USP Virtual Symposium

Emerging technologies

in **probiotic, live biotherapeutic product** and **microbiome analysis**

Oct. 6-7, 8:30-11:30am ET

Challenges and opportunities of DNA based identification, characterization and authentication of probiotic strains

Antonio Del Casale - MICROBION, Italy

microbion

About MICROBION

Contract Research Organization (CRO):

- serving agri-food and pharma-nutraceutical industry
- customization of DNA technologies
- industrial microbiology including
  - ✓ beneficial microbes
  - ✓ microbial contaminants
  - ✓ microbial communities

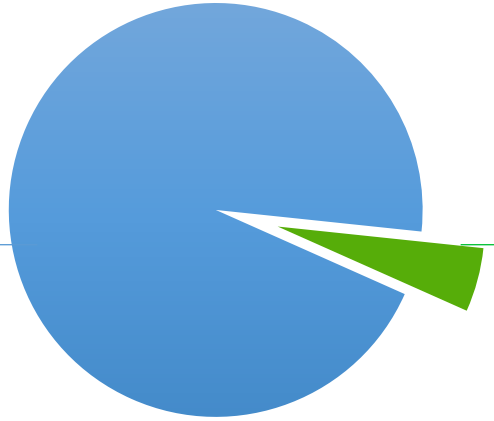
We are

✓Problem solvers

✓Innovation enablers

About MICROBION

Microbial Biodiversity



Naturally Occurring
- Biofilms, spores, heat-resistant
- Emerging resistant strains

Spoilage and outbreaks
Root-cause analysis
Food Quality & Safety

Product stability, shelf-life
Challenge-test
Process Validation

Risk assessment
Antibiotic resistances
Hazards

Identification, rapid kits
Strain tracking
Molecular Diagnostics

Industrial microbes
- Selected, manufactured and applied
- Controlled spontaneous communities

Bio-fertilisers
Bio-pesticides
Agriculture

Cultured food & beverages
Spontaneous fermentations
Food

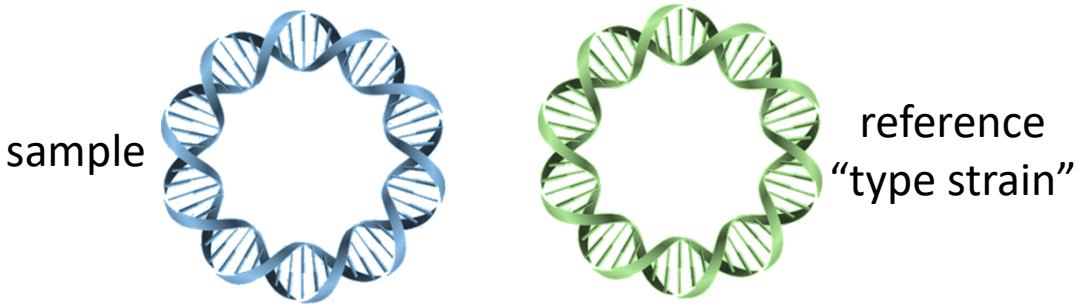
Probiotics, pre/post-biotics
Microbiota modulation
Health

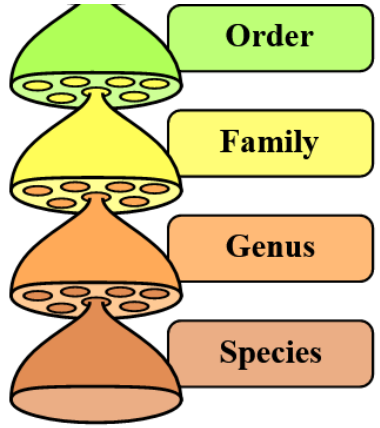
Bio-fuels, bio-plastics
Bio-preservatives
Green Chemistry

What is a bacterial species?

