Molecular methods to characterize the microbiota in the mouse tissues

Olivier Bouchez, GeT-PlaGe, INRA Toulouse
http://get.genotoul.fr
@GeT_Genotoul

Who are we?

- **Genomic and transcriptomic core facility spreads on**
  - **5 sites GeT in Toulouse**
  - **1 site Oncopole**

- **National Infrastructure within the « France Génomique » program**
- **IBISA Label**
- **INRA strategic core facility**
- **ISO9001:2008 & NFX50-900 Certifications**

- **Partnership with Genotoul Bioinformatic core-facility**

GeT Génome et Transcriptome

The GeT Core Facility

Our missions:

- ✓ To provide innovating technologies for genome analysis to the scientific community
- ✓ To develop new protocols, new methodologies, acquire expertise and train to those technologies
- ✓ To animate workshops for user network

Technologies available to the scientific community :

- ✓ Quantitative PCR: Fluidigm BioMark, Life tech QS6, Viia7…
- ✓ Genotyping: Fluidigm BioMark, Agilent, Affymetrix…
- ✓ Microarrays (expression): Agilent, Affymetrix
- ✓ Automation: TECAN EVOs, Agilent Bravo, Fluidigm Access Array
- ✓ Sanger Sequencing: ABI3130XL, ABI3730
- ✓ Single Cell Analysis: Fluidigm C1
- ✓ Next Generation Sequencing: Life tech Ion Proton, PGM, Illumina MiSeq, HiSeq3000, PacBio RSII

GET
Génome et
Transcriptome

What is Metagenomics?

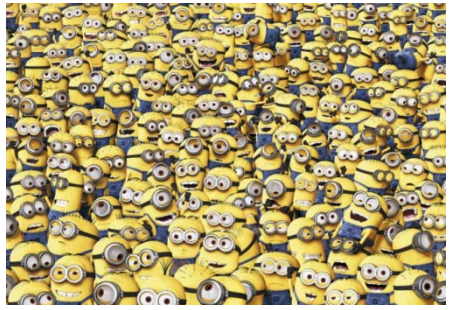
Metagenomics

Metagenomics is the study of **genetic material** recovered directly from biological samples.

Traditionnal microbiology : clonal cultures, 16S gene cloning and sequencing, ... => the vast majority of microbial diversity has been missed by cultivation-based methods

Recent studies : shotgun or PCR directed sequencing to get largely unbiased samples of all genes from all members

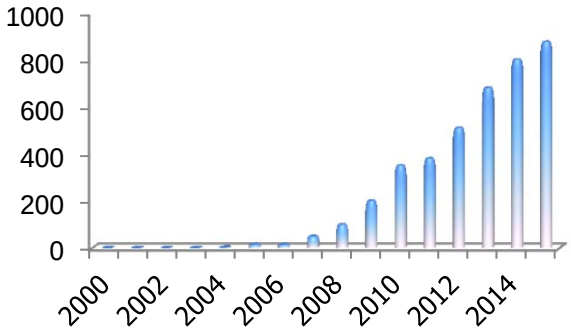
Because of its ability to **reveal the previously hidden diversity** of microscopic life, metagenomics offers a **powerful lens** for viewing the microbial world.



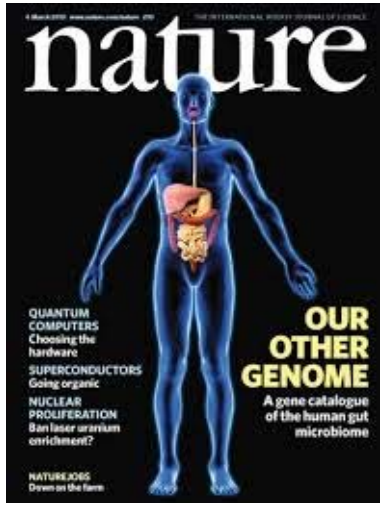
https://en.wikipedia.org/wiki/Metagenomics

Metagenomics

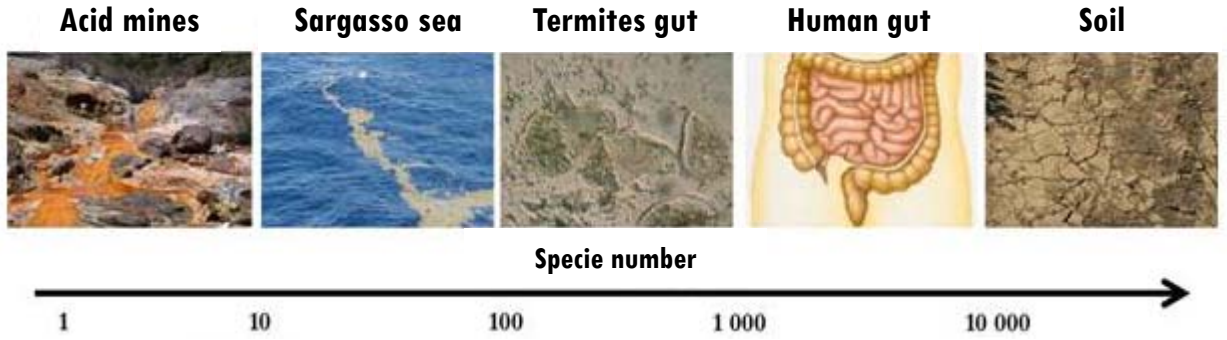
✓ Human Gut : 100 billions bacteria/1000 species (10 x more than human cell number)



✓ Growing publication number



✓ Soil: several thousand of different species/g of soil



Metagenomics

Bacterial communities are involved in:

- ✓ Diabetes
- ✓ Cancers
- ✓ Autism
- ✓ Inflammatory responses
- ✓ Digestion
- ✓ Site remediation
- ✓ Fermentation
- ✓ Methanisation
- ✓ Plant production
- ✓ ….

