Molecular identification of lactic acid bacteria *Enterococcus*, *Lactobacillus* and *Streptococcus* based on *pheS*, *rpoA* and *atpA* multilocus sequence analysis (MLSA)

Dr. Sabri M. Naser

Department of Biology and Biotechnology

An-Najah National University

Nablus, Palestine

Overview

- General introduction
- Own experimental work
- Results
- Conclusions
- Future perspectives

GENERAL INTRODUCTION

- Lactic acid bacteria (LAB)
- Occurrence of LAB
- Applications
- LAB genera
- Identification of LAB species
- Aims and conceptual framework

Lactic Acid Bacteria

A heterogeneous group:

- **Gram-positive**
- **Catalase-negative**
- **Non-spore forming**
- **Anaerobic bacteria**
- **Strictly fermentative with lactic acid as the key metabolite**

Occurrence of LAB

- Naturally found in dairy, meat, plant and cereal fermentation environments
- Inhabitants in the GIT, the oral cavity, and the vaginal cavity of humans and animals
- Most are commonly referred to as GRAS (Generally Regarded As Safe)
- Some are pathogenic e.g. *S. pneumoniae*
- Are of great economic importance for the dairy and other fermented food products

Applications of LAB

- **Starter cultures**
- **Health promoting products (probiotics)**
- **Flavour, texture and food preservation**

Overview of most important LAB genera

Genus name	No. of Species
Carnobacterium	10
Enterococcus	**35**
Fructobacillus	4
Lactobacillus	**123**
Lactococcus	6
Leuconostoc	13
Oenococcus	2
Pediococcus	11
Streptococcus	**68**
Tetragenococcus	4
Weisella	12

Phenotypic methods used for the identification and delineation of LAB species

- Determination of carbohydrate fermentation
- Enzyme patterns
- Fatty acid analysis
- Determination of cell wall structure
- SDS-PAGE analysis of whole-cell proteins

Limitations of the phenotypic methods

- Labour-intensive
- Variations within species and variations between laboratories
- Low taxonomic resolution

Genotypic methods

Technique | Discriminatory Power

Genus | Species | Strain

Sequencing (e.g. 16S rRNA gene)

RFLP

AFLP

RAPD-PCR

Rep-PCR

PFGE

Ribotyping

DNA-DNA hybridization

MLST

Limitations of the genotypic methods

Genomic fingerprinting methods

- Lack data portability and low inter-laboratory reproducibility

The 16S rRNA gene

- Often lacks resolution when compared with DNA-DNA hybridization

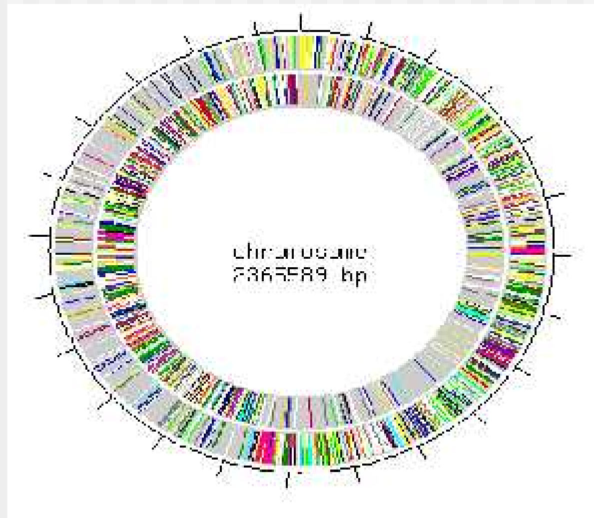
DNA-DNA hybridization

- It is the slowest and the most labour-intensive step in the species description

Complete genome sequencing of LAB

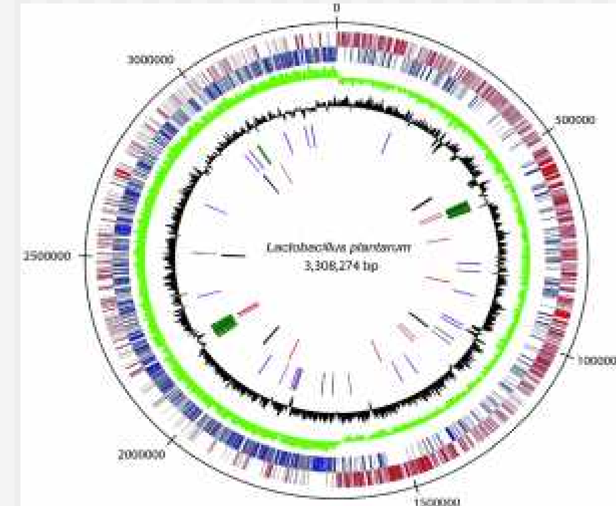
Lc. lactis IL 1403

(Bolotin *et al.*, 2001) (2,365,589 bp)



Lb. plantarum WCFS1

(Kleerebezem *et al.*, 2003) (3,308,274 bp)



OWN EXPERIMENTAL WORK

- Why choose multilocus sequence analysis?
- Aims and conceptual framework
- Methodology

Why choose multilocus sequence analysis?

