

Deciphering intra-species bacterial diversity using *gyrB* amplicon sequencing

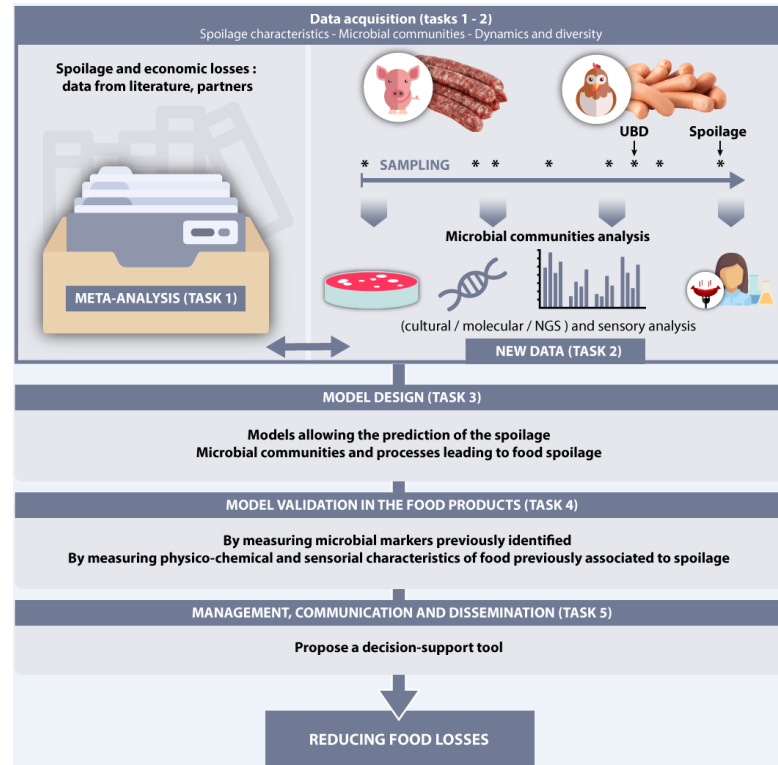
A comparative analysis with 16S rDNA V3-V4 amplicon
sequencing