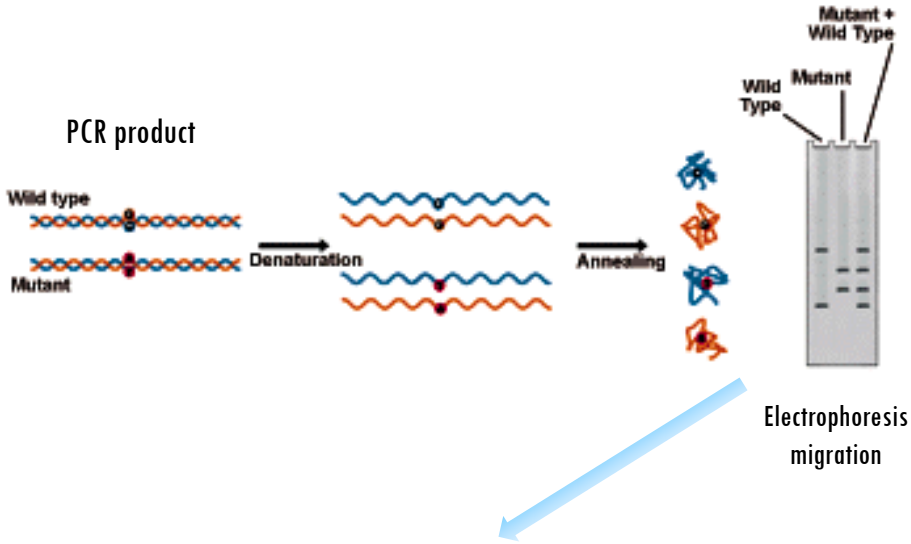
genotoul

GET
Génome et
Transcriptome

Examples of previous technologies used in microbiology Studies

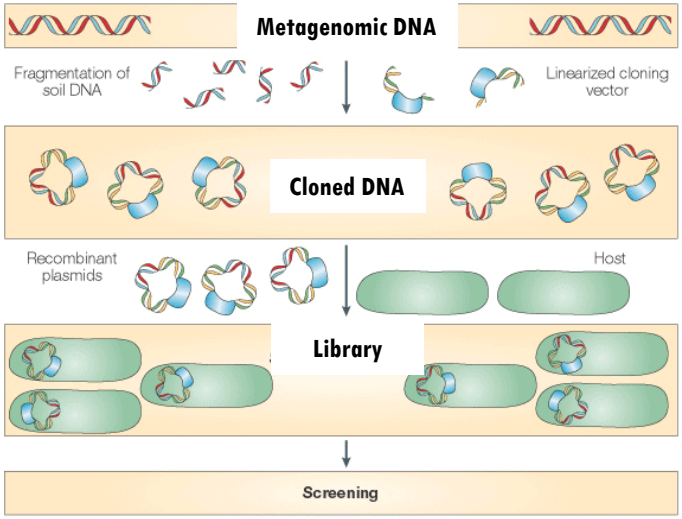
SSCP & Cloning

SSCP = single strand conformation polymorphism



Genetic fingerprinting : allows the observation of differences between communities, but not assignation

Cloning



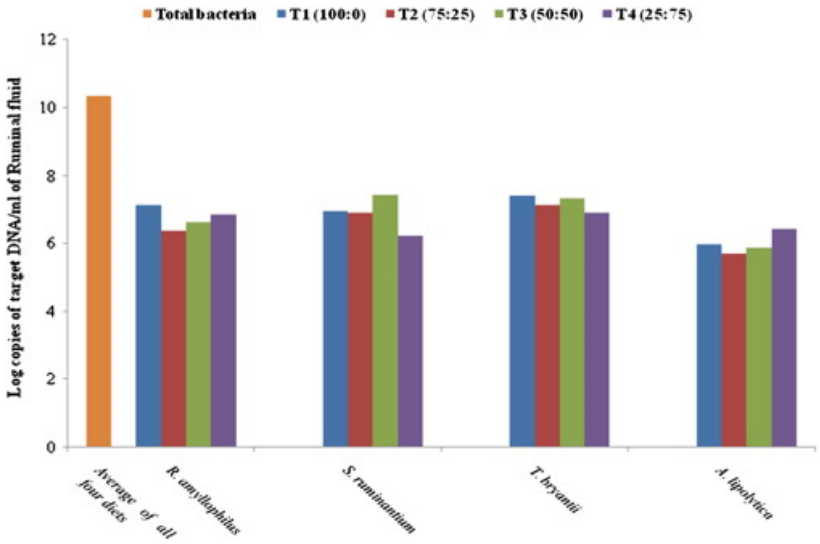
Nature Reviews | Microbiology

Study clone by clone : limited information, very time consuming and expensive

Modified from Rolf Daniel, Nature Reviews Microbiology, 2005

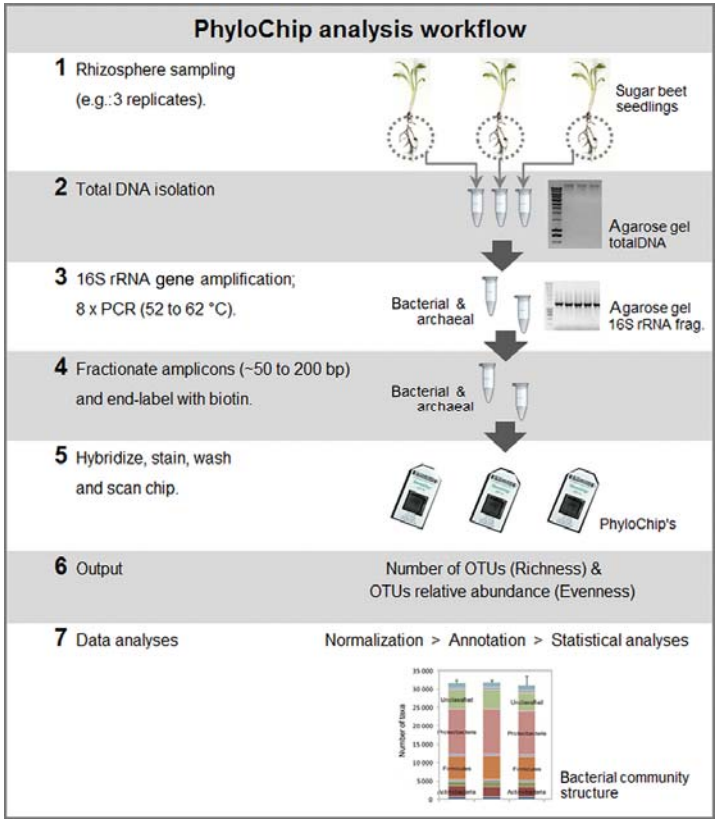
qPCR & PhyloChip

qPCR



Quantification of selected species
Very sensitive

Singh *et al.*, Meta Gene, 2014



Studies only the species represented on the Chip (~8000, 16S)
Quantification not so easy because of fluorescence saturation

Mendes *et al.*, FEMS Microbiology Reviews, 2013

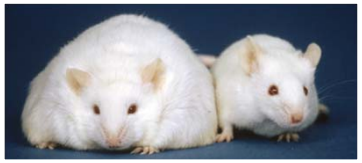
GET
Génome et
Transcriptome

Next Generation Metagenomics

How to do Next Generation Metagenomics?

