REGIONE AUTONOMA FRIULI VENEZIA GIULIA

ISTITUTO DI RICOVERO E CURA
A CARATTERE SCIENTIFICO
Burlo Garofolo di Trieste
BURLO

E.C.M.

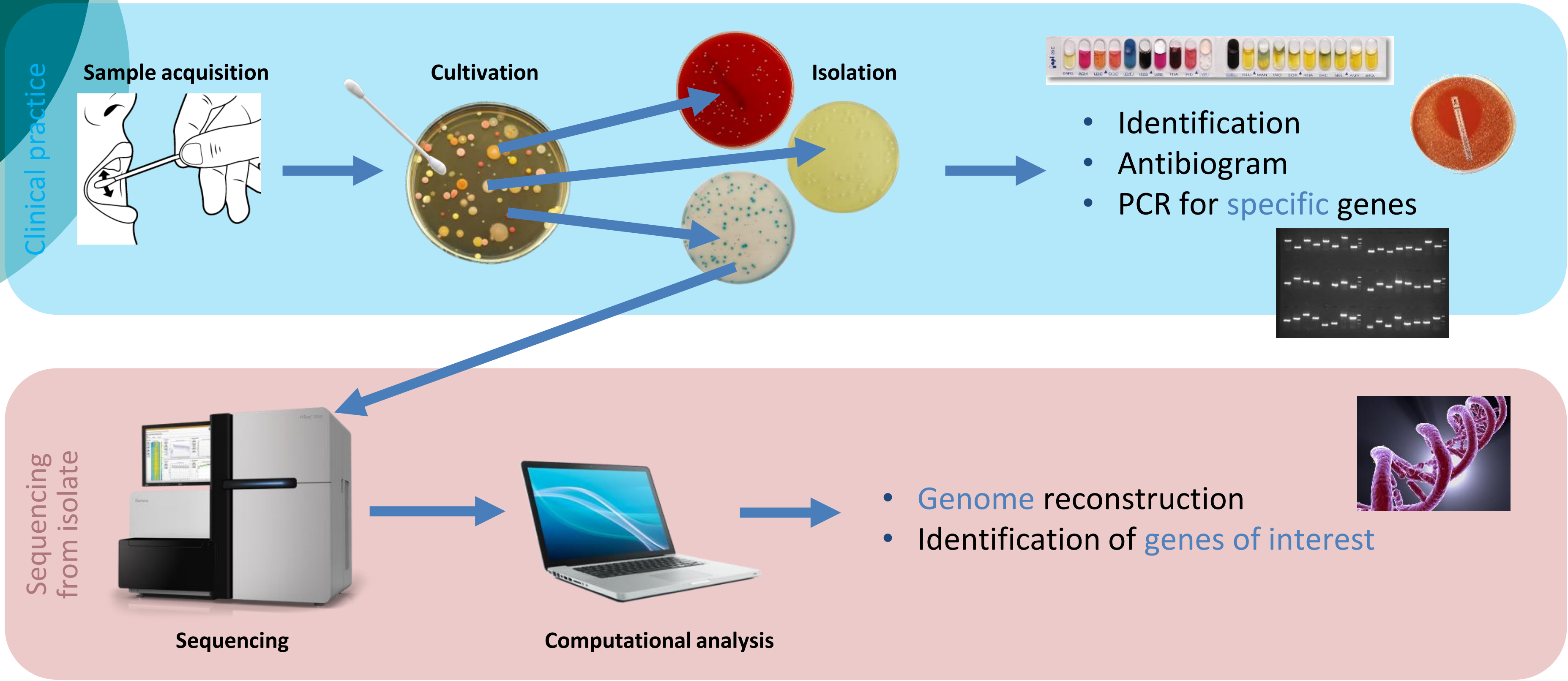
IRCBG_20081
“Metagenomica e culturomica: aspetti complementari della microbiologia omica"

Metagenome-assembled genomes (MAGs) to uncover the hidden human microbiome diversity