Group of “similar” strains sharing:

- DNA-DNA hybridization >70%
- 16S rRNA gene sequence identity >98,7%
- **Average Nucleotide Identity (genome) ≥96%**

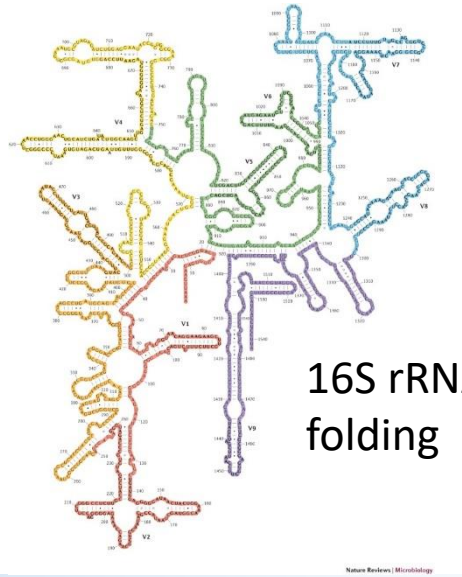




16S rRNA sequence the routine identification marker:

- present in every species
- slow mutation rate, with conservative regions and variable regions
- is known and available for every classified species (public database)

- **<u>do not distinguish</u>** some groups of species (and sub-species)
- Sanger sequencing gives about 50% of 16S rRNA
- public database contain misidentified and not approved species



16S rRNA folding

Solutions for accurate species identification

Solution for accurate species identification:

First step

- ✅ "full" 16S rRNA sequence (> 1200 bp)
- ✅ comparison with a qualified database (no BLAST)
- ❌ if no good match….
  - ➡ phylogeny

Additional steps if 16S rRNA do not give clear results

(<u>without</u> genome sequencing)

- ✅ group-specific genetic markers (e.g. *rpoB*, *dnaB*, *purH*)
- ✅ multi-locus approach
- ❌ if no good match….
  - ➡ phylogeny

(<u>with</u> genome sequence)

- ✅ ANI vs genome from databases
- ❌ if no good match….
  - ➡ phylogenomics

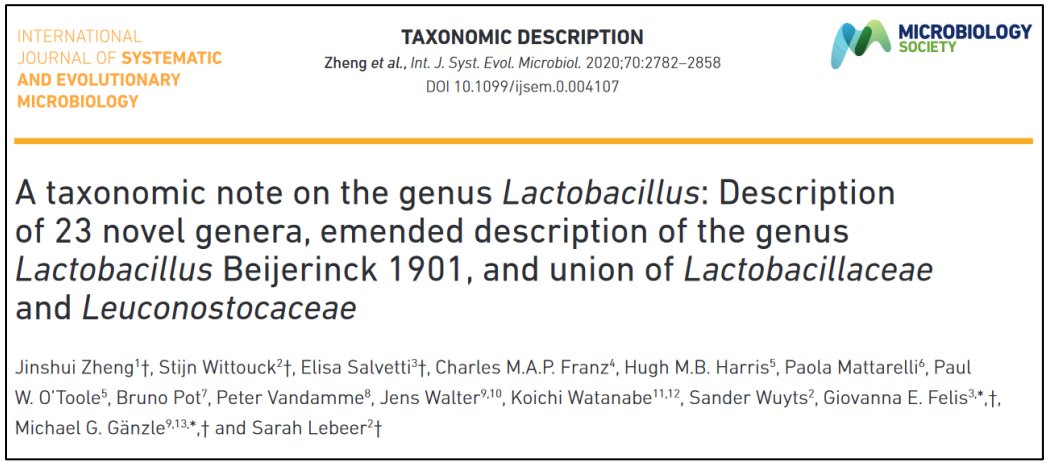
  (find the taxonomic "place")

Taxonomy is always subject to updates

The more we know, the more we need to find spaces in the “catalogues” of microbial biodiversity

Lactobacillus species divided in 23 new genera (2020)

…who’s the next one?!



INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

TAXONOMIC DESCRIPTION

Zheng *et al.*, *Int. J. Syst. Evol. Microbiol.* 2020;70:2782–2858
DOI 10.1099/ijsem.0.004107

MICROBIOLOGY SOCIETY

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*

Jinshui Zheng[1]†, Stijn Wittouck[2]†, Elisa Salvetti[3]†, Charles M.A.P. Franz[4], Hugh M.B. Harris[5], Paola Mattarelli[6], Paul W. O'Toole[5], Bruno Pot[7], Peter Vandamme[8], Jens Walter[9,10], Koichi Watanabe[11,12], Sander Wuyts[2], Giovanna E. Felis[3]*,†, Michael G. Gänzle[9,13]*,† and Sarah Lebeer[2]†

Bacillus clausii (1995)

↓

Alkalihalobacillus clausii (2020)

↓

Shouchella clausii (2022)

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

VALIDATION LIST

Oren and Garrity, *Int. J. Syst. Evol. Microbiol.* 2022;72:005167
DOI 10.1099/ijsem.0.005167

MICROBIOLOGY SOCIETY

Valid publication of new names and new combinations effectively published outside the IJSEM. Validation List no. 203

Aharon Oren[1,*] and George M. Garrity[2,*]

What is a bacterial strain?