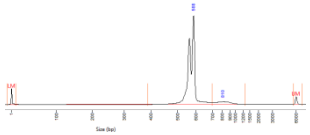
Next Generation Sequencing: sequencing of several millions to billions of short DNA fragments in parallel

Biological samples



DNA/RNA extraction

DNA/cDNA Targeted Metagenomics



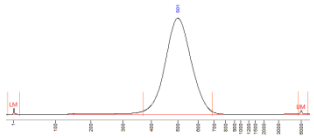
PCR amplication of the target

Paired-end sequencing

Identification of Microbial communities

Shotgun Metagenomics

Library preparation



Paired-end sequencing

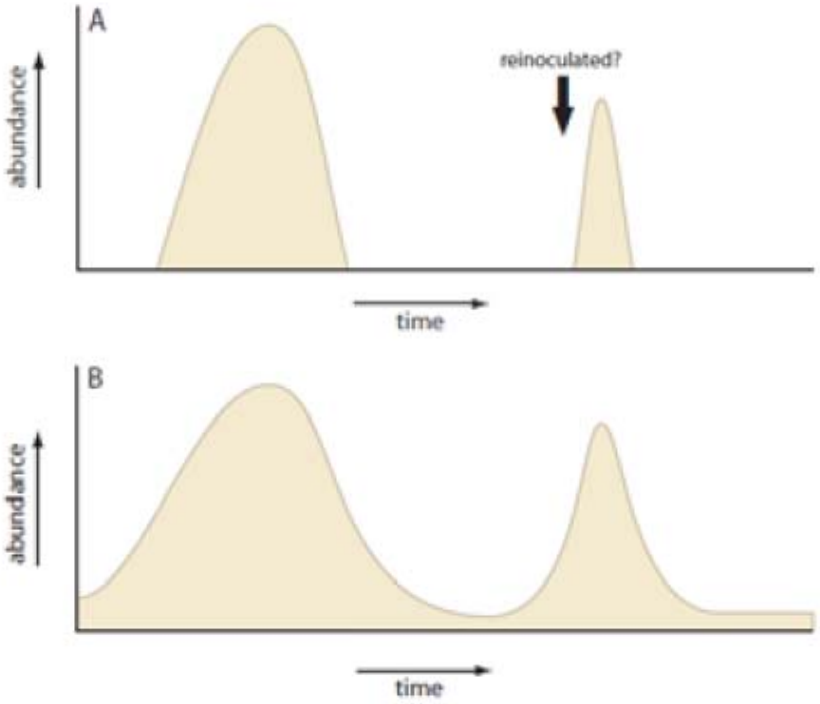
Assembly and interpretation

How to choose the approach I need?

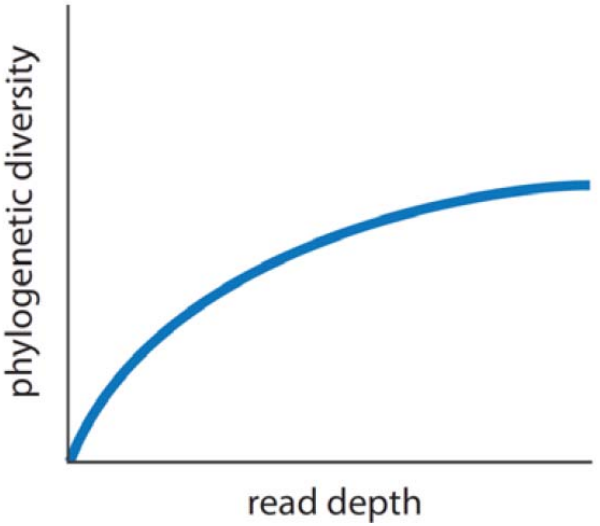
	Targeted sequencing	Shotgun Sequencing
Informations	Taxonomic composition and phylogenetic structure of microbial communities (OTU)	Fonctionnal characterisation of bacterial communities, draft genomes construction
Applications	Population identification	New/unknown member detection, new gene identification
Sensitivity	Extremely sensitive	Needs much more sequencing depth to obtain the same sensitivity as targeted sequencing (=€€€€€)
Biases	PCR induced biases: the amplification efficiency could be different between communities	Sequence composition (GC %): sequencing in GC rich regions is not optimal

Sequencing depth impact on result interpretation

Sequencing depth = number of sequences per sample



Enough sequencing depth?



Identification of under-represented sequences

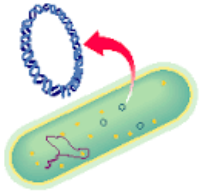
genotoul

GET
Génome et
Transcriptome

Targeted Metagenomics

Targeted Metagenomics known biases

- ✓ **DNA extraction kits/protocols**



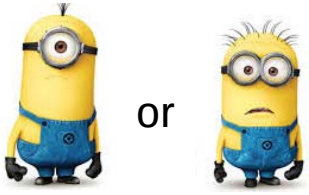
RESEARCH ARTICLE

Comparative Evaluation of DNA Extraction Methods from Feces of Multiple Host Species for Downstream Next-Generation Sequencing

Marcia L. Hart[1], Alexandra Meyer[2], Philip J. Johnson[3], Aaron C. Ericsson[1,4,5]*

- ✓ **PCR**
  - **Polymerase efficiency**
  - **Polymerase contaminations**
  - **Use of degenerated primers: non homogenous amplification**
- ✓ **Databases exhaustivity**

- ✓ **Analysis softwares**
- ✓ **16S copy numbers (1 to 15, depending on the bacteria species)**
- ✓ **Horizontal gene transferts (Ex : *B. globisporus* and *B. psychrophilus* show 99,8% identity on 16S, but only 23 to 50 % identity at genome level)**



Targeted Metagenomics known biases

D'Amore *et al. BMC Genomics* (2016) 17:55
DOI 10.1186/s12864-015-2194-9

BMC Genomics

RESEARCH ARTICLE Open Access

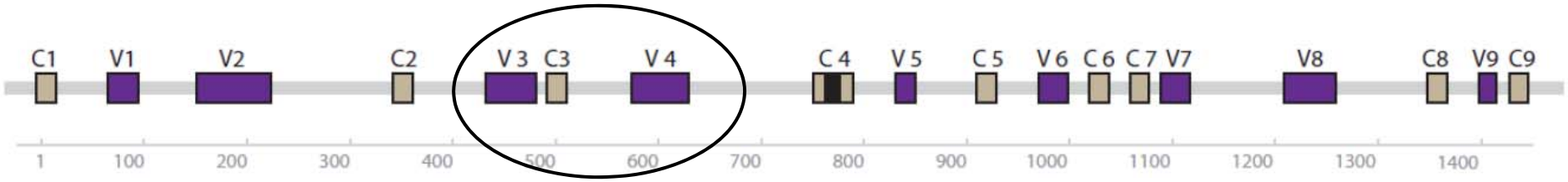
A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling

Rosalinda D'Amore[1†], Umer Zeeshan Ijaz[2†], Melanie Schirmer[2], John G. Kenny[1], Richard Gregory[1], Alistair C. Darby[1], Migun Shakya[3], Mircea Podar[4], Christopher Quince[5*] and Neil Hall[1*]

Impact of the sequencer and protocols on the results:

- ✓ **The choice of sequencing platform and experimental design needs to be taken into consideration in the early stage of a project**
- ✓ **A pilot experiment should be necessary to choose the target (ie which 16S region is the best for my study to identify associated microbial communities)**

Illumina MiSeq 16S sequencing example



Which region of the 16S to target?

