{013103236}\\ WP_{068850117}\\ WP_{148605255}\\ WP_{01167295}\\ WP_{020741133}\\ WP_{020742291}\\ WP_{02074613143}\\ WP_{02074618}\\ WP_{0126462}\\ WP_{0126462}\\ WP_{01264462}\\ WP_{025143646}\\ WP_{029462496}\\ WP_{029386}\\ WP_{029462496}\\ WP_{029386}\\ WP_{029462496}\\ WP_{029386}\\ WP_{029386}\\ WP_{029386}\\ WP_{029386}\\ WP_{029386}\\ WP_{029386}\\ WP_{023152}\\ WP_{01623152}\\ WP_{017261733}\\ WP_{005973403}\\ WP_{059773403}\\ WP_{059791401}\\ WP_{056975191}\\ WP_{056975191}\\ WP_{0123100221}\\ \end{array}$		NH-T WVLKPVD-N NHET WVLFPVD-N N-HET
Leu	cies from other conostocaceae genera	Leuconostoc inhae Leuconostoc lachis Leuconostoc lachis Leuconostoc litchii Leuconostoc miyukkmchii Leuconostoc miyukkmchii Leuconostoc miyukkmchii Leuconostoc suionicum Denococcus kitaharae Oenococcus kitaharae Oenococcus sicerae Weissella bombi Weissella bombi Weissella charia Weissella charia Weissella charia Weissella charia Weissella charia Weissella charia Weissella coryae Weissella oryae Weissella uvarum Weissella uvarum Weissella viridescens ^T Agrilactobacillus composti ^T Catellicoccus marimamalium ^T Dellagiloa algida ^T Furfurilactobacillus rossiae ^T Histeria monocytogenes ^T Planococcus citreus ^T Scopulibacillus darangshiensis ^T	$\begin{array}{c} Cu \overline{u} 17996 \\ wP_0(3103236 \\ wP_0(6850117 \\ wP_148605255 \\ wP_0(1679295 \\ wP_220742133 \\ wP_220742133 \\ wP_220742133 \\ wP_2074618 \\ wP_072613143 \\ wP_007746618 \\ wP_171284862 \\ wP_12866604 \\ wP_12866604 \\ wP_02462496 \\ wP_02462496 \\ wP_02462496 \\ wP_09494624 \\ wP_0948644703 \\ wP_022791401 \\ wP_022791401 \\ wP_022791401 \\ wP_022791401 \\ wP_022791401 \\ wP_0235452397 \\ wP_035452397 \\ wP_035452397 \\ wP_035452397 \\ wP_056975191 \\ wP_046747.1 \\ wP_1320221 \\ wP_132742639 \end{array}$		NH-T WULRPVD-N NHET WULPVD-N N-HET N-HET
Leu	cies from other conostocaceae genera	Leuconostoc inhae Leuconostoc lachis Leuconostoc lachis Leuconostoc lachis Leuconostoc mischii Leuconostoc mischi Leuconostoc mischi Leuconostoc rapi Leuconostoc rapi Leuconostoc rapi Leuconostoc suionicum Denococcus sitaharae Oenococcus sierae Weissella benhinensis Weissella benhinensis Weissella benhi Weissella cibaria Weissella cibaria Weissella cibaria Weissella cibaria Weissella coryzae Weissella thailandensis Weissella viridescens ^T Agrilactobacillus composti ^T Catellicoccus marimammalium ^T Dellaglioa algida ^T Furfurilactobacillus rossiae ^T Holzapfelia floricola ^T Listeria monocytogenes ^T Planococcus citreus ^T	$\begin{array}{c} CWB_{17996}\\ cWP_{013103236}\\ WP_{068850117}\\ WP_{148605255}\\ WP_{01167295}\\ WP_{020741133}\\ WP_{020742291}\\ WP_{02074613143}\\ WP_{02074618}\\ WP_{0126462}\\ WP_{0126462}\\ WP_{01264462}\\ WP_{025143646}\\ WP_{029462496}\\ WP_{029386}\\ WP_{029462496}\\ WP_{029386}\\ WP_{029462496}\\ WP_{029386}\\ WP_{029386}\\ WP_{029386}\\ WP_{029386}\\ WP_{029386}\\ WP_{029386}\\ WP_{023152}\\ WP_{01623152}\\ WP_{017261733}\\ WP_{005973403}\\ WP_{059773403}\\ WP_{059791401}\\ WP_{056975191}\\ WP_{056975191}\\ WP_{0123100221}\\ \end{array}$		NH-T WVLKPVD-N NHET WVLFPVD-N N-HET

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.