Trieste, 11 dicembre 2020

Docente: Edoardo Pasolli
edoardo.pasolli@unina.it

Introduction to metagenomics



Clinical practice

Sample acquisition

Cultivation

Isolation

- Identification
- Antibiogram
- PCR for specific genes

Sequencing from isolate

Sequencing

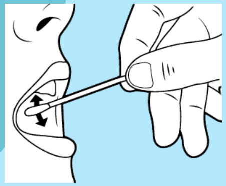
Computational analysis

- Genome reconstruction
- Identification of genes of interest

Introduction to metagenomics

Clinical practice

Sample acquisition



Metagenomics

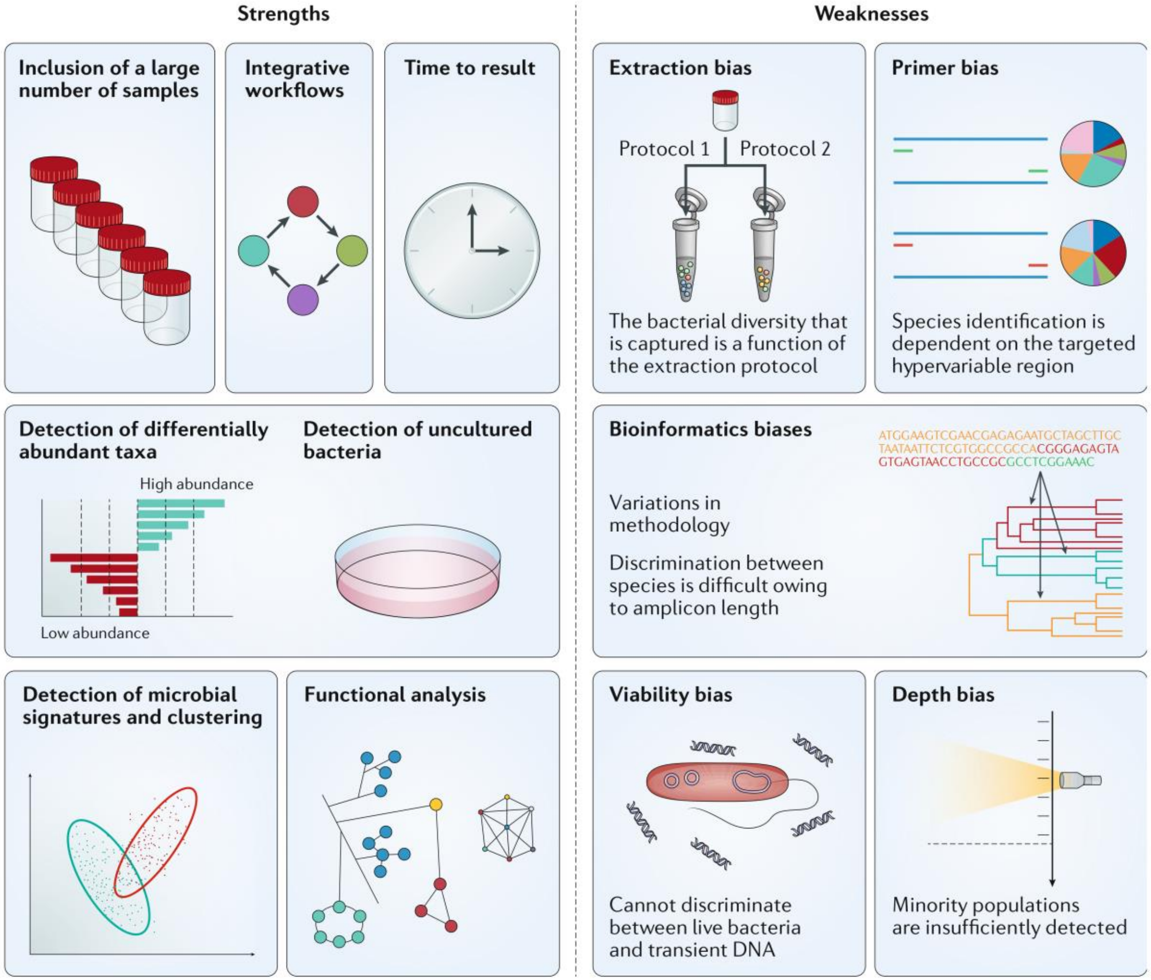
Metagenomics

Sequencing

Computational analysis

- Unknown microbes (“microbial dark matter”)
- Reconstruction of multiple genomes