Olivier Rué

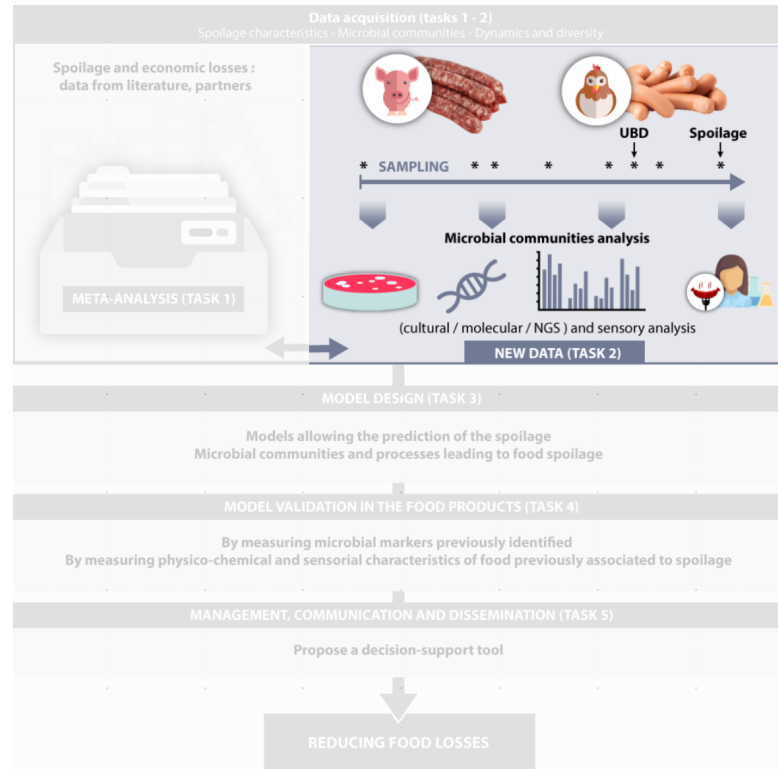
July 4, 2019



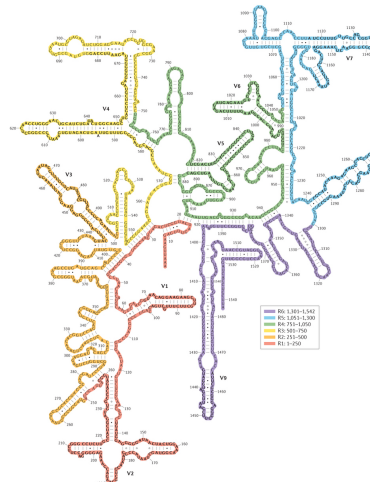
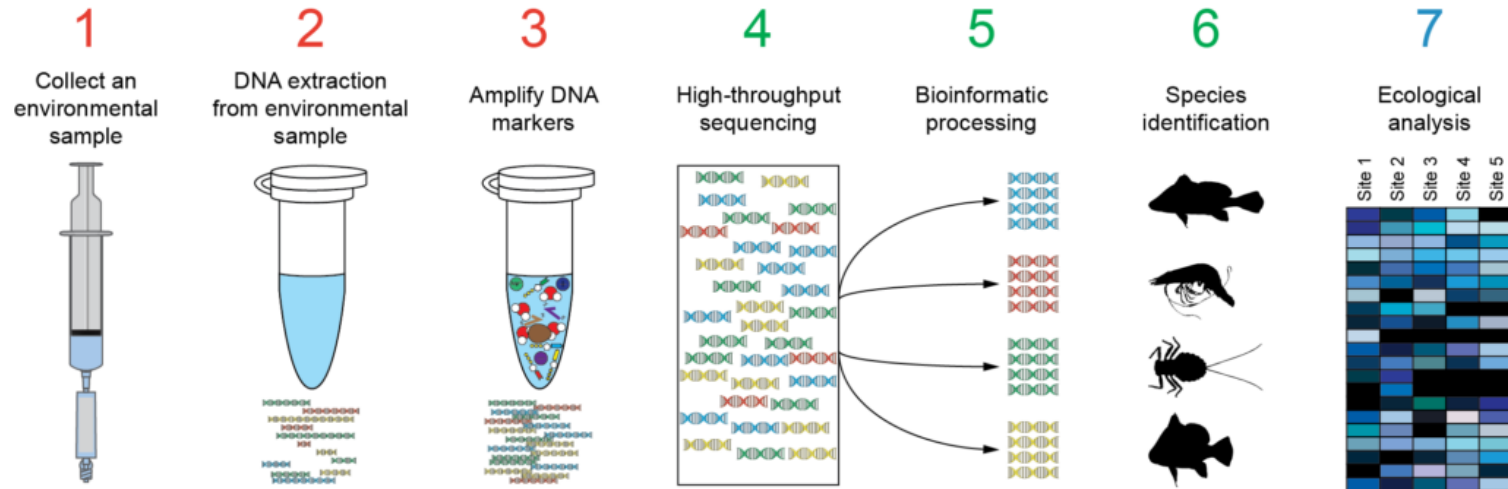
REDucing food LOSSES by microbial spoilage prediction



REDucing food LOSSES by microbial spoilage prediction



Diversity using metabarcoding



Meat spoilage

Knowledge

- Predominance of Bacteria
- Between 15 to 80 species
- Limited phylogenetic diversity
 - mainly Firmicutes and Proteobacteria
- Large majority is known and cultivable