(a)			366		40
	Oenococcus oeni ^T	WP 011677688		ENGG A	40 NNLRSARCLVERHEDVVQDVI
Oenococcus		WP 128685115		(-	R
(414)	Oenococcus kitaharae	WP_007747102			RR
(4/4)	Oenococcus alcoholitolerans Convivina intestini ^I	KG031908	RMKE-TKRKL		G-VKKRKID-ADED-M
	Fructobacillus fructosus ^T	WP_089938971 WP 187753584	RMKE-TAR-L		G-VKKRKIDKADED-M
	Leuconostoc carnosum	CAA65078	RMKE-TTRSL		-G-VKKRKIDKADED-M
	Leuconostoc fallax	WP_010008856	RMRE-TKRNL		-G-VKKRKID-QDND-M
	Leuconostoc mesenteroides ^T	MBD9365386	RMKE-TAR-L		-G-VKKRKIDKADED-M
	Leuconostoc miyukkimchii Leuconostoc palmae	WP_220741015 WP 220741916	RMKE-TTRKL RMKE-TTRKL		-G-VKKRKIDKADED-M -G-VKKRKIDKADED-M
Species from other	Leuconostoc rapi	WP 204769341	RMKE-TTRKL		-G-VKKRKIDKADED-M
Leuconostocaceae	Weissella oryzae	WP_045476354	RMKE-VARDL	-	-G-I-A-KRKID-RD-D-M
genera	Weissella soli	WP_147153328	LMREKRDL		-A-I-A-KRKIDDRDED-M
gener a	Weissella thailandensis Weissella uvarum	WP_115471058 WP_205144937	RMKE-VNREL RMHE-V-R-L		-G-I-A-KRKID-RDED-M -G-IKA-KRKID-RD-D-M
	Weissella viridescens ^T	WP 072152375	RMHE-V-RNL		-G-I-A-KRKID-RD-D-M
	Weissella beninensis	WP_205144294	MKE-VARQL		-S-IKN-KRKID-RD-D-M
	Weissella cryptocerci	WP_133363122	MKE-VSREL		-S-IKN-KRKID-RD-D-M
	Weissella fabalis	WP_168722271	MKE-VSREL		-S-IKN-KRKID-RD-D-M
	Apilactobacillus kunkeei ^T Enterobacter cloacae ^T	WP_054606803 TAT60443	MKE-VKR-S		S-IKKRQIKDEE-F TTIKA-KKME-AW-I-
	Pseudomonas aeruginosa ^T	WP_162945939	GK-ELR-L		TTIKA-KKME-AW-I-
Other Bacteria	Lactiplantibacillus plantarum ^T	TAR50926	MKE-VKREL		S-IKN-KRKI-HAD-D-WG
(0/500)	Tetrasphaera japonica ^T	AAY57910	VMKVD-DH		Q-IKKRMQRSW
	Wigglesworthia glossinidia ^T	WP 014354013	GK-ELN-F		TTIKA-KKME-SW-I-
		_			
(b)			281		31
	Weissella jogaejeotgali	WP_075270024	RIFLMTPIH		G LIGNTEKP WNEWQVNLFF
	Weissella bombi Weissella ceti	WP_092461930 WP_038528101			F
	Weissella cibaria	WP 043708208			D
	Weissella coleopterorum	WP_166010241	-VA		- WRMDPKYKTL-
Weissella	Weissella confusa	WP_003608218			D
main clade	Weissella halotolerans Weissella hellenica	WP_022791740 WP_042493316) –v		V-HSR-DG
	Weissella minor	WP 057788441			VDH DAL-
	Weissella paramesenteroides	WP_002828548	3		F
	Weissella soli	WP_114981094			- I-M-PDDA
	Weissella thailandensis Weissella uvarum	WP_115471179 WP 205145741			G-VDHSL-
	Weissella viridescens ^T	WP_057746093			DAL-
	Weissella beninensis	WP_205143917	7K-S-V-		
Weissella clade-2	Weissella cryptocerci	WP_133364512			
(0/3)	Weissella fabalis	WP_168722037			
(0.0)	Convivina intestini ^T Fructobacillus fructosus ^T	WP_089938766 WP_01069155			
Species from other	Oenococcus oeni ^T	WP 186413956			
	Leuconostoc carnosum	WP 014974591			
Leuconostocaceae	Leuconostoc mesenteroides ^T	WP_002815199	9P-A	F	IDAL-