Strengths and weaknesses of metagenomic studies



Inclusion of a large number of samples

Integrative workflows

Time to result

Detection of differentially abundant taxa

Detection of uncultured bacteria

Detection of microbial signatures and clustering

Functional analysis

Extraction bias

The bacterial diversity that is captured is a function of the extraction protocol

Primer bias

Species identification is dependent on the targeted hypervariable region

Bioinformatics biases

Variations in methodology

Discrimination between species is difficult owing to amplicon length

Viability bias

Cannot discriminate between live bacteria and transient DNA

Depth bias

Minority populations are insufficiently detected

Lagier et al., **Culturing the human microbiota and culturomics**, *Nature Reviews Microbiology*, 2018

The shotgun metagenomics framework

Caption of the previous figure. Summary of a metagenomics workflow.

Step (1): study design and experimental protocol. The importance of this step is often underestimated in metagenomics.

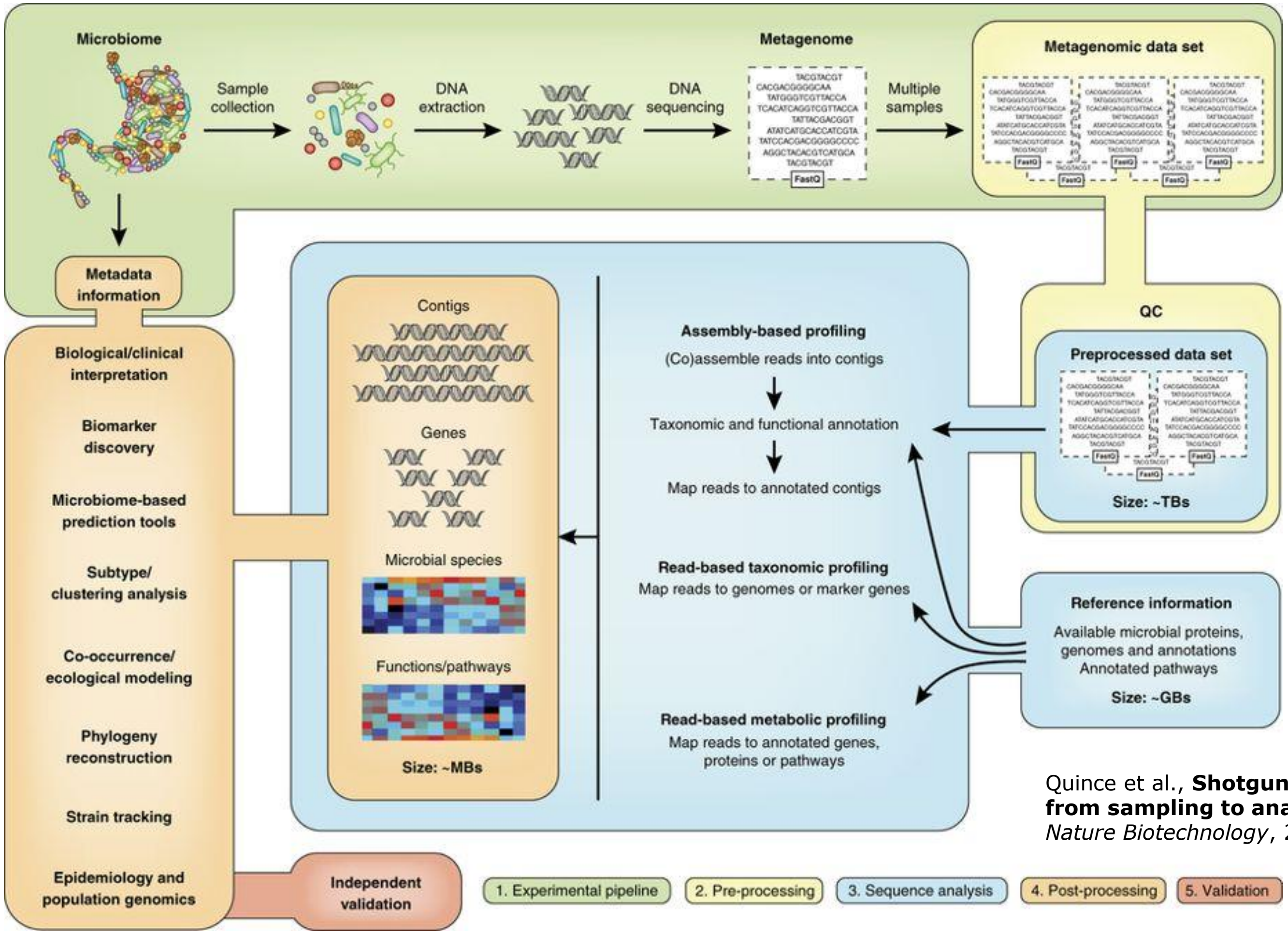
Step (2): computational pre-processing. Computational quality control (QC) steps minimize fundamental sequence biases or artifacts such as removal of sequencing adaptors, quality trimming, removal of sequencing duplicates (using for example, FastQC, Trimmomatic or Picard tools). Foreign or non-target DNA sequences are also filtered, and samples are subsampled to normalize read numbers if the diversity of taxa or functions is compared.