16S rDNA profiling limitations

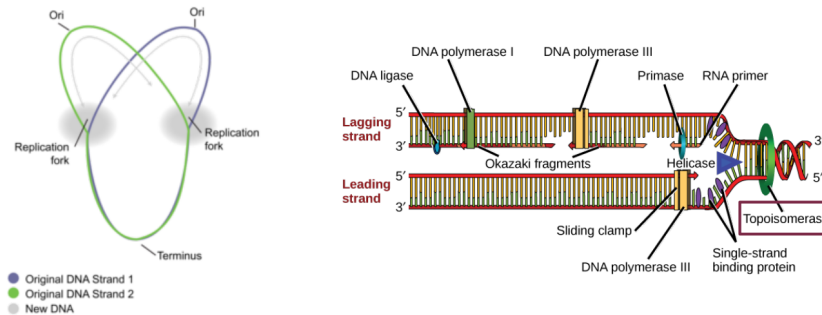
- Multiple copies of 16S
- Amplification biases
- Over-estimation of diversity
- 🚫 **Genus-level resolution**

⚠️ Food quality is highly correlated to **strain's** diversity ! ⚠️

Need an alternative solution to assess with precision the strain's diversity

Potential advantages of *gyrB* (gyrase β subunit)

Characteristics



- Gene encoding for DNA gyrase (essential to DNA replication)
 - **Ubiquist**
 - Present in Bacteria, Eukaryota and sometimes Archae
 - 280 nt long - less secondary structures (chimera)
 - **One copy** (rare exceptions)
 - No amplification of chloroplast and mitochondrial rDNA
 - **Higher nucleotide mutation drift** than 16S

Potential advantages of *gyrB* (gyrase β subunit)

gyrB metabarcoding ?

- Experienced in seed microbiota
- Highly degenerated primers (~2,000 times) designed in a conserved AA region



Emergence Shapes the Structure of the Seed Microbiota

Matthieu Barret,^{a,b,c} Martial Briand,^{a,b,c} Sophie Bonneau,^{a,b,c} Anne Prévieux,^{a,b,c} Sophie Valière,^{d,e} Olivier Bouchez,^{d,f}
Gilles Hunault,^g Philippe Simoneau,^{a,b,c} Marie-Agnès Jacques^{a,b,c}

INRA, UMR1345 Institut de Recherche en Horticulture et Semences, Beaucauzé, France^a; Agrocampus Ouest, UMR1345 Institut de Recherche en Horticulture et Semences, Beaucauzé, France^b; Université d'Angers, UMR1345 Institut de Recherche en Horticulture et Semences, SFR4207 QUASAV, Beaucauzé, France^c; GeT-PlaGe, Genotoul, INRA Auzeville, Castanet-Tolosan, France^d; INRA, UAR1209, Département de Génétique Animale, INRA Auzeville, Castanet Tolosan, France^e; UMR INRA/INPT ENSAT/NPT ENVT, Génétique, Physiologie et Systèmes d'Élevage, INRA Auzeville, Castanet Tolosan, France^f; Université d'Angers, Laboratoire d'Hémodynamique, Interaction Fibrose et Invasivité Tumorale Hépatique, UPRES 3859, IFR 132, Angers, France^g

Proof of concept

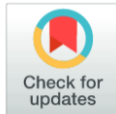
RESEARCH ARTICLE

Deciphering intra-species bacterial diversity of meat and seafood spoilage microbiota using *gyrB* amplicon sequencing: A comparative analysis with 16S rDNA V3-V4 amplicon sequencing

Simon Poirier¹, Olivier Rué², Raphaëlle Peguilhan¹, Gwendoline Coeuret¹, Monique Zagorec³, Marie-Christine Champomier-Vergès¹, Valentin Loux², Stéphane Chaillou^{1*}

1 MICALIS, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France, **2** MaIAGE, INRA, Université Paris-Saclay, Jouy-en-Josas, France, **3** Secalim, INRA, Oniris, Nantes, France

* stephane.chaillou@inra.fr



Questions :

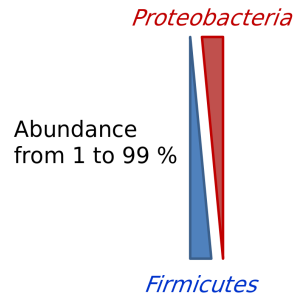
1. Reflect the real diversity ?