International Journal of Systematic and Evolutionary Microbiology (2002), 52, 1043–1047 DOI: 10.1099/ijs.0.02360-0

TAXONOMIC NOTE

Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology

Erko Stackebrandt,[1] Wilhelm Frederiksen,[2] George M. Garrity,[3] Patrick A. D. Grimont,[4] Peter Kämpfer,[5] Martin C. J. Maiden,[6] Xavier Nesme,[7] Ramon Rosselló-Mora,[8] Jean Swings,[9] Hans G. Trüper,[10] Luc Vauterin,[11] Alan C. Ward[12] and William B. Whitman[13]

[1] DSMZ–Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, D-38124 Braunschweig, Germany

[2] Statens Seruminstitut

Multilocus sequence typing (MLST)

- **Genotypic characterization using the allelic mismatches of housekeeping genes (internal fragments ~ 450 bp)**
- **Allows definition of strains within named species (typing at intraspecis level)**
- **Population and molecular epidemiological studies**

Multilocus sequence analysis (MLSA)

***In silico* studies based on complete genomes**

Sequences from housekeeping genes

accurately predict genome relatedness

Species identification

Development of MLSA schemes

Multilocus sequence analysis (MLSA)

MLSA is a polygenic scheme that compares the partial DNA sequences from multiple conserved protein coding loci for assessing the diversity and relation of different isolates across related taxa (i.e. identification at species level).

Why multiple genes?

- A single-gene approach may lead to inaccurate estimation of genomic relatedness at species level
- MLSA provides a buffer against the distorting effects of recombination and horizontal gene transfer at a single locus
- Different genes have different discriminatory powers

Why protein-coding genes?

- Show a wider sequence variation
- More rapidly evolving than the more conserved 16S rRNA genes

The selection criteria of candidate genes to be included in MLSA

- Present in single-copy
- Widely distributed among bacterial genomes (at least in the taxon under study)
- Genes in which recombination might confer a selective advantage, or closely linked genes should be avoided
- Contain a sufficient amount of resolution (neither be too conserved nor too variable)

Three core housekeeping genes present in LAB

The genes encoding

- **Phenylalanyl-tRNA synthase alpha subunit (*pheS*, 1100bp)**
- **RNA polymerase alpha subunit (*rpoA*, 1000bp)**
- **ATP synthase alpha subunit (*atpA*, 1500bp)**

The main goal of this study

- **Evaluation of the contribution of MLSA to the species delineation and identification in LAB, particularly the genera *Enterococcus*, *Lactobacillus* and *Streptococcus***

- **Provide rapid, electronically portable, highly reproducible and inexpensive genomic markers that serve as valuable alternative(s) to 16S rRNA gene sequencing**

Evaluation of MLSA to LAB species delineation and identification

Enterococcus

35 species/ 119 strains

Lactobacillus

100 species/ 203 strains

Streptococcus

60 species/152 strains

Selection of protein-coding genes

pheS

rpoA

atpA

Evaluations

Technical

Bioinformatic analysis: TaxonGap

Taxonomic

Methodology

- Oligonucleotide primer design
- PCR optimization
- DNA amplification and sequencing
- Sequence analysis: BioNumerics

Oligonucleotide primer design

- *rpoA*, *pheS* and *atpA* gene sequences of LAB from publicly available data of whole-genome sequence projects
- Kodon program

Species	Strain
Enterococcus faecalis	V583
Lactobacillus plantarum	WCFS1
Lactococcus lactis subsp. *lactis*	IL1403
Streptococcus pneumoniae	TIGR4
	R6
Streptococcus agalactiae	NEM316
	2603 V/R
Streptococcus pyogenes	MGAS8232
	SSI-1
	MGAS315
	SF370
Streptococcus mutans	UA159

Oligonucleotide primer design

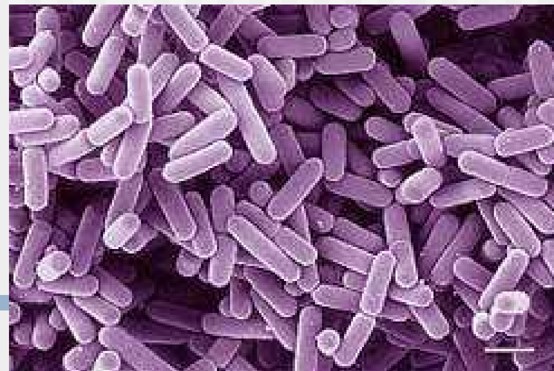
Gene	Primer name	Sequence (5'→3')	Position
pheS	pheS-21-F	CAYCCNGCHCGYGAYATGC	557
	pheS-22-R	CCWARVCCRAARGCAAARCC	1031
	pheS-23-R	GGRTGRACCATVCCNGCHCC	968
rpoA	rpoA-21-F	ATGATYGARTTTGAAAAACC	1
	rpoA-22-R	ACYTTVATCATNTCWGVYTC	844
	rpoA-23-R	ACHGTRTTRATDCCDGCRCG	802
atpA	atpA-20-F	TAYRTYGGKGAYGGDATYGC	97
	atpA-22-F	GCWCCYGGTRTYATGCARCG	397
	atpA-23-R	CGYTGCATRAYACCRGGWGC	397
	atpA-24-F	GATGAYYTWTCAAARCAAGC	781
	atpA-25-R	GCTTGYTTTGAWARRTCATC	781
	atpA-27-R	CCRCGRTTHARYTTHGCYTG	1219