Step (3): sequence analysis. This should comprise a combination of 'read-based' and 'assembly-based' approaches depending on the experimental objectives. Both approaches have advantages and limitations.

Step (4): post-processing. Various multivariate statistical techniques can be used to interpret the data.

Step (5): validation. Conclusions from high-dimensional biological data are susceptible to study-driven biases, so follow-up analyses are vital.

The shotgun metagenomics framework



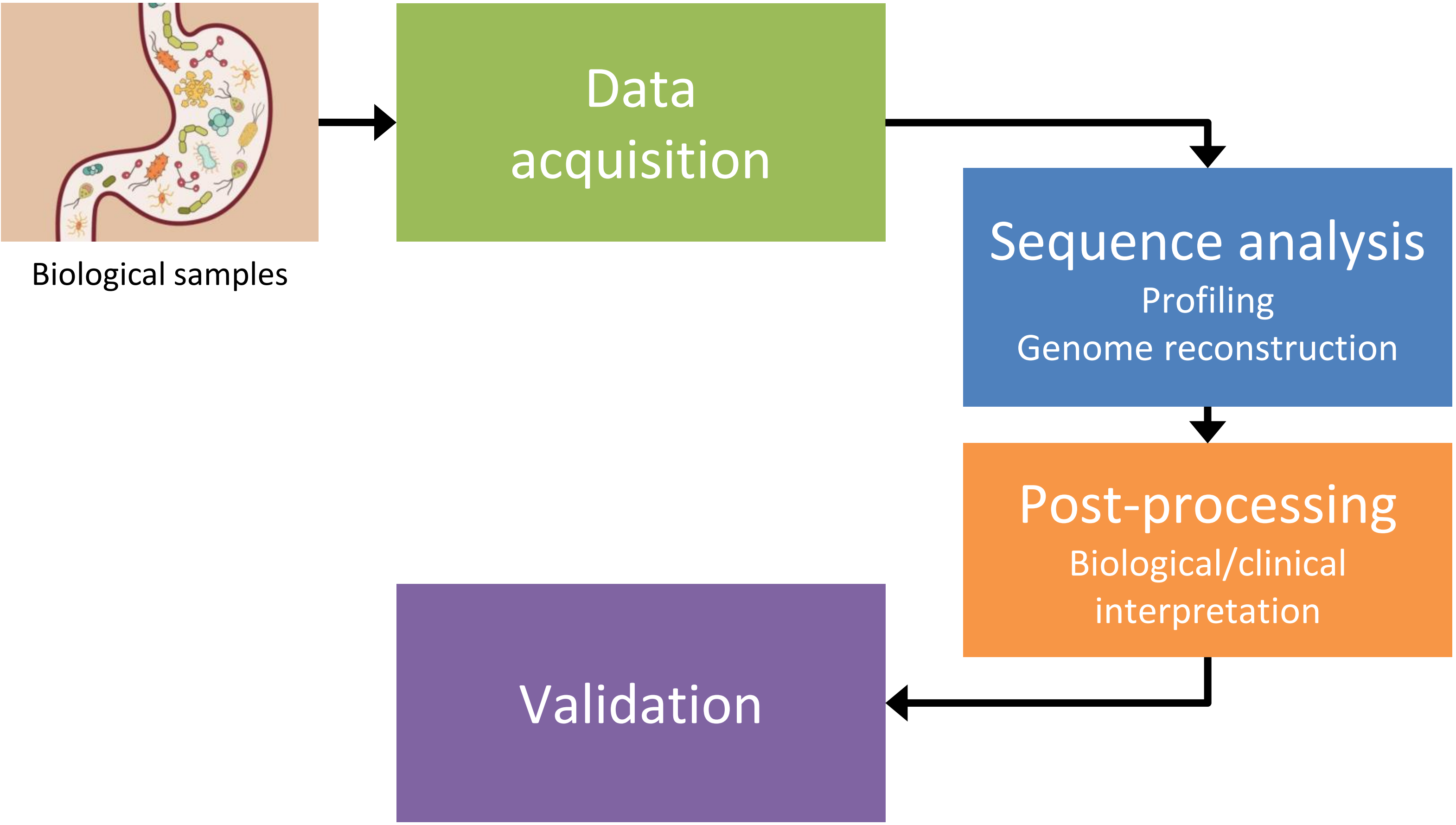
Quince et al., **Shotgun metagenomics, from sampling to analysis**, *Nature Biotechnology*, 2017

Strengths and weaknesses of metagenomic studies

Caption of the previous figure. The strengths and weaknesses of metagenomic studies.

High- throughput sequencing methods have enabled comparisons of large cohorts in record time, as the time it takes to generate results has substantially decreased over the past decade owing to integrative workflows. Sequencing enables uncultured bacteria to be identified and associates microbial signatures with a particular physiological state or disease as the relative abundance of each taxon can be measured. In addition, bacterial species can be grouped without taxonomic assignment. Shotgun sequencing enables the function of microbial communities to be inferred through the analysis of genomes and the coding potential. However, these methods are limited by the heterogeneity of the protocols used. Discrepant results can be obtained depending on the method used to extract DNA or the primers that are used for amplification. The variety of methodologies proposed for bioinformatics analyses (for example, operational taxonomic unit clustering, taxonomic assignment or statistical analysis) can substantially affect the results. Sequencing methods cannot discriminate between live bacteria and transient DNA , and despite recent progress, they cannot easily detect minority populations.