- ✓ **Today, the v3-v4 is to most used**
- ✓ **400 to 500 pb PCR product**

Illumina MiSeq

Sequencing by synthesis

Max lenght: 2 x 300 pb

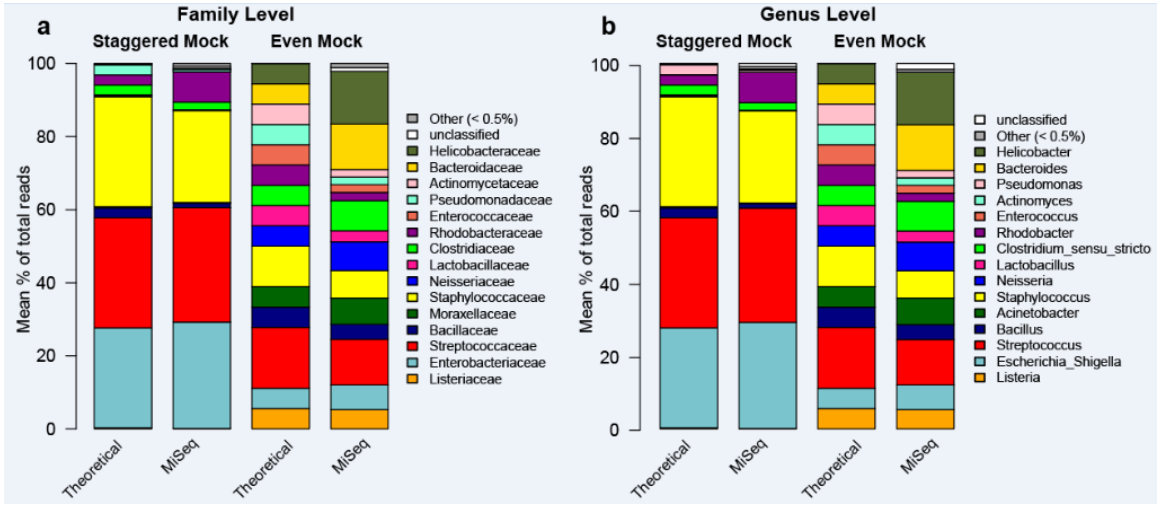
Max Throughput: 15 GB

Run time: 55 h

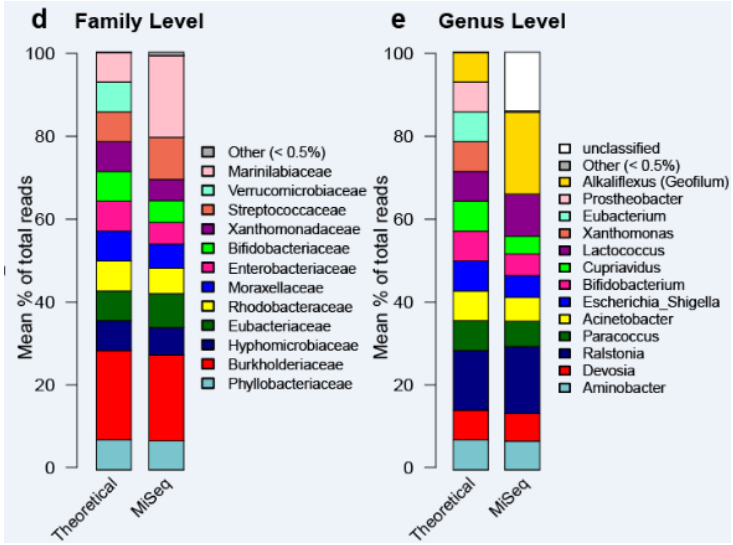
Pipeline validation against mock communities