genera	Leuconostoc suionicum	WP_072613969			
	Acetilactobacillus jinshanensis ^T	WP_133441441			
	Alicyclobacillus acidocaldarius ^T	WP_012810642 WP 034531845			
Other Bacteria	Apilactobacillus kunkeei ^T				
	Dellaglioa algida ^T	WP_057973871	K-S	-HMS	KIDIV-
(0/500)		WP_057973871	LK-S	-HMS	KIDIV- KIDIV-
(0/500)	Dellaglioa algida ^T Fructilactobacillus fructivorans ^T	WP_057973871 WP_056998014	LK-S	-HMS	KIDIV- KIDIV-
	Dellaglioa algida ^T Fructilactobacillus fructivorans ^T	WP_057973871 WP_056998014	LK-S	-HMS	:KIDIV- :SKIDIV- :SRID-T-
(0/500) (C)	Dellaglioa algida [†] Fructilactobacillus fructivorans [†] Lacticaseibacillus casei [†]	WP_057973871 WP_056998014	86	- HMS MI - HMK	KIDIV- KIDIV-
(0/500)	Dellaglioa algida [†] Fructilactobacillus fructivorans [†] Lacticaseibacillus casei [†] 2 Weissella fabalis Weissella beninensis	WP_057973871 WP_056998014 WP_039638954 WP_16872222 WP_20514382	LK-S IS IS 96 01 WGGVLTRSQQRV 24 -P-KKP-MQ-	-HMS MI -HMF VNEI -T-I	KIDIV-
(0/500) (C)	Dellaglioa algida [†] Fructilactobacillus fructivorans [†] Lacticaseibacillus casei [†] 2 weissella fabalis Weissella chonimensis Weissella cryptocerci	WP_057973871 WP_056998014 WP_039638954 WP_1687222 WP_20514382 WP_13336299	LK-S AK AS B6 WGGVLTRSQQRV 24 -P-KKP-MQ- 22 -E-QEQ 22 -E-QEQ	-HMS MI -HMF VNEI -T-I -T-V	KIDIV-
(0/500) (C) <i>Weissella</i> Clade-	<pre>Pellaglioa algida^T Fructilactobacillus fructivorans^T Lacticaseibacillus casei^T 2 {</pre>	WP_057973871 WP_05699801 WP_039638954 WP_039638954 WP_16872220 WP_20514380 WP_13336295 WP_09246281	LK-S iS 86 96 91 WGGVLTRSQQRV 24 -P-KKP-MQ- 92 -E-QEQ 5 -E-KKLTL	-HMS MI -HMF -T-I -T-I :SDQQ I	KIDIV-
(0/500) (C) <i>Weissella</i> Clade-	Dellaglioa algida [†] Fructilactobacillus fructivorans [†] Lacticaseibacillus casei [†] 2 weissella fabalis Weissella chonimensis Weissella cryptocerci	WP_05797387 WP_056998014 WP_039638954 WP_05154385 WP_0514385 WP_02514385 WP_024283 WP_04371165 WP_04371165 WP_059731	K-S K-S S 86 WGGVLTRSQQRV S 2	-HMS MI -HMF VNEI -T-I -T-V SDQQ I AD-Q V AD-Q I	KIDIV- SKIDIV- SRID-T- 12 KKHTHNHLVWAVTSAGKT LSFNQK
(0/500) (C) <i>Weissella</i> Clade-	<pre>Dellaglioa algida^T Fructilactobacillus fructivorans^T Lacticaseibacillus casei^T 2 { Weissella fabalis Weissella boninensis Weissella cipatia Weissella cibatia Weissella confusa Weissella confusa Weissella la diestrammenae</pre>	WP_057973871 WP_056998014 WP_039638954 WP_13687222(WP_2051438 WP_1333629 WP_0246281 WP_04371162 WP_0659731.	K-S K-S WGGVLTRSQQRV S WGGVLTRSQQRV S S	-HMS MI -HMF -T-I -T-V SDQQ I AD-Q V AD-Q I SQAQ I	KIDIV-