The shotgun metagenomics framework



Biological samples

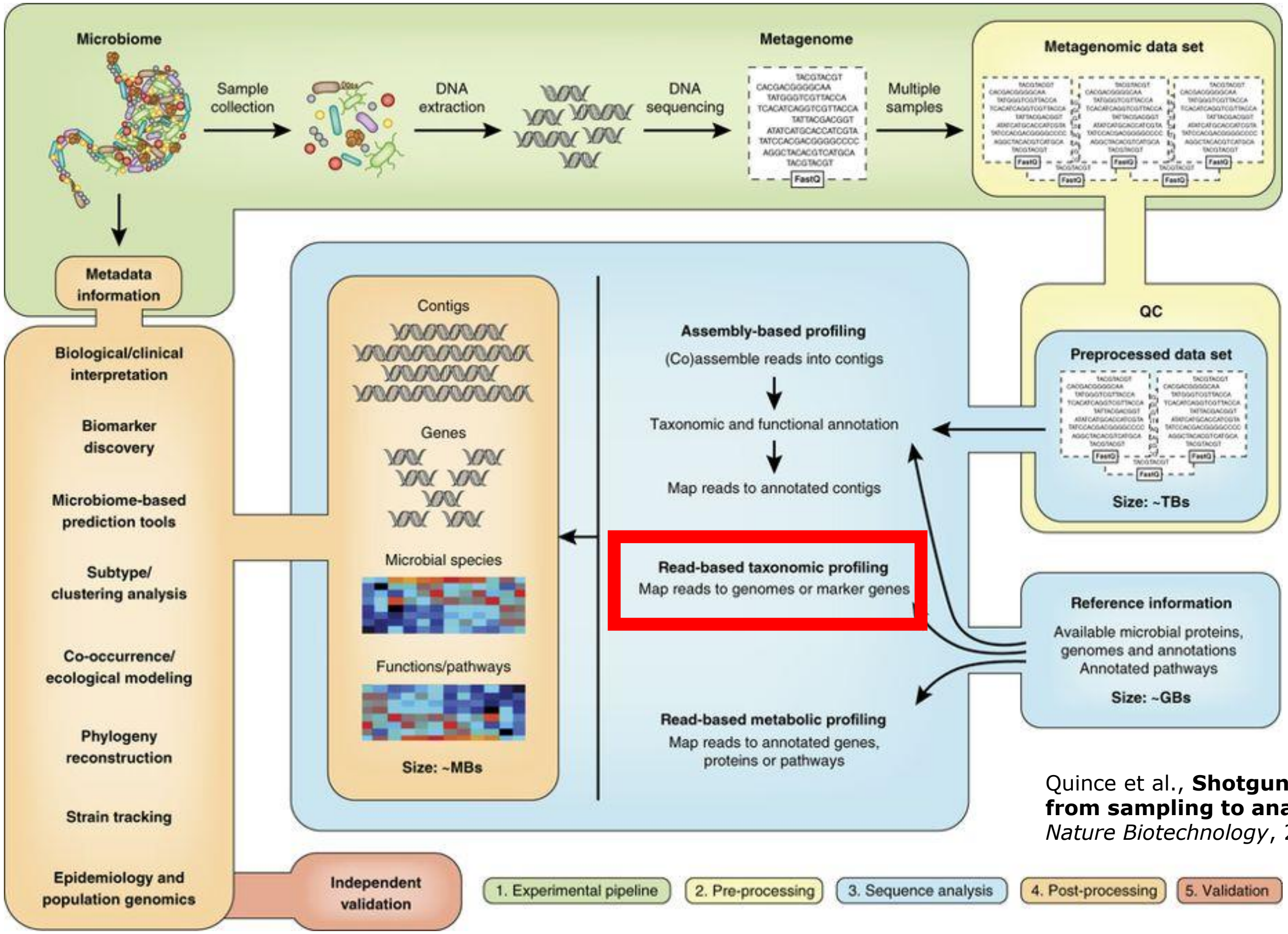
Data acquisition

Sequence analysis
Profiling
Genome reconstruction

Post-processing
Biological/clinical interpretation