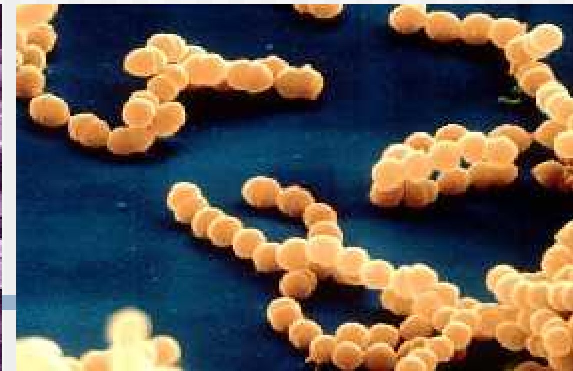
PCR optimization

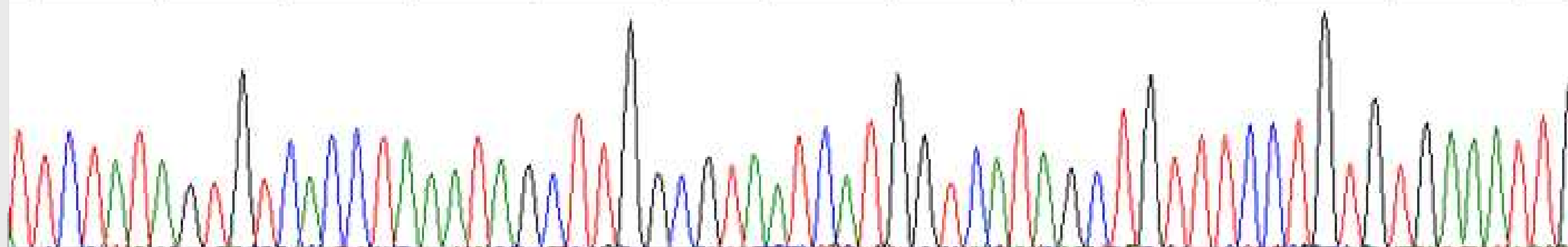
- **Different primer combinations for each gene (4-6)**
- **Different annealing temperatures (42-60°C)**
- **Different $MgCl_2$ concentrations (1-4 µM)**

DNA amplification and sequencing

- **Amplification of target genes by PCR**
- **Dideoxy-termination sequencing reactions using internal and/or amplification primers**
- ***pheS* (382-455 nt), *rpoA* (402-694 nt) and *atpA* (611-1102 nt)**