(0/500) (C) <i>Weissella</i> Clade-	<pre>Dellaglioa algida[†] Fructilactobacillus fructivorans[†] Lacticaseibacillus casei[†] 2 { Weissella fabalis Weissella combinensis Weissella combinensis Weissella cibaria Weissella cibaria Weissella diestrammenae Weissella halotolerans </pre>	WP_057973871 WP_05999014 WP_039638954 WP_039638954 WP_02514385 WP_0224528 WP_0224528 WP_04271165 WP_05697314 WP_18752856 WP_18752856	86 WGGVLTRSQQRV	-HMS MI -HMF -T-I -T-V SDQQ I SDQQ I SQAQ I AQQL F	KIDIV- SKIDIV- SRID-T- 12 KKHTHNHLVWAVTGAGKT LSHNQK MG-QG
(0/500) (C) <i>Weissella</i> Clade- (3/3)	<pre>Dellaglioa algida^T Fructilactobacillus fructivorans^T Lacticaseibacillus casei^T 2 {</pre>	WP_05797387 WP_056998014 WP_039638954 WP_039638954 WP_020514362 WP_00246281 WP_04371166 WP_04371165 WP_04371165 WP_18752863 WP_18752863 WP_16218800 WP_162248965	K-S K-S S 86 WGGVLTRSQQRV -E-QEQ -E-QEQ -E-QEQ -E-QEQ -F-KKPR -T-T-PA -T-T-EAR -E-CPR-NK -E-CPR-NK	-HMS -HMK -HMK VNEI -T-I -T-V SDQQ I AD-Q V AD-Q V AD-Q I SQAQ I SQAQ I SQAQ I S-QQ I	KIDIV-
(0/500) (C) <i>Weissella</i> Clade- (3/3) <i>Weissella</i>	<pre>Dellaglioa algida^T Fructilactobacillus fructivorans^T Lacticaseibacillus casei^T 2 { Weissella fabalis Weissella beninensis Weissella charia Weissella charia Weissella cibaria Weissella diestrammenae Weissella halotolerans Weissella halotoleran Weissella jogaejeotgali Weissella rit </pre>	WP_057973871 WP_05999014 WP_039638954 WP_039638954 WP_02514385 WP_0224528 WP_0224528 WP_04271165 WP_05697314 WP_18752856 WP_18752856	86 WGGVLTRSQQRV K-S WGGVLTRSQQRV 2	-HMS -HMF -HMF -T-I -T-V SDQQ I AD-Q V AD-Q I SQAQ I SQQL F SQQL F SD-Q I	KIDIV- SKIDIV- SRID-T- 12 KKHTHNHLVWAVTGAGKT LSHNQK MG-QG
(0/500) (C) <i>Weissella</i> Clade- (3/3)	<pre>Dellaglioa algida[†] Fructilactobacillus fructivorans[†] Lacticaseibacillus casei[†] 2</pre>	WP_05797387; WP_056998014 WP_039638954 WP_039638954 WP_025133629 WP_02246228 WP_02246228 WP_0371165 WP_05697314 WP_16213800 WP_1621380634 WP_0373557 WP_0336034	86 91 92 92 93 94 95 96 97 98 99 90 91 92 94 97 98 98 99 90 90 91 92 94 95 94 94 95 95 96 97 97 96 97 97 97 97 97 97 97 97 98 98 99 90 90 91 92 93 94 94 94	-HMS MI -HMK VNEI -T-I -T-V SDQQ I AD-Q V AD-Q V AD-Q I SQAQ I SQAQ I SQ-Q I SD-Q I SD-Q I AQQQ I AQ I	
(0/500) (C) Weissella Clade- (3/3) Weissella	 Dellaglioa algida[†] Fructilactobacillus fructivorans[†] Lacticaseibacillus casei[†] %eissella fabalis weissella compositionensis weissella compositionensis weissella confusa weissella diestrammenae weissella hallenica weissella hallenica weissella kandleri weissella koreensis weissella koreensis 	WP_057973871 WP_056998014 WP_039638954 WP_025498954 WP_039638954 WP_025133629 WP_0246281 WP_0371167 WP_0597314 WP_18752866 WP_16213800 WP_0360344 WP_05775573 WP_0380346 WP_05775573	$ \begin{array}{c} &K-S\\ 1 &K-S\\ 1 &S\\ 1 &S\\ 1 &S\\ 1 &S\\ 2 &S\\ 2 &SS\\ 2 &SS\\ 2 &\\ 3 &T$	-HMS MI -HMF -T-I -T-V SDQQ I AD-Q V AD-Q V AD-Q V SDAQ I SD-Q I SD-Q I SD-Q I SD-Q I -AAK- I	KIDIV- SKIDIV- SKIDIV- SRID-T- 122 KKHTHNHLVWAVTGAGKT LSHNQK