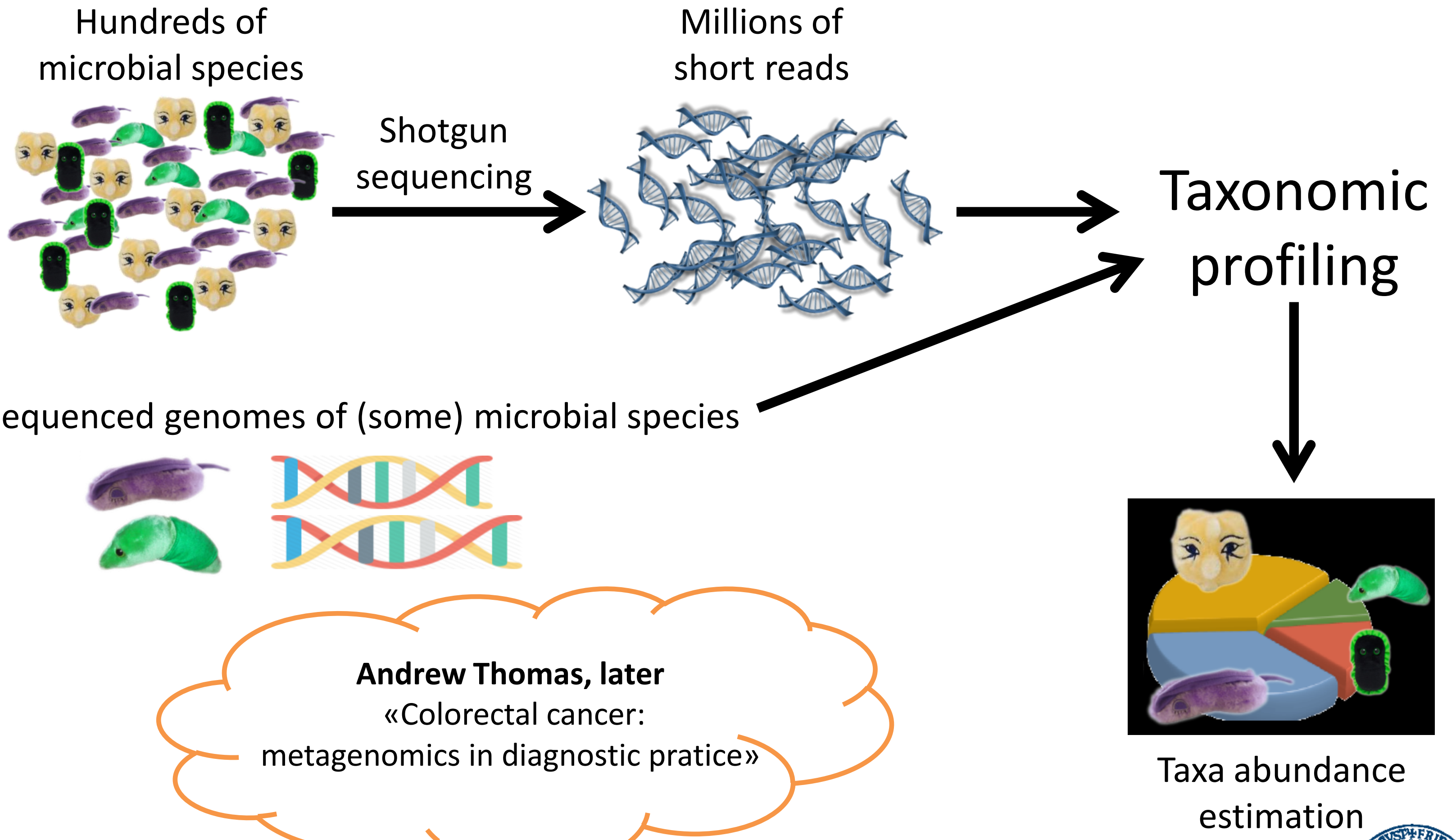
Validation

The shotgun metagenomics framework



Quince et al., **Shotgun metagenomics, from sampling to analysis**, *Nature Biotechnology*, 2017