Legal definition:

IDA deposit (Budapest treaty) → Accession number ← Clinical trials, Industrial properties, Intellectual property rights

Biology definition:

No formal definition
No formal similarity thresholds
Mutations happen continuously

↓

Use of Mater Cell Banks and Working Cell Banks
Does 1 SNP make a new strain? …**yes but actually NO!**

Sure facts, different strains show (one or more):

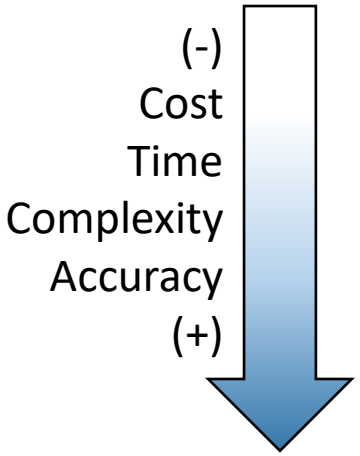
- different phenotypes
- different DNA fingerprinting profiles

↓

We can prove difference, not (easily) identity

"We know what we are,
but know not what we may be"
W. Shakspeare

Solutions for strain authentication



Available solutions:

- PCR-based fingerprinting profiles
- single gene sequence analysis
- multi-gene sequence analysis (e.g. MLST or MLSA)
- whole genome optical mapping (similar to PFGE)
- whole genome sequencing + phylogenomic approaches (e.g. wgMLST pan-genome and/or MLSA core-genome)

Under development:

- definition of species mutation rate

- definition of number of SNP to qualify as different strain in each species

Design of species/strain-specific assays

Can we design "specific" assays for bacteria in blends?

Species specificity:
- ~~selective media for plate counts (?)~~
- PCR assays based on 16S rRNA

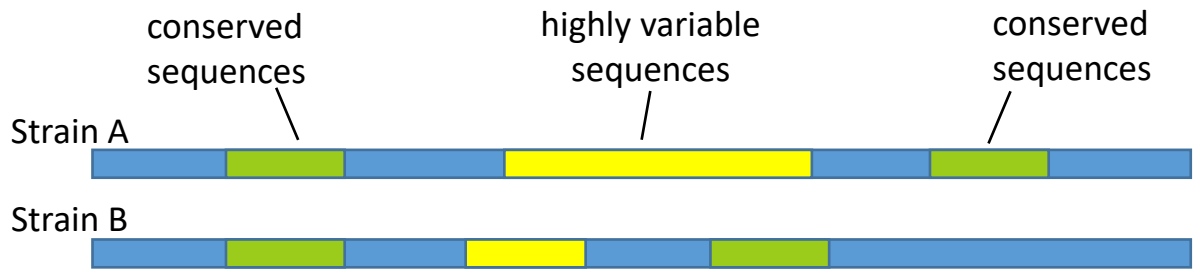
yes, but…
not for close **related species** and subspecies
e.g. blend of *L. casei/paracasei/rhamnosus*
B. animalis subsp. *animalis/lactis*

More specificity, always for a **defined set of strains**:
- PCR assays based on signature sequences (+/-)
- Highly Polymorphic and Modular Extragenic (HPME)

set of primers/probes for multi-purpose assay

MICROBION's HPME technology

(patent pending WO2018014979A1)