(0/500) (C) <i>Weissella</i> Clade- (3/3) <i>Weissella</i>	<pre>Dellaglioa algida[†] Fructilactobacillus fructivorans[†] Lacticaseibacillus casei[†] 2</pre>	WP_05797387; WP_056998014 WP_039639954 WP_2051438; WP_1338229; WP_0246284 WP_0246284 WP_0246286 WP_16213806 WP_16213806 WP_16213806 WP_03860344 WP_03860346 WP_1570575 WP_15705973 WP_157059757 WP_157059757 WP_157059757 WP_15705975775775775775775775775775775775775775	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	-HMS MI -HMK VNEI -T-I -T-V SDQQ I AD-Q I SQAQ I SD-Q I SQQQ I AQQQ I SD-Q I AQQQ I SD-Q I AQQQ I	KIDIV- SKIDIV- SKIDIV- SRID-T- 12 KKHTHNHLVWAVTGAKK LSHNQK TA LSTKTS-TY QA L-AGRS-TY QA LTNHSS-TY QA LTNHSS-TY LG VQTGKPY1H
(0/500) (C) Weissella Clade- (3/3) Weissella	<pre>Dellaglioa algida[†] Fructilactobacillus fructivorans[†] Lacticaseibacillus casei[†] 2 { Weissella fabalis Weissella bombi Weissella chyptocerci Weissella charia Weissella confusa Weissella confusa Weissella chololerans Weissella halotolerans Weissella halotolerans Weissella kandleri Weissella kandleri Weissella kandleri Weissella konensis Weissella konensis Weissella konensis Weissella kandleri Weissella konensis Weissella konensis Weissella minor Weissell</pre>	WP_057973871 WP_056998014 WP_039638954 WP_025498014 WP_039638954 WP_025133629 WP_0246281 WP_0246281 WP_0371167 WP_0597314 WP_18752866 WP_18752866 WP_0575573 WP_0380344 WP_05775573 WP_03398966 WP_15705033	$ \begin{array}{c}K-S\\K-S\\K-K\\S\\S\\S\\S\\S\\\\\\$	-HMS MI -HMK -T-I -T-V SDQQ I AD-Q V AD-Q V AD-Q I SDAQ I SD-Q I	KIDIV- SKIDIV- SKIDIV- SKIDIV-
(0/500) (C) Weissella Clade- (3/3) Weissella	 Dellaglioa algida[†] Fructilactobacillus fructivorans[†] Lacticaseibacillus casei[†] 2 {	WP_05797387 WP_056998014 WP_039638954 WP_039638954 WP_025133629 WP_0246284 WP_0246284 WP_0371165 WP_0246284 WP_04249365 WP_04249365 WP_04249365 WP_05775575 WP_0398966 WP_15705033 WP_1662366 WP_05234855 WP_0228285 WP_0228285 WP_0228565 WP_022855 WP_028555 WP_02855 WP_028555WP_028555 WP_028555	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	-HMS MI -HMF -TI -T-V SDQQ I AD-Q V AD-Q I SQQQ I AQQQ I AQQQ I AQQQ I AQQQ I AQQQ I AQQQ I SD-Q I SD-Q I SQQL I SQQL I SQQL I SQQL I SQQL I SQQL I	KIDIV- SKIDIV- SKIDIV- SRID-T- 12 KKHTHNHLVWAVTGAGKI LSHNQK