2. Quantification biases ?

3. Sub-species level reachable ?

Design of the study

4 Mock communities



15 food samples as 'unknown' tests

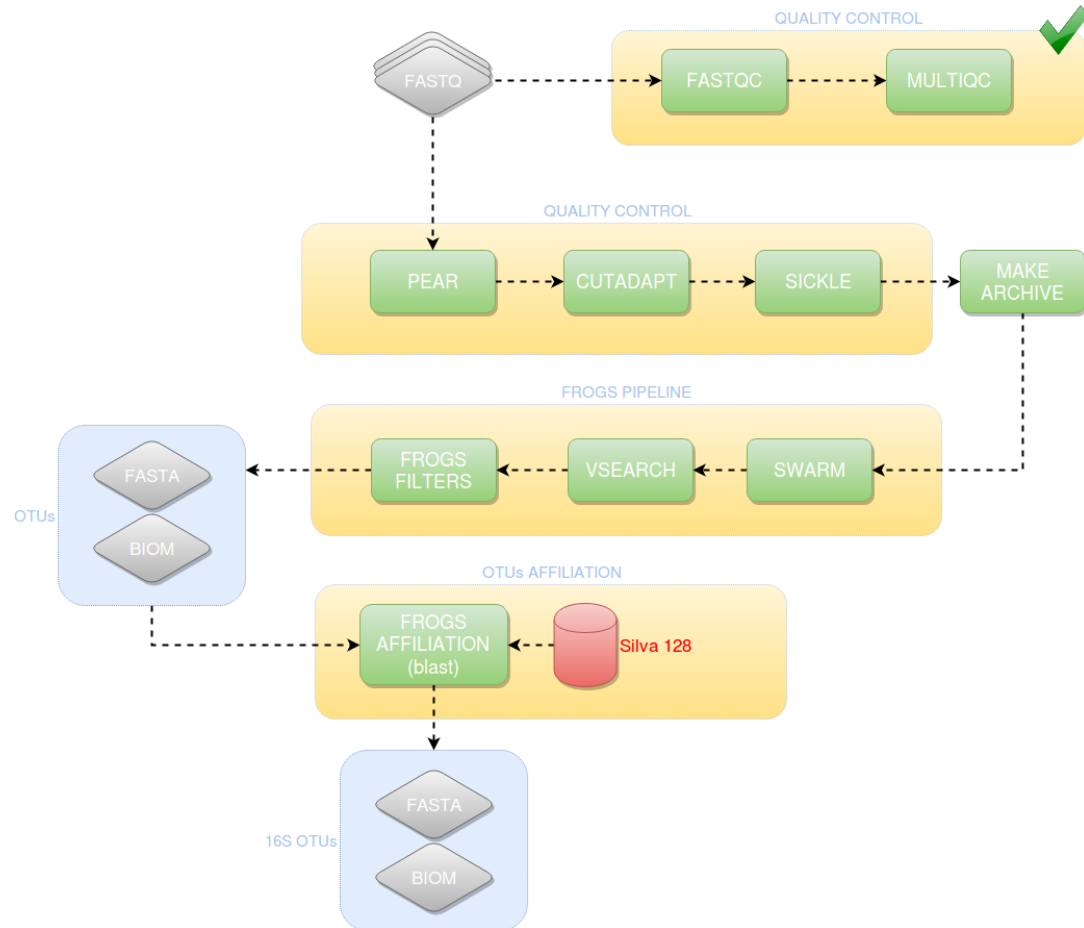


One mixture with large strain's diversity expected



Bioinformatics workflow

FROGS to obtain Operational Taxonomic Units



gyrB databank

No ready-to-use resources

- *gyrB* public databanks not complete

Use of whole sequenced genomes

- A lot of complete genomes in public repositories
- *gyrB* well annotated

 e!Ensembl


biopython

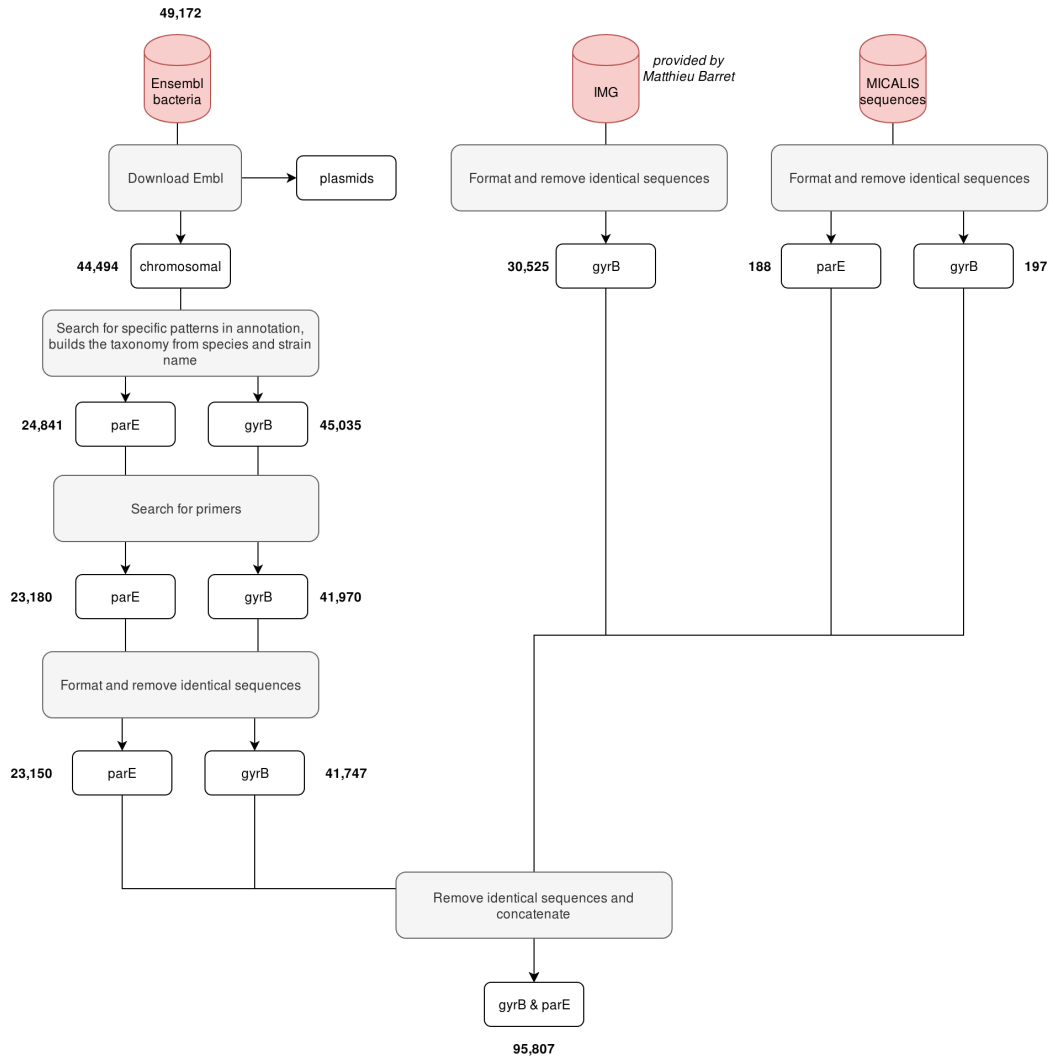
Other sequences to add

- Unpublished genomes, home-made databanks...
- Sequence, species and strain needed
 - NCBI taxonomy resources



Offer the ability to easily growth the content on demand 

gyrB databank



Results

Reflect the real diversity

	Number of observed OTUs (species richness)		Diversity of genus-level taxa (Genus richness)	
	<i>16S</i>	<i>gyrB</i>	<i>16S</i>	<i>gyrB</i>
Mock1 (15 species)	59	42	15	15
Mock2 (15 species)	77	53	16	15
Mock3 (15 species)	56	37	15	14
Mock4 (5 species)	65	50	8	5
Mock5 (complex community)	110	229	33	40