Results obtained in collaboration with the VAIOMER start-up

Lluch *et al.* 2015, PloS-ONE



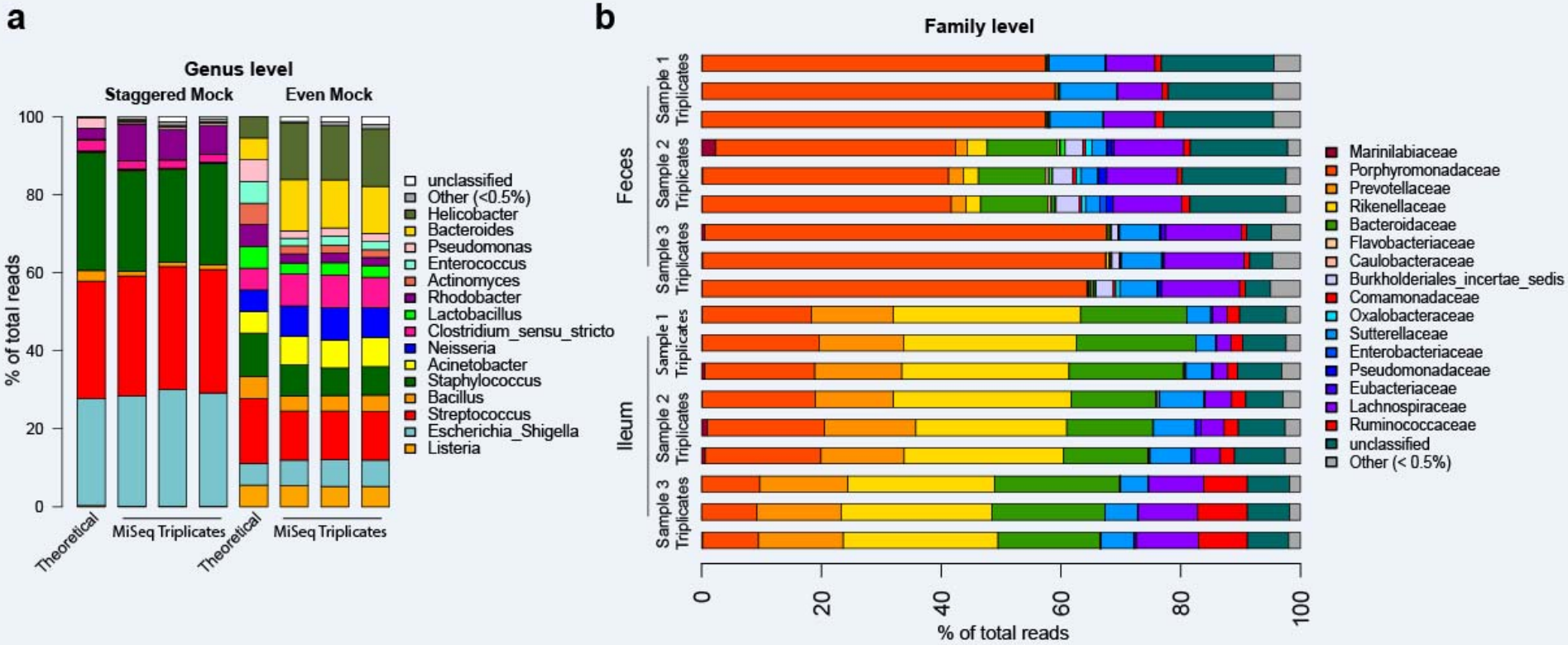
gDNA mock communities



Plasmid mock communities

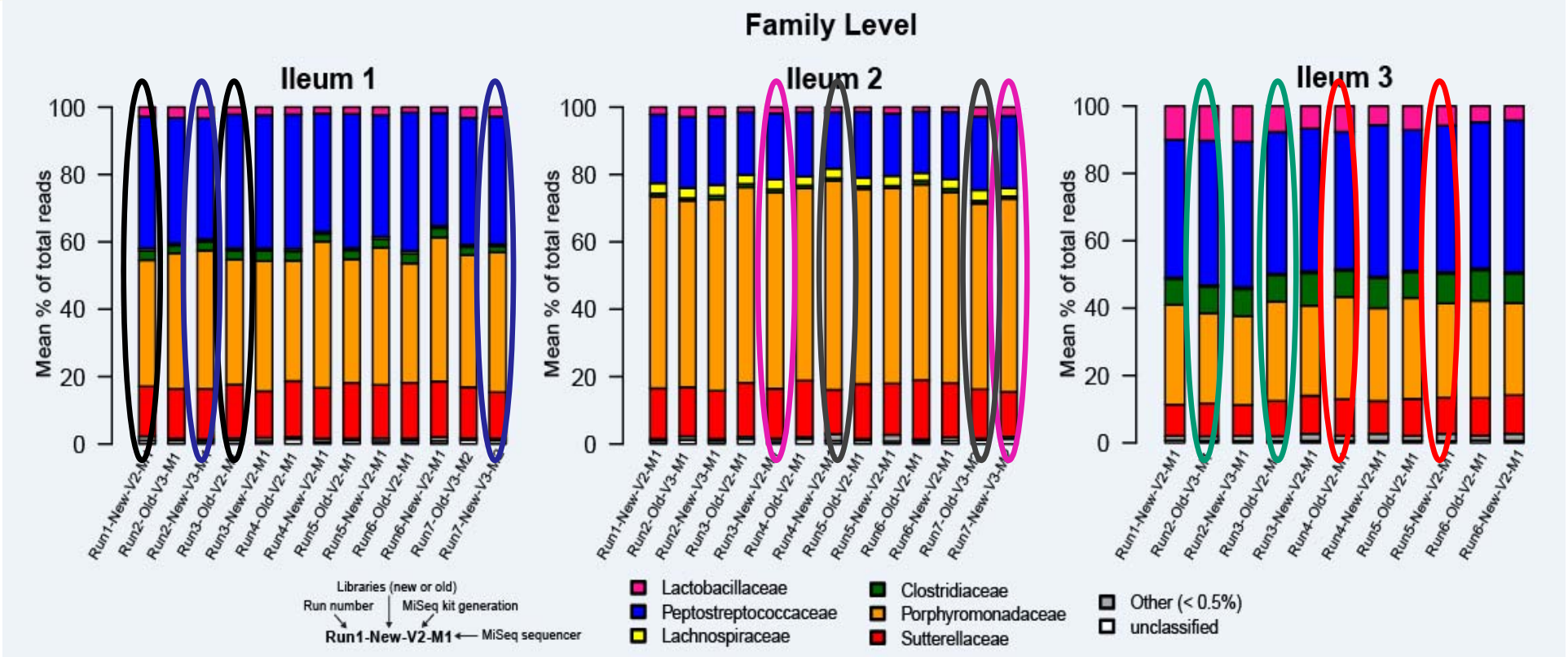
Good correlation between theory and reality