(0/500) (C) Weissella Clade- (3/3) Weissella	 Pellaglioa algida[†] Fructilactobacillus fructivorans[†] Lacticaseibacillus casei[†] %eissella fabalis weissella comptocerci Weissella comptocerci Weissella confusa Weissella confusa Weissella hellenica Weissella hellenica Weissella koreensis Weissella koreensis Weissella muntiaci Weissella muntiaci Weissella paramesenteroides Weissella soli Weissella soli 	WP_05797387 WP_056998014 WP_039638954 WP_039638954 WP_025133629 WP_02246281 WP_04246361 WP_04246365 WP_0577577 WP_05775577 WP_05775577 WP_05775577 WP_05775575 WP_05775575 WP_05775575 WP_05775575 WP_05775575 WP_0724855 WP_07022856 WP_07022856 WP_07022956 WP_0722565 WP_0722556 WP_0722556 WP_0722556	$ \begin{array}{c}\kappa - s - s - s - s - s - s - s - s - $	-HMS MI -HMF VNEI -T-I SDQQ I AD-Q V AD-Q V AD-Q V SSQQ I AQQL F SSQQ I AQQQ F SSD-Q I A AQQQ F SSD-Q I A AQQQ F SSD-Q I A AQQ F SSD-Q I A AQQ F SSD-Q I SSD-Q I A A A A A A A A A A A A A A A A A A A	KIDIV- SKIDIV- SKIDIV- SRID-T- 12 KKHTHNHLVWAVTGAGKI LSHNQK
(0/500) (C) <i>Weissella</i> Clade- (3/3) <i>Weissella</i> Main Clad	<pre>Dellaglioa algida[†] Fructilactobacillus fructivorans[†] Lacticaseibacillus casei[†] 2</pre>	WP_0599303 WP_036998014 WP_036998014 WP_0363984 WP_0363984 WP_0363984 WP_0246228 WP_0246228 WP_0246228 WP_024626 WP_16213806 WP_16213806 WP_038034 WP_038034 WP_038034 WP_038036 WP_1570573 WP_146236 WP_022889 WP_0702295 WP_0702295 WP_0399403	K-S K-S K-S K-S WGGVLTRSQQRV 24 -P-KKP-MQ- 2	-HMS MI -HMK VNEI -T-V SDQQ I SDQQ I SDQQ I SD-Q I SQQQ I SD-Q I SCQQ I SD-Q I SCQQ I SD-Q I SCQQ I SCQ SCQ I SCQ SCQ I SCQ	
(0/500) (C) Weissella Clade- (3/3) Weissella Main Clad	<pre>Dellaglioa algida[†] Fructilactobacillus fructivorans[†] Lacticaseibacillus casei[†] 2</pre>	WP_05797387 WP_056998014 WP_039638954 WP_039638954 WP_025133629 WP_0246228 WP_0246228 WP_0246228 WP_0246284 WP_037116 WP_05697311 WP_16213800 WP_04249365 WP_02453634 WP_05234855 WP_05234855 WP_05234855 WP_0228289 WP_14132500WP_14132500 WP_14132500WP_14132500 WP_14132500WP_14132500 WP_14132500WP_14132500 WP_14132500WP_14132500 WP_14132500WP_14132500 WP_14132500WP_14132500 WP_14132500WP_14132500 WP_14132500WP_141325000 WP_141325000 WP_141325000000000000000000000000000000000000	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	-HMS MI -HMF -T-I -T-V SDQQ I AD-Q V SDQQ I SQAQ I S SQAQ I S S SQAQ I S S S S S S S S S S S S S S S S S S S	
(0/500) (C) Weissella Clade- (3/3) Weissella Main Clad	 Dellaglioa algida[†] Fructilactobacillus fructivorans[†] Lacticaseibacillus casei[†] 2 {	WP_05797387 WP_056998014 WP_039638954 WP_039638954 WP_025133629 WP_0246281 WP_05471167 WP_0597314 WP_18752865 WP_04249365 WP_04249365 WP_0360344 WP_0575573 WP_01398966 WP_14862366 WP_02284855 WP_02284855 WP_02284855 WP_02284855 WP_02284855 WP_02284855 WP_02284855 WP_02284855 WP_03984033 WP_14132501 WP_03994033 WP_18775394 OXK56622.1	$ \begin{array}{c}K-S\\K-S\\K-K\\S\\S\\S\\S\\S\\S\\S\\S\\S\\$	-HMS MX MX 	