Reflect the real diversity

	Number of observed OTUs (species richness)		Diversity of genus-level taxa (Genus richness)	
	<i>16S</i>	<i>gyrB</i>	<i>16S</i>	<i>gyrB</i>
Mock1 (15 species)	59	42	15	15
Mock2 (15 species)	77	53	16	15
Mock3 (15 species)	56	37	15	14
Mock4 (5 species)	65	50	8	5
Mock5 (complex community)	110	229	33	40

Reflect the real diversity

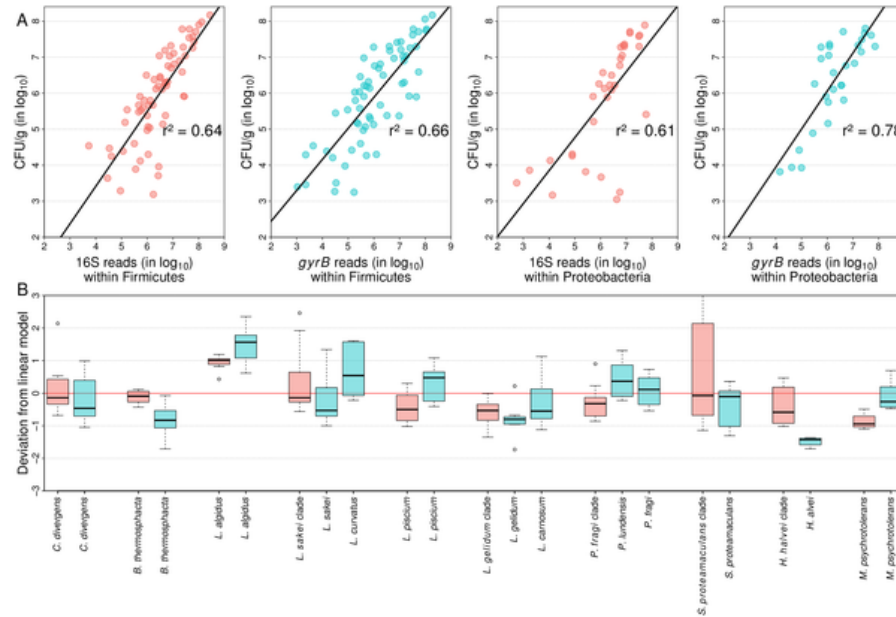
	Number of observed OTUs (species richness)		Diversity of genus-level taxa (Genus richness)	
	<i>16S</i>	<i>gyrB</i>	<i>16S</i>	<i>gyrB</i>
Mock1 (15 species)	59	42	15	15
Mock2 (15 species)	77	53	16	15
Mock3 (15 species)	56	37	15	14
Mock4 (5 species)	65	50	8	5
Mock5 (complex community)	110	229	33	40

Reflect the real diversity

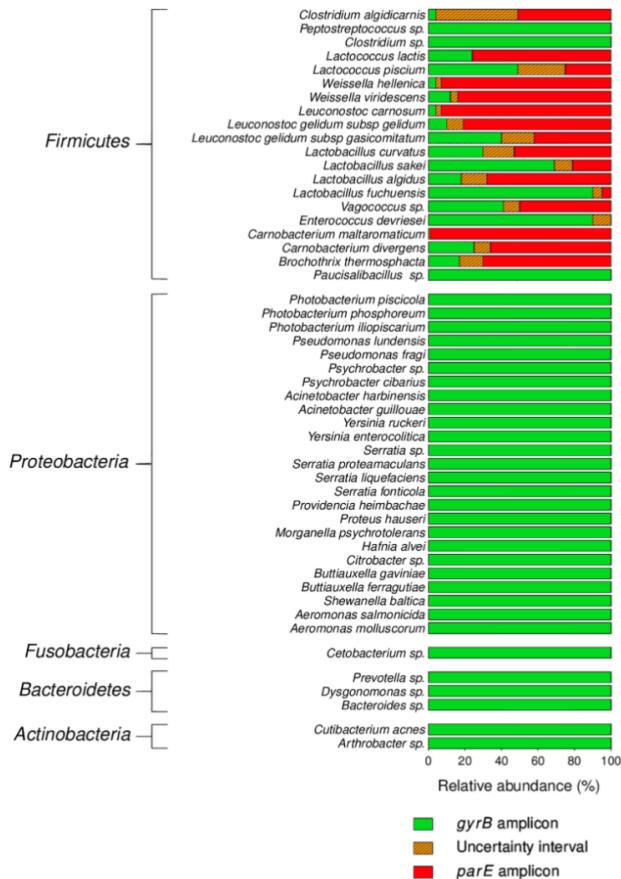
	Number of observed OTUs (species richness)		Diversity of genus-level taxa (Genus richness)	
	<i>16S</i>	<i>gyrB</i>	<i>16S</i>	<i>gyrB</i>
Mock1 (15 species)	59	42	15	15
Mock2 (15 species)	77	53	16	15
Mock3 (15 species)	56	37	15	14
Mock4 (5 species)	65	50	8	5
Mock5 (complex community)	110	229	33	40