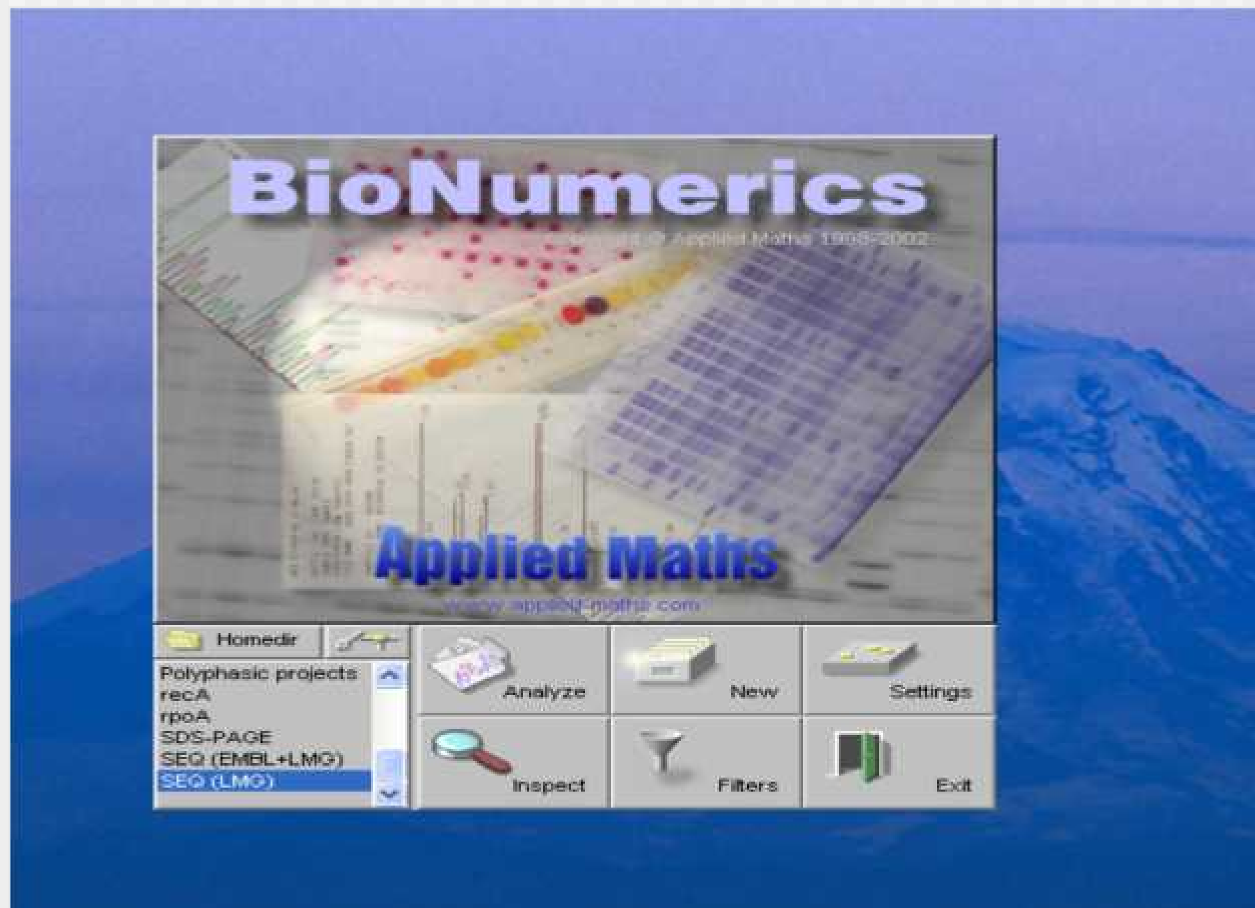




```
   1 ATTGAACGCT GGCGGCATGC CTTACACATG CAAGTCGAAC GGCAGCACGG GTGCTTGCAC cTGGTGGcGA GTGGCGAACG GGTGAGTAAT   90
  91 ACATCGGAAC ATGTCCTGTA GTGGGGGATA GCCCGGcGAA AGCCGGATTA ATACCGCATA CGATCyACGG ATGAAAGCGG GGGACCTTCG  180
 181 GGCCTCGCGC TATAGGGTTG GCCGATGGCT GATTAGcTAG TTGGTGGGGT AAAGGCCTAC CAAGGcGACG ATCAGTAGCT GGTcTGAGAG  270
 271 GACGACCAGC CACACTGGGA CTGAGACACG GCCCAGACTC CTACGGGAGG CAGCAGTGGG GAATTTTGGA CAATGGGCGA AAGCCTGATC  360
 361 CAGCAATGCC GCGTGTGTGA AGAAGGCCTT cGGGTTGTAA AGCACTTTTg TCCGGAAAGA AATCCtTGGc TCTAATACAG TCGGGGGATG  450
 451 ACGGTACCGG AAGAATAAGC ACCGGCTAAC TACGTGCCAG CAGCCGCGGT AATACGTAGG GTGCGAGCGT TAATCGGAAT TACTGGGCGT  540
 541 AAAGCGTGCG CAGGCGGTCT GTTAAGACAG ATGTGAAATC CCCGGGCTCA ACCTGGGAAC TGCATTTGTG ACTGGCAGGC TAGAGTATGG  630
 631 CAGAGGGGGG TAGAATTCCA CGTGTAGCAG TGAAATGCGT AGAGATGTGG AGGAATACCG ATGGCGAAGG CAGCCCCCTG GGCCAATACT  720
 721 GACGCTCATG CACGAAAGCG TGGGGAGCAA ACAGGATTAG ATACCCTGGT AGTCCACGCC CTAAACGATG TCAACTAGTT GTTGGGGATT  810
 811 CATTTCCTTA GTAACGTAGC TAACGCGTGA AGTTGACCGC CTGGGGAGTA CGGTCGCAAG ATTAAAACTC AAAGGAATTG ACGGGGACCC  900
 901 GCACAAGCGG TGGATGATGT GGATTAATTC GATGCAACGC GAAAAACCTT ACCTACCCTT GACATGGTCG GAATCCTGAA GAGATTCGGG  990
 991 AGTGCTCGAA AGAGAACCGG CGCACAGGTG CTGCATGGCT GTCGTCAGCT CGTGTCGTGA GATGTTGGGT TAAGTCCCGC AACGAGCGCA 1080
1081 ACCCTTGTCC TTAGTTGCTA CGCAAGAGCA CTCTAAGGAG ACTGCCGGTG ACAAACCGGA GGAAGGTGGG GATGACGTCA AgTCCTCATG 1170
1171 GCCCTTATGG GTAGGGCTTC ACACGTCATA CAATGGTCGG AACAgAGGGT TGCCAaCCCG CGAGGGGGAG CTAATCCCAG AAAACCGATC 1260
1261 GTAGTCCGGA TTGCACTCTG CAACTCGAGT GCATGAAGCT GGAATCGCTA GTAATCGCGG ATCAGCATGC CGCGGTGAAT ACGTTCCCGG 1350
1351 GTCTTGTACA CACCGCCCGT CACACCATGG GAGTGGGTTT TACCAGAAGT GGCTAGTCTA ACCGCAAGGA GGACGGTCAC CACGGTAGGA 1440
1441 TTCATGACTG GGGTGAAGTC GTAACAAGGT AGCCGTATCG GAAGG                                                1530
```