(0/500) (C) Weissella Clade- (3/3) Weissella Main Clad	<pre>Dellaglioa algida[†] Fructilactobacillus fructivorans[†] Lacticaseibacillus casei[†] 2</pre>	WP_05797387; WP_056998014 WP_039639954 WP_039639954 WP_025133629 WP_0224528 WP_0224528 WP_024528 WP_024528 WP_0371165 WP_0373165 WP_0373165 WP_036034 WP_1621380 WP_1621380 WP_1570503 WP_137059403 WP_1413250 WP_07022956 WP_07022956 WP_1413250 WP_1877539 WP_1877539 0.1556622.1	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	-HMS MI -HMF VNEI -T-I -T-V SDQQ V AD-Q I AQQL F AQQL F AQQL F AQQL I AQQL F AQQL I AQQL I A AQQ I A AQQ I A AQQ I A AQQ I A AQ A A A A A A A A A A A A A A A A A	KIDIV- SKIDIV- SKIDIV- SRID-T- 12 KKHTHNHLVWAVTGAGKT LSHNCK
(0/500) (C) Weissella Clade- (3/3) Weissella Main Clad Species from othe Leuconostocacea	<pre>Dellaglioa algida[†] Fructilactobacillus fructivorans[†] Lacticaseibacillus casei[†] 2</pre>	WP_05797387 WP_056998014 WP_039638954 WP_039638954 WP_025133629 WP_0246281 WP_05471167 WP_0597314 WP_18752865 WP_04249365 WP_04249365 WP_0360344 WP_0575573 WP_01398966 WP_14862366 WP_02284855 WP_02284855 WP_02284855 WP_02284855 WP_02284855 WP_02284855 WP_02284855 WP_02284855 WP_03984033 WP_14132501 WP_03994033 WP_18775394 OXK56622.1	$ \begin{array}{c}K-S\\K-S\\K-K-S\\K-K-K-K-K-K-K-K-K-K-K-K-K-K-K-K-K$	-HMS MI -HMR 	
(0/500) (C) Weissella Clade- (3/3) Weissella Main Clad Species from othe Leuconostocaceau genera	<pre>Dellaglioa algida[†] Fructilactobacillus fructivorans[†] Lacticaseibacillus casei[†] 2</pre>	WP_05797387 WP_056998014 WP_039638954 WP_039638954 WP_025133629 WP_0246284 WP_0246284 WP_0246284 WP_0246284 WP_0246286 WP_16213804 WP_0246286 WP_05360344 WP_05360344 WP_05234855 WP_0722955 WP_0722955 WP_14132505 WP_0722956 WP_14132505 WP_0722956 WP_14132505 WP_0722956 WP_0726229 WP_0726295 WP_07262837 WP_0726285 WP_07262837 WP_07262857 WP_07267WP_07267 WP_07267 WP_07267WP_07267 WP_07267WP_07267 WP_07267WP_07267 WP_07267W	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	-HMS MI -HMR 	KIDIV- SKIDIV- SKIDIV- SRID-T- 12 KKHTHNHLVWAVTGAGKT LSHNQK
(0/500) (C) Weissella Clade- (3/3) Weissella Main Clad Species from othe Leuconostocacea	<pre>Dellaglioa algida[†] Fructilactobacillus fructivorans[†] Lacticaseibacillus casei[†] 2</pre>	WP_05797387 WP_056998014 WP_039638954 WP_039638954 WP_025133629 WP_0246284 WP_0246284 WP_0246284 WP_0246284 WP_0246286 WP_16213804 WP_0246286 WP_05360344 WP_05360344 WP_05234855 WP_0722955 WP_0722955 WP_14132505 WP_0722956 WP_14132505 WP_0722956 WP_14132505 WP_0722956 WP_0726229 WP_0726295 WP_07262837 WP_0726285 WP_07262837 WP_07262857 WP_07267WP_07267 WP_07267 WP_07267WP_07267 WP_07267WP_07267 WP_07267WP_07267 WP_07267W	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	-HMS MI -HMF VNEI -T SDQQ I ADQ-Q V SQAQ I SQAQ I S SQAQ I S S SQAQ I S S SQAQ I S S S S S S S S S S S S S S S S S S S	

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.