Quantification biases

- quantitative PCR

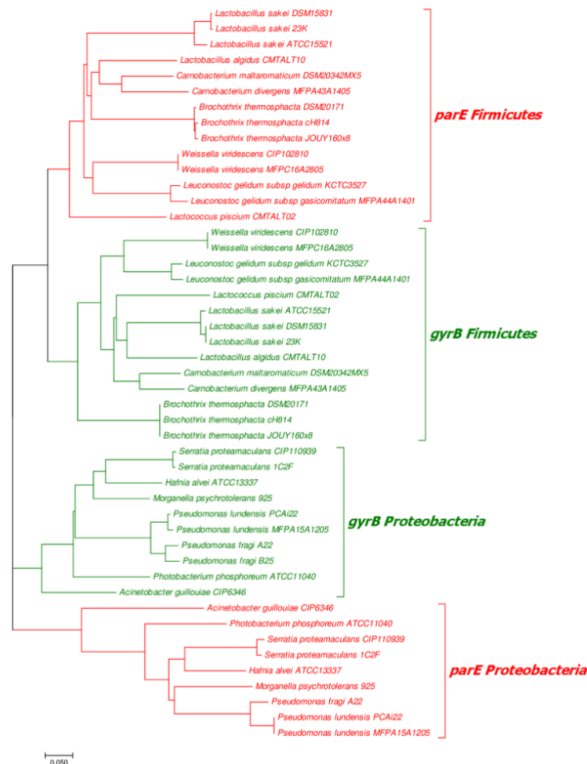


Amplification biases



- Amplification of *parE* (*gyrB* paralog)
- No impact on quantification

Amplification biases



- Amplification of *parE* (*gyrB* paralog)
- No impact on quantification
- **Specific to Firmicutes**

Fig 3. Unrooted phylogenetic tree constructed with the partial *gyrB* and *parE* sequences extracted from the genomes of bacterial species present in the mock communities. Branch length scale indicates the number of nucleotide substitutions per site. The tree is divided into three main branches: *parE* genes from *Proteobacteria*, *gyrB* genes from *Proteobacteria*, and *parE/gyrB* genes from *Firmicutes*. The two sub-branches discriminating between *parE* and *gyrB* genes from *Firmicutes* are equally distant from the *gyrB* branch of *Proteobacteria*.