Sequence analysis: BioNumerics

Construction of *pheS*, *rpoA* and *atpA* databases

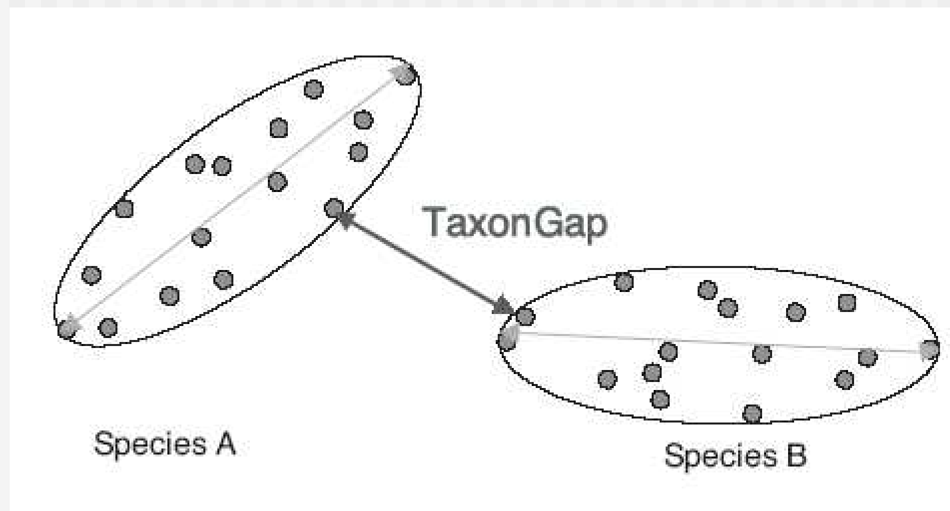


RESULTS

- Interpretation of MLSA data: TaxonGap
- Intraspecies variation
- Interspecies gaps
- Congruency of MLSA data with 16S rRNA gene
- Description of new taxa
- Reclassifications

Interpretation of MLSA data: TaxonGap

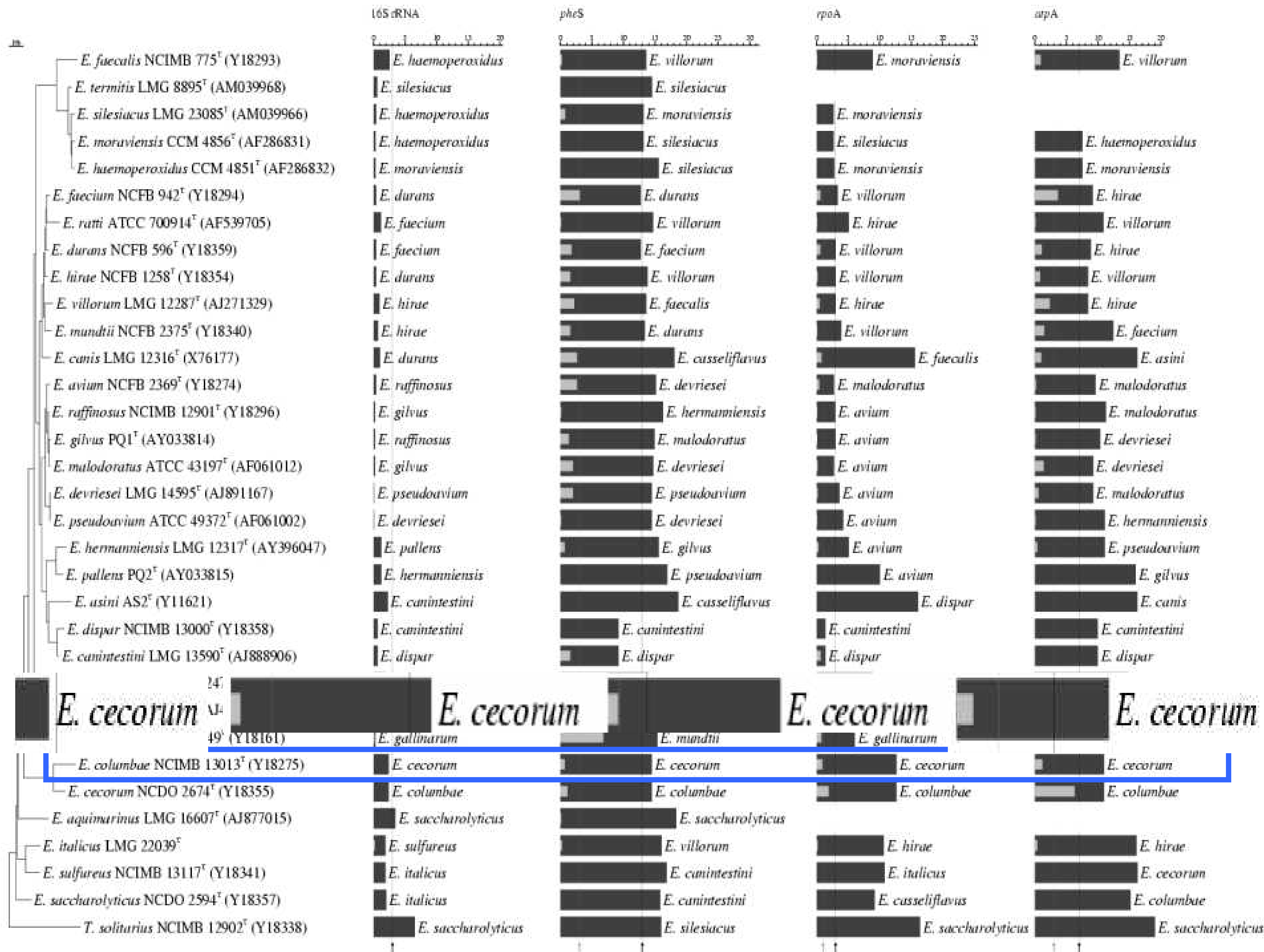
- The intraspecies diversity represents the maximum sequence distance among strains of the species
- The TaxonGap represents the minimum distance between a species A and its closest neighbour species B