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. \$32	
TatD family hydrolase	WP_180369227	2 aa Ins	48-70	Fig. \$33	
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. \$36	
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
Cytochrome d ubiquinol oxidase subunit II	WP_070230395	1 aa Ins	165–205	Fig. S40	clade
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 <i>Bacillales</i> species.					

Table 2. Conserved signature indels specific for the genus	<i>Oenococcus</i> and for the two <i>Weissella</i> species clades
--	---

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

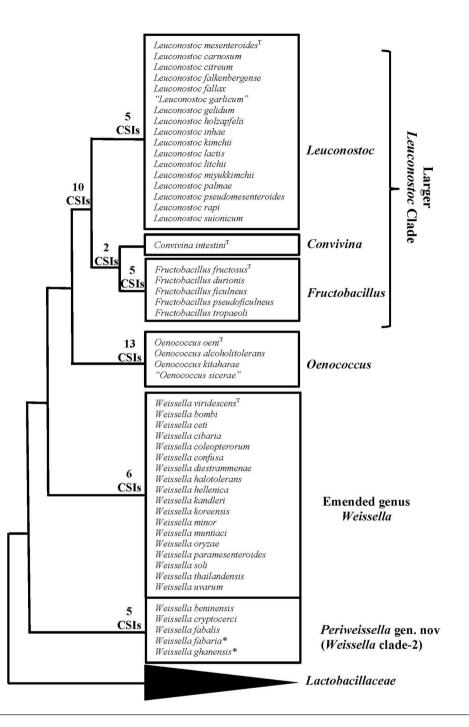


Fig. 7. A conceptual diagram summarizing the results of our phylogenomic and comparative genomic studies on members of the family *Leuconostocaceae*. The numbers of CSIs, which constitute molecular syanpomorphies, that are specifically shared by members of different obsevred clades are shown on the nodes. 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 *Periweissella* 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, 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.coccus. 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

- 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.
- 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.
- 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.
- 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.
- 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.
- Mokoena MP. Lactic acid bacteria and their bacteriocins: classification, biosynthesis and applications against uropathogens: a mini-review. *Molecules* 2017;22:E1255.
- 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.
- 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.
- Van Tieghem P. Sur la gomme de sucrerie (Leuconostoc mesenteroides). Annales des Sciences Naturelles Botanique 1878;7:180–203.
- Sneath PHA, McGowan V, Skerman VBD. Approved lists of bacterial names. Int J Syst Bacteriol 1980;30:225–420.

- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Nel S, Davis SB, Endo A, Dicks LMT. Phylogenetic analysis of Leuconostoc and Lactobacillus species isolated from sugarcane processing streams. *Microbiologyopen* 2020;9:e1065.
- 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.
- 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.
- 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.
- 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.
- Li YQ, Tian WL, Gu CT. Weissella sagaensis sp. nov., isolated from traditional Chinese yogurt. Int J Syst Evol Microbiol 2020;70:2485–2492.
- 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.
- 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 (*Diestram*mena coreana). Int J Syst Evol Microbiol 2013;63:2951–2956.
- Parte AC. LPSN the list of prokaryotic names with standing in nomenclature. Int J Syst Evol Microbiol 2018;68:1825–1829.
- 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.
- 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.
- 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.
- Konstantinidis KT, Tiedje JM. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr Opin Microbiol* 2007;10:504–509.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximumlikelihood approach. *Mol Biol Evol* 2001;18:691–699.
- Le SQ, Lartillot N, Gascuel O. Phylogenetic mixture models for proteins. *Philos Trans R Soc Lond B Biol Sci* 2008;363:3965–3976.
- 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.
- 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.
- 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.
- Parker C, Tindall BJ, Garrity GM. International Code of Nomenclature of Prokaryotes. Int J Syst Evol Microbiol 2019;69:S1–S111.
- 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.

- 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.
- 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.

Five reasons to publish your next article with a Microbiology Society journal

- 1. The Microbiology Society is a not-for-profit organization.
- 2. We offer fast and rigorous peer review average time to first decision is 4–6 weeks.
- 3. Our journals have a global readership with subscriptions held in research institutions around the world.
- 4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
- 5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.