<https://doi.org/10.1371/journal.pone.0204629.g003>

Sub-species level

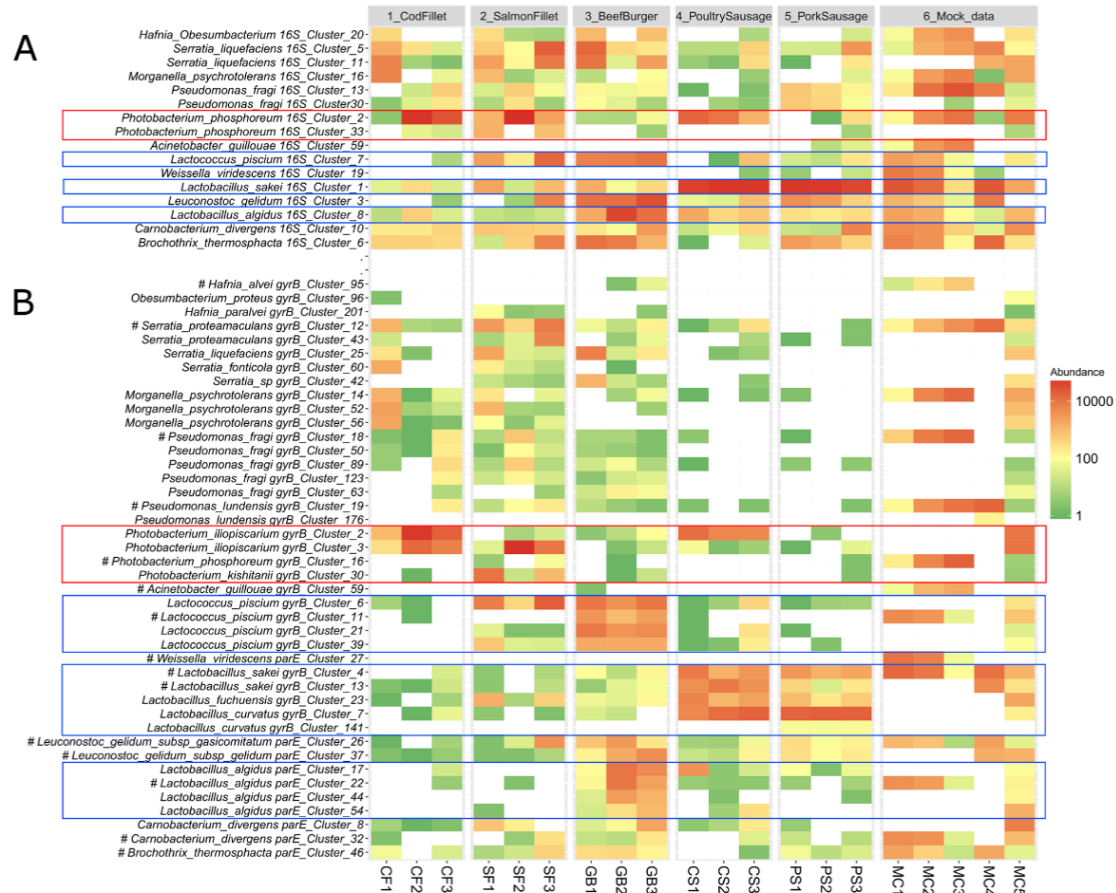
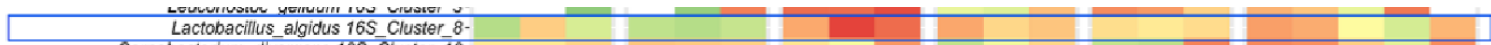


Fig 7. Heatmap showing the ability of (A) 16S rDNA and (B) *gyrB* amplicon analysis to estimate intraspecies population levels. Samples are ordered from left to right according to the sample type. The scale on the right of the heatmap depicts the color palette associated with the relative numbers of reads of the various OTUs. OTUs are labeled with their cluster number and the taxonomic assignment at the species level. The *gyrB*/*parE* OTUs associated with the strains used in the mock communities are labeled with (#). Boxes are drawn around the main phylogenetic clades that are detailed in the text.

Sub-species level ✓

1_CodFillet 2_SalmonFillet 3_BeefBurger 4_PoultrySausage 5_PorkSausage 6_Mock_data

16S



gyrB



Conclusions / Take home messages

- At the genus level : 16S (V3-V4) and *gyrB* performs identically 👍
- Taxonomic assignment performed 100% at the species level with *gyrB* 👍
- Tracking and tracing of strain's diversity was possible 👍
- Easy use of complete genomes for building a specific *gyrB* databank 👍

Guidelines

- For not well known food or other microbiota: **always include a 16S rDNA control !**
- Validated for Firmicutes and Proteobacteria.
 - Needs validation for other phyla.

Thanks



- Simon Poirier (Post-Doc MICALIS, publication author)
- INRA units and teams



- REDLOSSES consortium



- JOBIM programme comitee