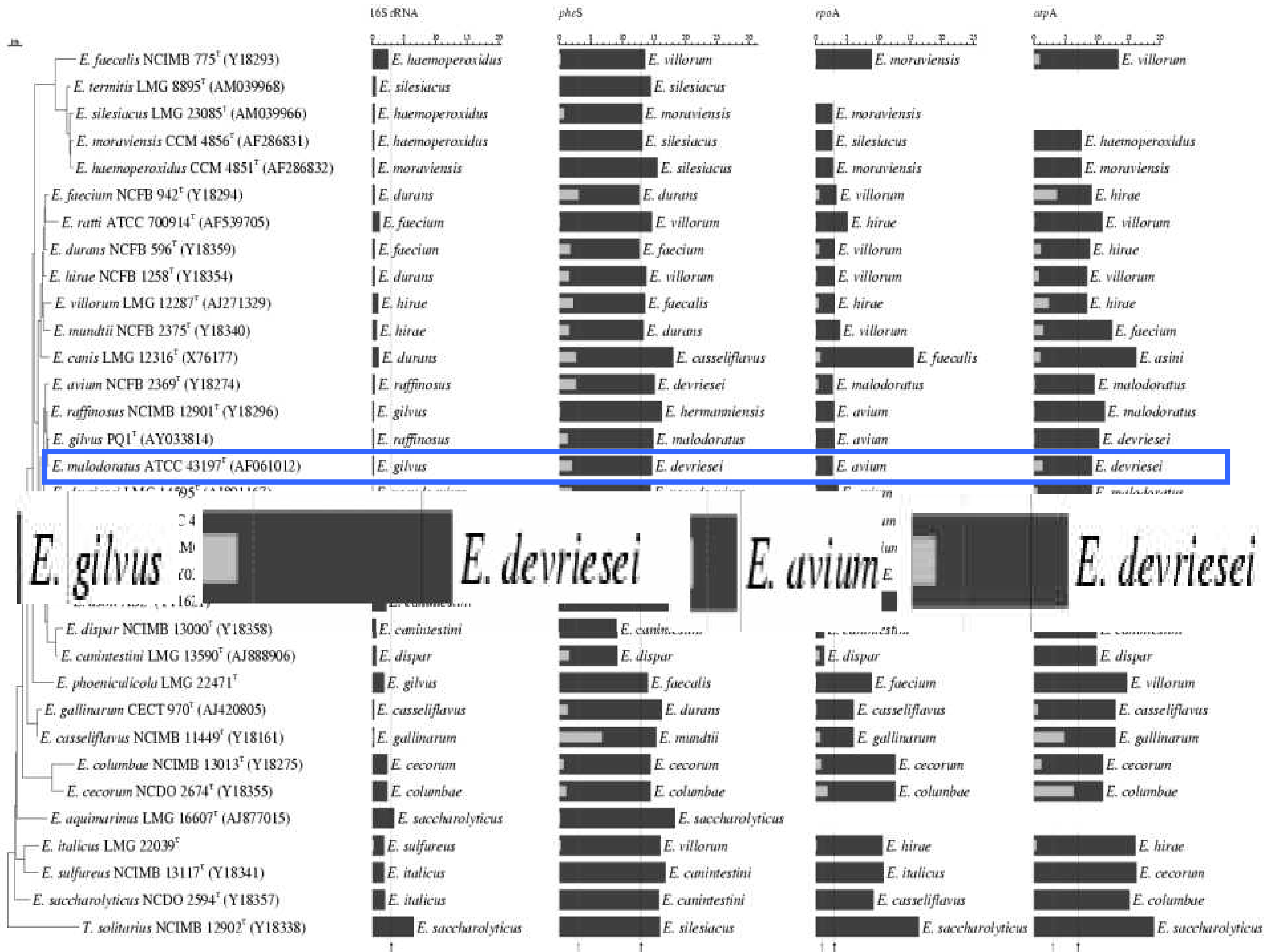
Interpretation of MLSA data: TaxonGap

- Based on a pairwise distance matrix derived from the aligned sequences for all investigated strains
- Subsequently calculates the intraspecies and interspecies variations of the genes in the MLSA scheme

16S rRNA
pheS
rpoA
atpA
E. faecalis NCIMB 775T (Y18293)
E. termitis LMG 8895T (AM039968)
E. silesiacus LMG 23085T (AM039966)
E. moraviensis CCM 4856T (AF286831)
E. haemoperoxidus CCM 4851T (AF286832)
E. faecium NCFB 942T (Y18294)
E. durans
E. durans
E. villorum
E. hirae
E. canis LMG 12316T (X76177)
E. avium NCFB 2369T (Y18274)
E. raffinosus NCIMB 12901T (Y18296)
E. gilvus PQ1T (AY033814)
E. malodoratus ATCC 43197T (AF061012)
E. devriesei LMG 14595T (AJ891167)
E. pseudoavium ATCC 49372T (AF061002)
E. hermanniensis LMG 12317T (AY396047)
E. pallens PQ2T (AY033815)
E. asini AS2T (Y11621)
E. dispar NCIMB 13000T (Y18358)
E. canintestini LMG 13590T (AJ888906)
E. phoeniculicola LMG 22471T
E. gallinarum CECT 970T (AJ420805)
E. casseliflavus NCIMB 11449T (Y18161)
E. columbae NCIMB 13013T (Y18275)
E. cecorum NCDO 2674T (Y18355)
E. aquimarinus LMG 16607T (AJ877015)
E. italicus LMG 22039T
E. sulfureus NCIMB 13117T (Y18341)
E. saccharolyticus NCDO 2594T (Y18357)
T. solitarius NCIMB 12902T (Y18338)