Replicability and reproductibility



Low variation between replicates

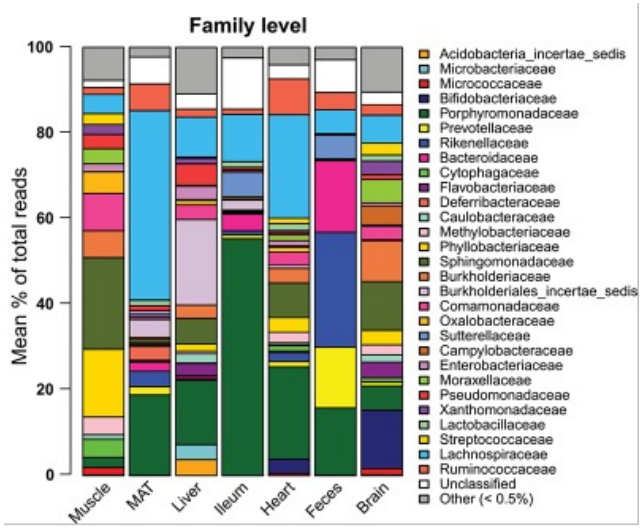
Replicability and reproducibility

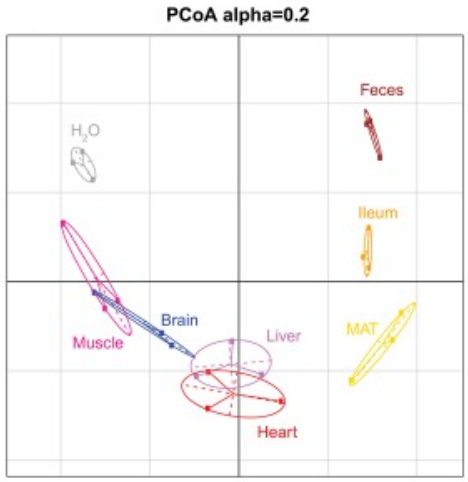


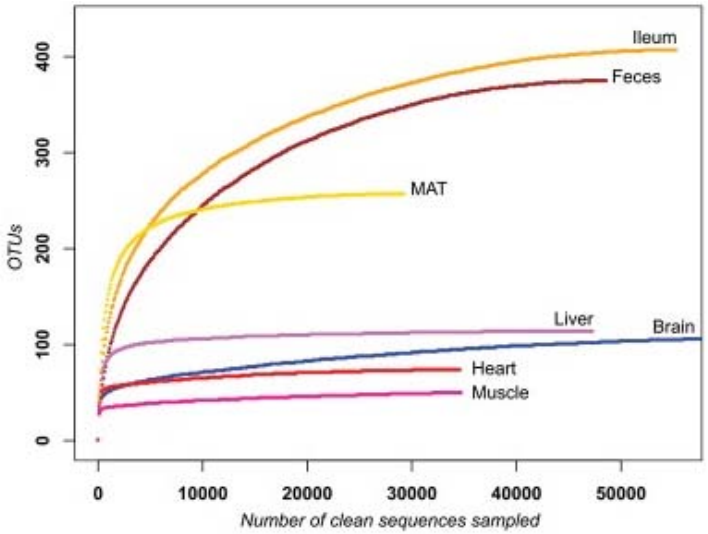
Same library, same chemistry, same sequencer
Same library, different chemistry, same sequencer
Different library, same chemistry, same sequencer
Same library, same chemistry, different sequencer
Same library, different chemistry, different sequencer
Different library, different chemistry, different sequencer

Very low variability, good reproducibility

16S metagenomics on diverse tissue samples







Detection of bacteria in diverse mouse tissues

Taxonomic profiles different between tissues and different from H_2O control

Very high diversity in Ileum, Feces and adipose tissues (MAT)

GeT Génome et Transcriptome

...nalyses

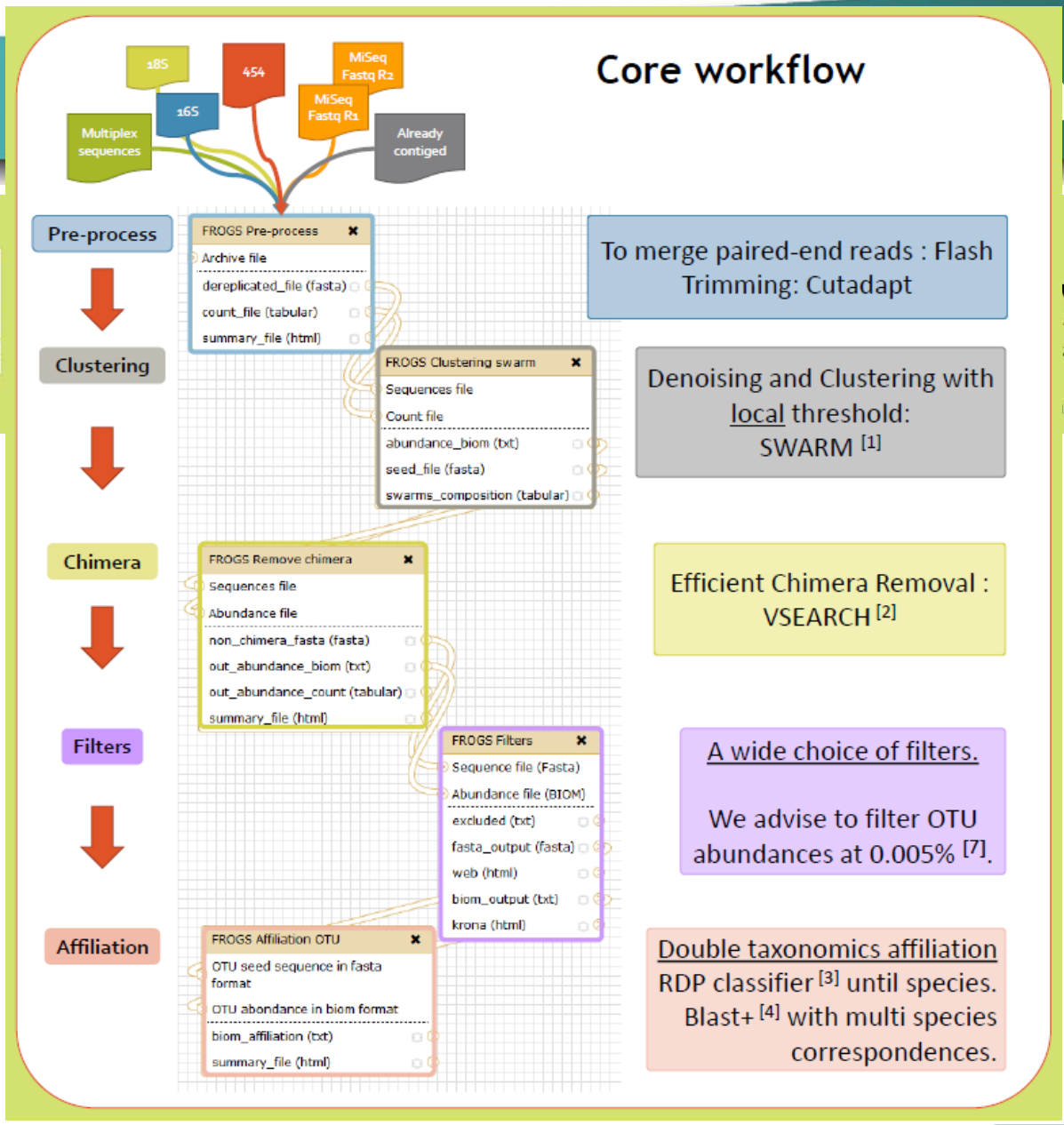
...y[5] Solution

...a MARIADASSOU, Guillermina

...ldine.pascal@toulouse.inra.fr

INRA SCIENCE & IMPACT

OTU : Operationnal



genotoul bioinfo

Formations available!

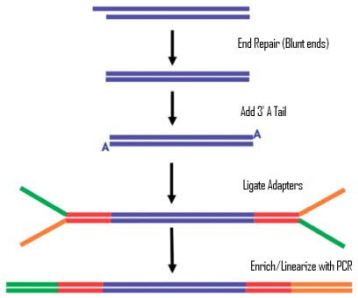
frogs@toulouse.inra.fr

genotoul

GeT
Génome et
Transcriptome

Shotgun Metagenomics

Principle



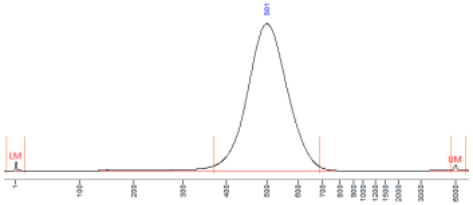
DNA/RNA

Shearing (sonication, nebulization, enzymatic, mecanic)