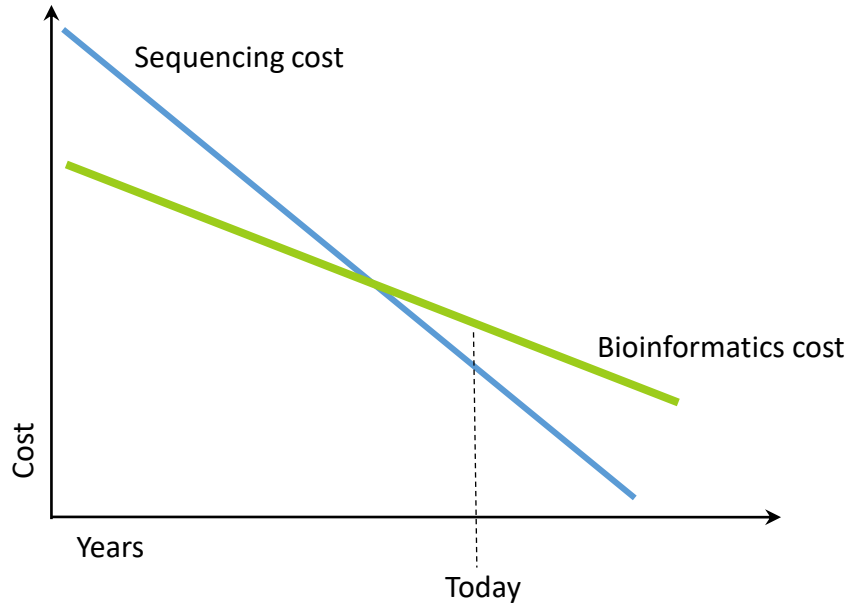


different PCR products for
each species in the blend:
- different lengths for easy and fast detection
- different sequences providing additional specificity

- check blend composition
- quantify species/strain in the blend

Genome sequencing

- genome sequencing is becoming a routine analysis
- bioinformatics is the new bottleneck

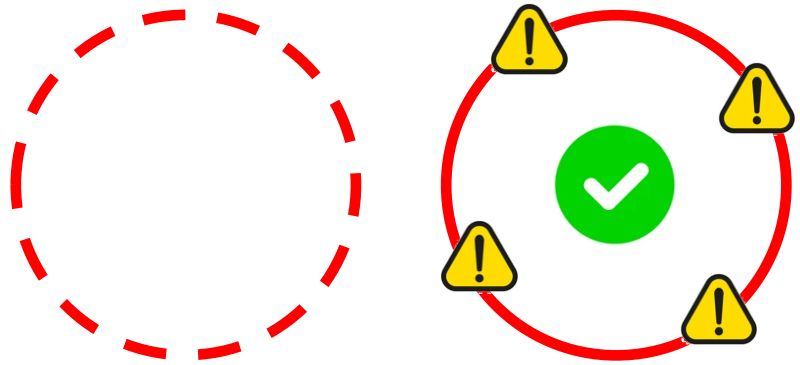


Short reads

- low cost
- high fidelity
- fragmented genome
- no plasmids, repetitive sequences, etc...

Long reads

- fully assembled genome
- assembled plasmids, repetitive sequences, etc...
- higher cost
- lower fidelity



“hybrid” approach

Genome anaysis

European Food Safety Authority

2018 - Guidance on the characterisation of microorganisms used as feed additives or as production organisms

2020 - EFSA statement on the requirements for whole genome sequence analysis of microorganisms intentionally used in the food chain

Genome analyses

EFSA guidelines genome quality:

- > 30x coverage (suggested > 100x)
- < 500 fragments **(best is 1 +plasmids)**
- < 5% contaminants reads
- total contigs length +/- 20% of expected genome size
- report assembly strategy, assembler software version, statistics and parameters of annotation

EFSA guidelines genome analyses:

- identification ➡ 16S rRNA **and** ANI
- anti-microbial resistances ➡ **recommended 2** databases
- virulence factors (if not a QPS species)

MICROBION suggested analyses:

- secondary metabolites (e.g. bacteriocines)
- plasmids, prophages and transposons search
- biogenic ammines of family *Lactobacillaceae*
- toxic metabolites (e.g. toxins of *Bacillus* spp.)
- exopolysaccharides
- other “probotic” traits (e.g. adhesion genes, cross-feeding)

Take home messages

a. species identification is not easy task,
mind your marker/s and your database/s
(be ready to taxonomy changes)

b. absolute strain specificity is not possible,
but is possible within a given set of strains
(can easily prove difference, not identity)

c. cheap (fragmented) genome sequencing is just ok for R&D,
not enough for products and regulatory compliance

Antonio Del Casale
a.delcasale@microbion.it
tel. +39 (0)457 1240

Follow us on LinkedIn