16S rRNA
pheS
rpoA
atpA
E. faecalis NCIMB 775T (Y18293)
E. termitis LMG 8895T (AM039968)
E. silesiacus LMG 23085T (AM039966)
E. moraviensis CCM 4856T (AF286831)
E. haemoperoxidus CCM 4851T (AF286832)
E. faecium NCFB 942T (Y18294)
E. ratti ATCC 700914T (AF539705)
E. durans NCFB 596T (Y18359)
E. hirae NCFB 1258T (Y18354)
E. villorum LMG 12287T (AJ271329)
E. mundtii NCFB 2375T (Y18340)
E. canis LMG 12316T (X76177)
E. avium NCFB 2369T (Y18274)
E. raffinosus NCIMB 12901T (Y18296)
E. gilvus PQ1T (AY033814)
E. malodoratus ATCC 43197T (AF061012)
E. devriesei LMG 14595T (AJ891167)
E. pseudoavium ATCC 49372T (AF061002)
E. hermanniensis LMG 12317T (AY396047)
E. pallens PQ2T (AY033815)
E. asini AS2T (Y11621)
E. dispar NCIMB 13000T (Y18358)
E. canintestini LMG 13590T (AJ888906)
(Y18161)
E. columbae NCIMB 13013T (Y18275)
E. cecorum NCDO 2674T (Y18355)
E. aquimarinus LMG 16607T (AJ877015)
E. italicus LMG 22039T
E. sulfureus NCIMB 13117T (Y18341)
E. saccharolyticus NCDO 2594T (Y18357)
T. solitarius NCIMB 12902T (Y18338)
E. haemoperoxidus
E. silesiacus
E. haemoperoxidus
E. haemoperoxidus
E. moraviensis
E. durans
E. faecium
E. faecium
E. durans
E. hirae
E. hirae
E. durans
E. raffinosus
E. gilvus
E. raffinosus
E. gilvus
E. pseudoavium
E. devriesei
E. pallens
E. hermanniensis
E. canintestini
E. canintestini
E. dispar
E. gallinarum
E. cecorum
E. columbae
E. saccharolyticus
E. sulfureus
E. italicus
E. italicus
E. saccharolyticus
E. villorum
E. silesiacus
E. moraviensis
E. silesiacus
E. silesiacus
E. durans
E. villorum
E. faecium
E. villorum
E. faecalis
E. durans
E. casseliflavus
E. devriesei
E. hermanniensis
E. malodoratus
E. devriesei
E. pseudoavium
E. devriesei
E. gilvus
E. pseudoavium
E. casseliflavus
E. canintestini
E. dispar
E. mundtii
E. cecorum
E. columbae
E. saccharolyticus
E. villorum
E. canintestini
E. canintestini
E. silesiacus
E. moraviensis
E. moraviensis
E. silesiacus
E. moraviensis
E. villorum
E. hirae
E. villorum
E. villorum
E. hirae
E. villorum
E. faecalis
E. malodoratus
E. avium
E. avium
E. avium
E. avium
E. avium
E. avium
E. avium
E. dispar
E. canintestini
E. dispar
E. gallinarum
E. cecorum
E. columbae
E. hirae
E. italicus
E. casseliflavus
E. saccharolyticus
E. villorum
E. haemoperoxidus
E. moraviensis
E. hirae
E. villorum
E. hirae
E. villorum
E. hirae
E. faecium
E. asini
E. malodoratus
E. malodoratus
E. devriesei
E. devriesei
E. malodoratus
E. hermanniensis
E. pseudoavium
E. gilvus
E. canis
E. canintestini
E. dispar
E. cecorum
E. columbae
E. hirae
E. cecorum
E. columbae
E. saccharolyticus


16S rRNA
pheS
rpoA
atpA
E. faecalis NCIMB 775T (Y18293)
E. termitis LMG 8895T (AM039968)
E. silesiacus LMG 23085T (AM039966)
E. moraviensis CCM 4856T (AF286831)
E. haemoperoxidus CCM 4851T (AF286832)
E. faecium NCFB 942T (Y18294)
E. ratti ATCC 700914T (AF539705)
E. durans NCFB 596T (Y18359)
E. hirae NCFB 1258T (Y18354)
E. villorum LMG 12287T (AJ271329)
E. mundtii NCFB 2375T (Y18340)
E. canis LMG 12316T (X76177)
E. avium NCFB 2369T (Y18274)
E. raffinosus NCIMB 12901T (Y18296)
E. gilvus PQ1T (AY033814)
E. malodoratus ATCC 43197T (AF061012)
E. dispar NCIMB 13000T (Y18358)
E. canintestini LMG 13590T (AJ888906)
E. phoeniculicola LMG 22471T
E. gallinarum CECT 970T (AJ420805)
E. casseliflavus NCIMB 11449T (Y18161)
E. columbae NCIMB 13013T (Y18275)
E. cecorum NCDO 2674T (Y18355)
E. aquimarinus LMG 16607T (AJ877015)
E. italicus LMG 22039T
E. sulfureus NCIMB 13117T (Y18341)
E. saccharolyticus NCDO 2594T (Y18357)
T. solitarius NCIMB 12902T (Y18338)
E. haemoperoxidus
E. silesiacus
E. haemoperoxidus
E. haemoperoxidus
E. moraviensis
E. durans
E. faecium
E. faecium
E. durans
E. hirae
E. hirae
E. durans
E. raffinosus
E. gilvus
E. raffinosus
E. gilvus
E. canintestini
E. dispar
E. gilvus
E. casseliflavus
E. gallinarum
E. cecorum
E. columbae
E. saccharolyticus
E. sulfureus
E. italicus
E. italicus
E. saccharolyticus
E. villorum
E. silesiacus
E. moraviensis
E. silesiacus
E. silesiacus
E. durans
E. villorum
E. faecium
E. villorum
E. faecalis
E. durans
E. casseliflavus
E. devriesei
E. hermanniensis
E. malodoratus
E. devriesei
E. dispar
E. faecalis
E. durans
E. mundtii
E. cecorum
E. columbae
E. saccharolyticus
E. villorum
E. canintestini
E. canintestini
E. silesiacus
E. moraviensis
E. moraviensis
E. silesiacus
E. moraviensis
E. villorum
E. hirae
E. villorum
E. villorum
E. hirae
E. villorum
E. faecalis
E. malodoratus
E. avium
E. avium
E. avium
E. dispar
E. faecium
E. casseliflavus
E. gallinarum
E. cecorum
E. columbae
E. hirae
E. italicus
E. casseliflavus
E. saccharolyticus
E. villorum
E. haemoperoxidus
E. moraviensis
E. hirae
E. villorum
E. hirae
E. villorum
E. hirae
E. faecium
E. asini
E. malodoratus
E. malodoratus
E. devriesei
E. devriesei
E. dispar
E. villorum
E. casseliflavus
E. gallinarum
E. cecorum
E. columbae
E. hirae
E. cecorum
E. columbae
E. saccharolyticus
E. gilvus
E. devriesei
E. avium
E. devriesei