Or reverse transcription

Sheared DNA or cDNA

End repair, A-tailing, adaptaterligation, optional PCR



Sequencing ready library

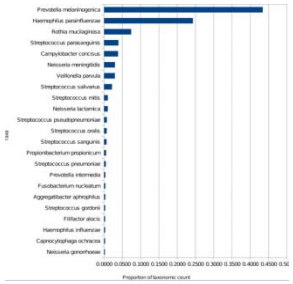
Sequencing

Illumina HiSeq3000

2,4 billion paired-end reads/run

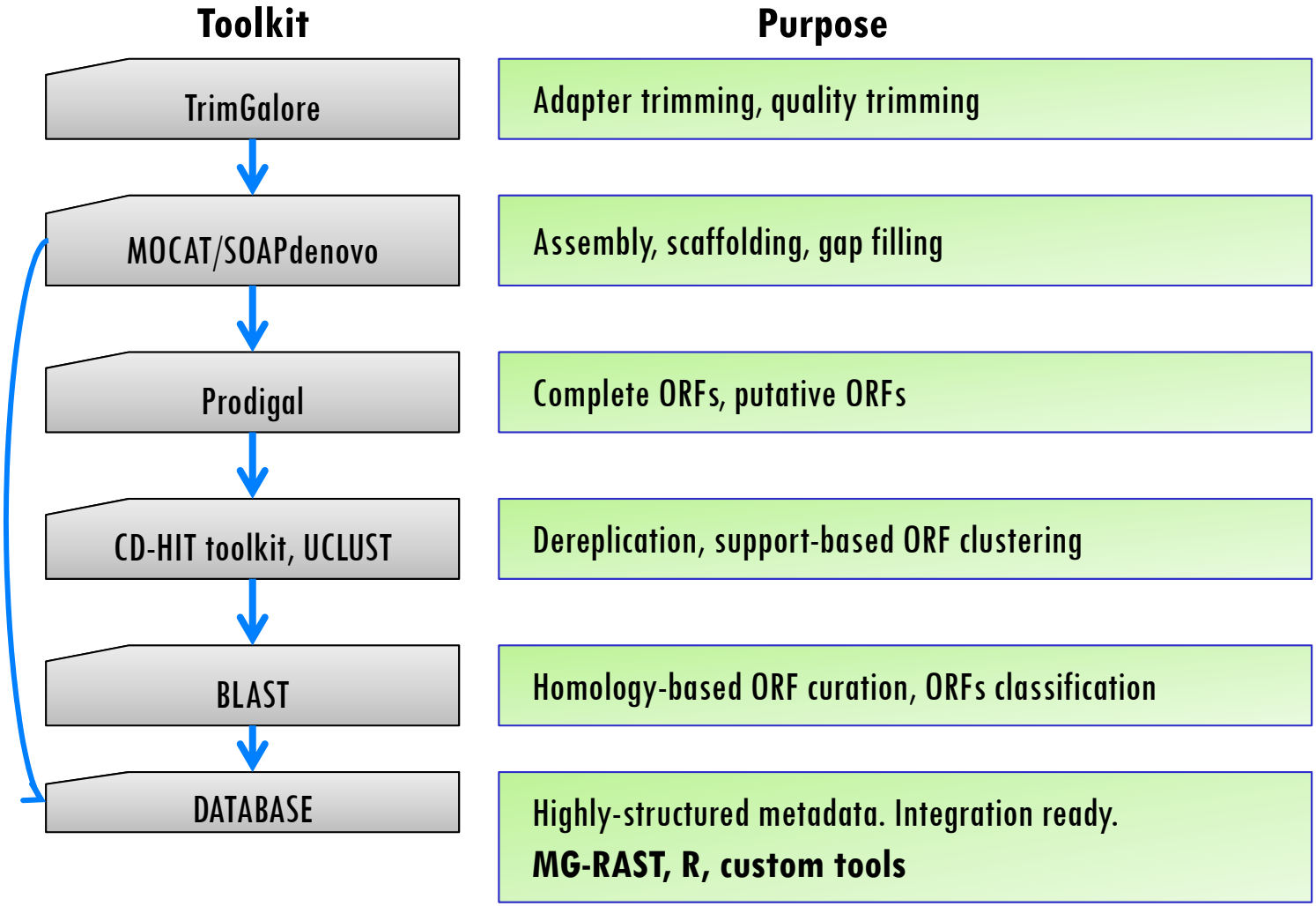
Sequence data

Bioinformatic analyses



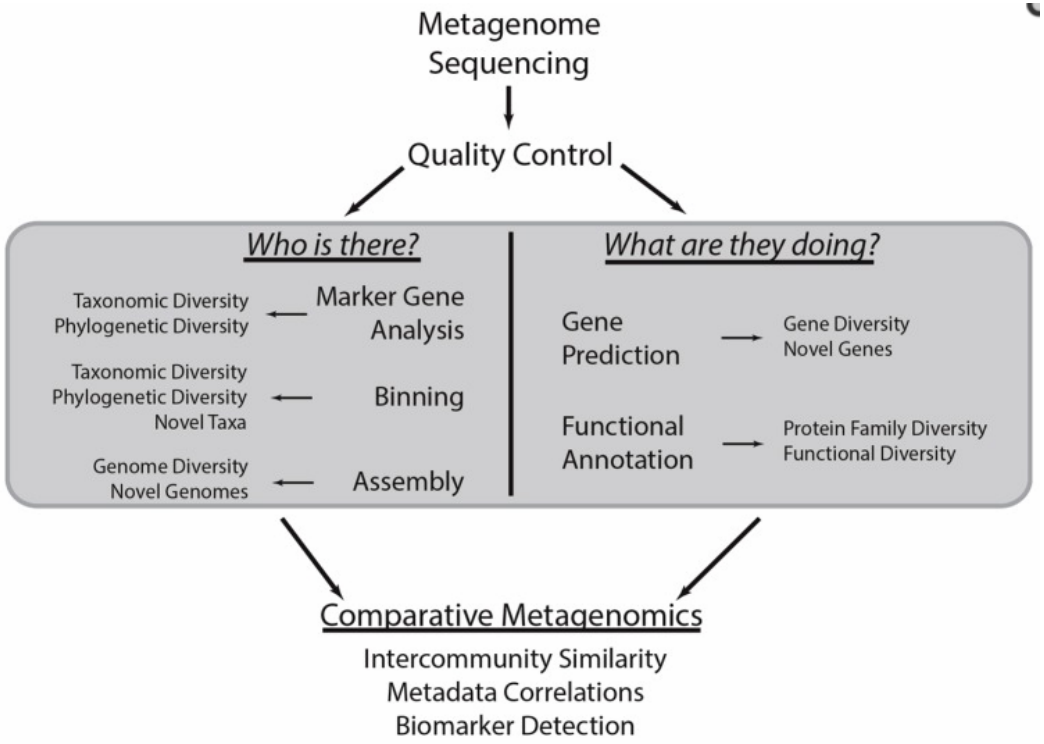
Interpretation

Example of an informatic pipeline



Pavel Senin

Data interpretation



frontiers in PLANT SCIENCE
REVIEW ARTICLE
published: 16 June 2014
doi: 10.3389/fpls.2014.00209

An introduction to the analysis of shotgun metagenomic data

*Thomas J. Sharpton**
Department of Microbiology and Department of Statistics, Oregon State University, Corvallis, OR, USA

First large study published by the MetaHIT consortium

Vol 464|4 March 2010|doi:10.1038/nature08821 nature

ARTICLES

A human gut microbial gene catalogue established by metagenomic sequencing

Junjie Qin[1]*, Ruiqiang Li[1]*, Jeroen Raes[2,3], Manimozhiyan Arumugam[2], Kristoffer Solvsten Burgdorf[4], Chaysavanh Manichanh[5], Trine Nielsen[4], Nicolas Pons[6], Florence Levenez[6], Takuji Yamada[2], Daniel R. Mende[2], Junhua Li[1,7], Junming Xu[1], Shaochuan Li[1], Dongfang Li[1,8], Jianjun Cao[1], Bo Wang[1], Huiqing Liang[1], Huisong Zheng[1], Yinlong Xie[1,7], Julien Tap[6], Patricia Lepage[6], Marcelo Bertalan[9], Jean-Michel Batto[6], Torben Hansen[4], Denis Le Paslier[10], Allan Linneberg[11], H. Bjørn Nielsen[9], Eric Pelletier[10], Pierre Renault[6], Thomas Sicheritz-Ponten[9], Keith Turner[12], Hongmei Zhu[1], Chang Yu[1], Shengting Li[1], Min Jian[1], Yan Zhou[1], Yingrui Li[1], Xiuqing Zhang[1], Songgang Li[1], Nan Qin[1], Huanming Yang[1], Jian Wang[1], Søren Brunak[9], Joel Doré[6], Francisco Guarner[5], Karsten Kristiansen[13], Oluf Pedersen[4,14], Julian Parkhill[12], Jean Weissenbach[10], MetaHIT Consortium†, Peer Bork[2], S. Dusko Ehrlich[6] & Jun Wang[1,13]

To understand the impact of gut microbes on human health and well-being it is crucial to assess their genetic potential. Here we describe the Illumina-based metagenomic sequencing, assembly and characterization of 3.3 million non-redundant microbial genes, derived from 576.7 gigabases of sequence from faecal samples of 124 European individuals. The gene set, ~150 times larger than the human gene complement, contains an overwhelming majority of the prevalent (more frequent) microbial genes of the cohort and probably includes a large proportion of the prevalent human intestinal microbial genes. The genes are largely shared among individuals of the cohort. Over 99% of the genes are bacterial, indicating that the entire cohort harbours between 1,000 and 1,150 prevalent bacterial species and each individual at least 160 such species, which are also largely shared. We define and describe the minimal gut metagenome and the minimal gut bacterial genome in terms of functions present in all individuals and most bacteria, respectively.

Mouse gut metagenome

RESOURCE

nature
biotechnology

A catalog of the mouse gut metagenome

Liang Xiao[1,16], Qiang Feng[1,2,16], Suisha Liang[1,16], Si Brask Sonne[2], Zhongkui Xia[1], Xinmin Qiu[1], Xiaoping Li[1], Hua Long[3], Jianfeng Zhang[1], Dongya Zhang[1], Chuan Liu[1], Zhiwei Fang[1], Joyce Chou[3], Jacob Glanville[3], Qin Hao[2], Dorota Kotowska[2], Camilla Colding[2], Tine Rask Licht[4], Donghai Wu[5], Jun Yu[6], Joseph Jao Yiu Sung[6], Qiaoyi Liang[6], Junhua Li[1], Huijue Jia[1], Zhou Lan[1], Valentina Tremaroli[7], Piotr Dworzynski[8], H Bjørn Nielsen[8], Fredrik Bäckhed[7,9], Joël Doré[10,11], Emmanuelle Le Chatelier[11], S Dusko Ehrlich[11,12], John C Lin[3], Manimozhiyan Arumugam[1,9], Jun Wang[1,2,13,14], Lise Madsen[1,2,15] & Karsten Kristiansen[1,2]

We established a catalog of the mouse gut metagenome comprising ~2.6 million nonredundant genes by sequencing DNA from fecal samples of 184 mice. To secure high microbiome diversity, we used mouse strains of diverse genetic backgrounds, from different providers, kept in different housing laboratories and fed either a low-fat or high-fat diet. Similar to the human gut microbiome, >99% of the cataloged genes are bacterial. We identified 541 metagenomic species and defined a core set of 26 metagenomic species found in 95% of the mice. The mouse gut microbiome is functionally similar to its human counterpart, with 95.2% of its Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologous groups in common. However, only 4.0% of the mouse gut microbial genes were shared (95% identity, 90% coverage) with those of the human gut microbiome. This catalog provides a useful reference for future studies.

Conclusions

Targeted Metagenomics:

- ✓ Known biases, need to be take into account during analysis
- ✓ Pilot experiment necessary to calibrate the study
- ✓ Very sensitive
- ✓ Inexpensive
- ✓ Easy preparation and sequencing
- ✓ Bioinformatic: dedicated tools, fast, easy to manipulate

Shotgun Metagenomics:

- ✓ Less biases
- ✓ More informations provided from this application
- ✓ Less sensitive
- ✓ Expensive
- ✓ Preparation and sequencing easy
- ✓ But bioinformatic is the bottleneck, slow, need to be expert to manipulate the data

3rd generation sequencers: the future?

Oxford NANOPORE Technologies

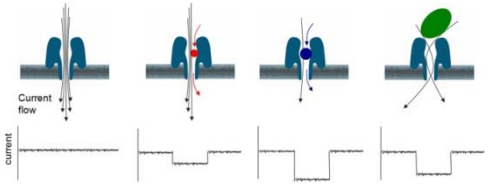
PACBIO®

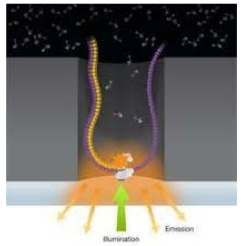
MinION

PromethION

Sequel

RSII





Longer reads (>10 kb) = better taxonomic resolution

genotoul

GeT
Génome et
Transcriptome

Thanks!

LANGUEDOC-ROUSSILLON
LA RÉGION MIDI-PYRÉNÉES

IBiSA
INFRASTRUCTURES BIOLOGIE SANTÉ ET AGRONOMIE

vaiomer

INRA
SCIENCE & IMPACT

FRANCE GÉNOMIQUE

LLOYD'S REGISTER QUALITY ASSURANCE
ISO9001