Intraspecies variation

- **The *rpoA*, *atpA*, and *pheS* gene sequence analyses provided an intraspecies variation 1-3%**

- **Strains of same species had a high degree of homogeneity**

Interspecies gaps

The interspecies variations between individual species and the nearest neighbour using TaxonGap software

- ***Enterococcus*: 12.75%, 2.7%, 7.5% *pheS, rpoA, atpA***
- ***Lactobacillus*: 6.7%, 3% *pheS*, *rpoA***
- ***Streptococcus*: 2%, 1.3%, 1.5% *pheS, rpoA, atpA***

Compared to 0.23-0.26% 16S rRNA gene!

Congruency of MLSA data with 16S rRNA gene

The ability of a specific gene to recognize the 16S rRNA gene – based species groups, varies among the investigated genera

The genus *Enterococcus*

Both *atpA* and *rpoA* genes are congruent to 16S rRNA gene, whereas the *pheS* gene shows no congruency

The genus *Lactobacillus*

Both *pheS* and *rpoA* exhibit species clustering mostly correlated with 16S rRNA gene

Congruency of MLSA data with 16S rRNA gene

The genus *Streptococcus*

Occupies an intermediate position where the *pheS* gene is the mostly correlated with 16S rRNA gene. Other loci show less congruency

Description of new taxa

- MLSA data is shown to be an efficient screening method for the detection of novel taxa
- In this study, seven novel species were initially detected by sequence analysis of gene(s) included in the MLSA scheme
- Further phenotypic and genotypic data confirmed the MLSA data

Description of new taxa

Genus	New taxa
Enterococcus	*Enterococcus aquimarinus* sp. nov.
	Enterococcus canintestini sp. nov
	Enterococcus devriesei sp. nov.
	Enterococcus silesiacus sp. nov.
	Enterococcus termitis sp. nov.
Lactobacillus	*Lactobacillus parabrevis* sp. nov.
	Lactobacillus amylotrophicus sp. nov.

Reclassifications

Genus	Emended description	Junior synonym name
Enterococcus	*E. casseliflavus*	*E. flavescens*
	E. italicus	*E. saccharominimus*
Lactobacillus	*L. helveticus*	*L. suntoryeus*
	L. acidipiscis	*L. cypricasei*

CONCLUSIONS

- **MLSA data can be used as reliable tools for the identification of clinical and environmental species of the genera *Enterococcus*, *Lactobacillus* and *Streptococcus***
- **The use of partial sequences of *pheS*, *rpoA* and *atpA* genes provides a rapid and low cost tool for species identification**
- **The sequencing of housekeeping loci provides unambiguous, electronically portable and highly reproducible data**

CONCLUSIONS

- ***pheS*, *rpoA* and *atpA* loci are informative in more than one group and provide tools for broader comparisons**
- **TaxonGap provides a straightforward evaluation of the discriminatory power of the genes in the MLSA scheme**

FUTURE PERSPECTIVES

The construction of a central curated database in which MLSA data of LAB can be stored and freely accessed online

- **The present MLSA scheme paved the way to extend the study of other LAB genera using *pheS*, *rpoA*, and *atpA* gene sequences**

- **All *Leuconostoc*, *Fructobacillus*, *Lactococcus*, *Weisella*, *Oenococcus*, and *Pediococcus* species are clearly delineated based on MLSA scheme. (De Bruyne, K. 2009. Ghent University, Belgium)**

- **The available *pheS*, *rpoA*, and *atpA* gene sequence data have already been used to identify isolates from raw milk, different food fermentations**

Acknowledgements

Prof. Dr. Ir. Jean Swings

Department of Microbiology, Ghent University, K.L. Ledeganckstraat 35, Ghent 9000, Belgium

Thank You