Mapping-based taxonomic profiling



Hundreds of microbial species

Shotgun sequencing

Millions of short reads

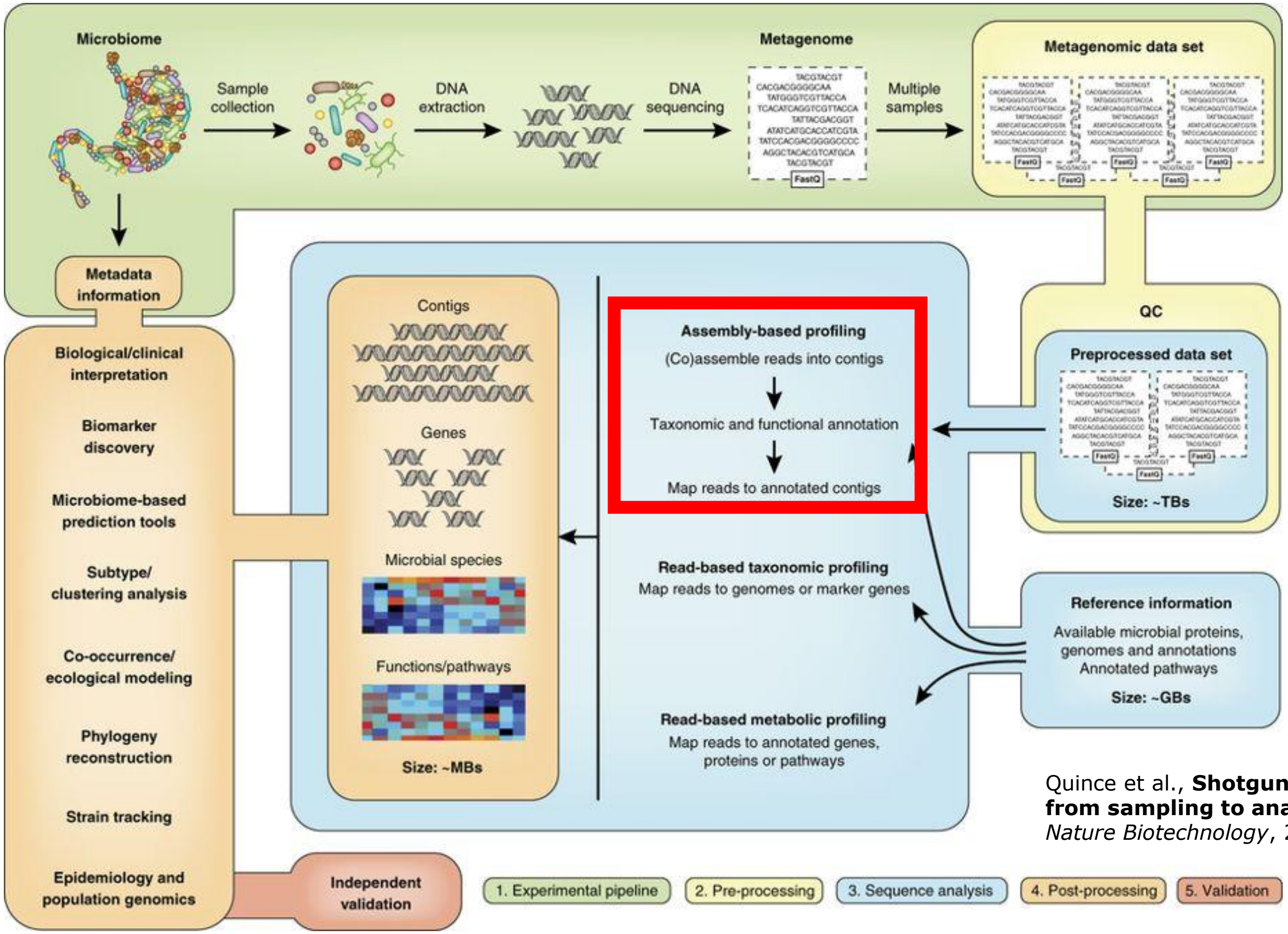
Taxonomic profiling

Sequenced genomes of (some) microbial species

Taxa abundance estimation

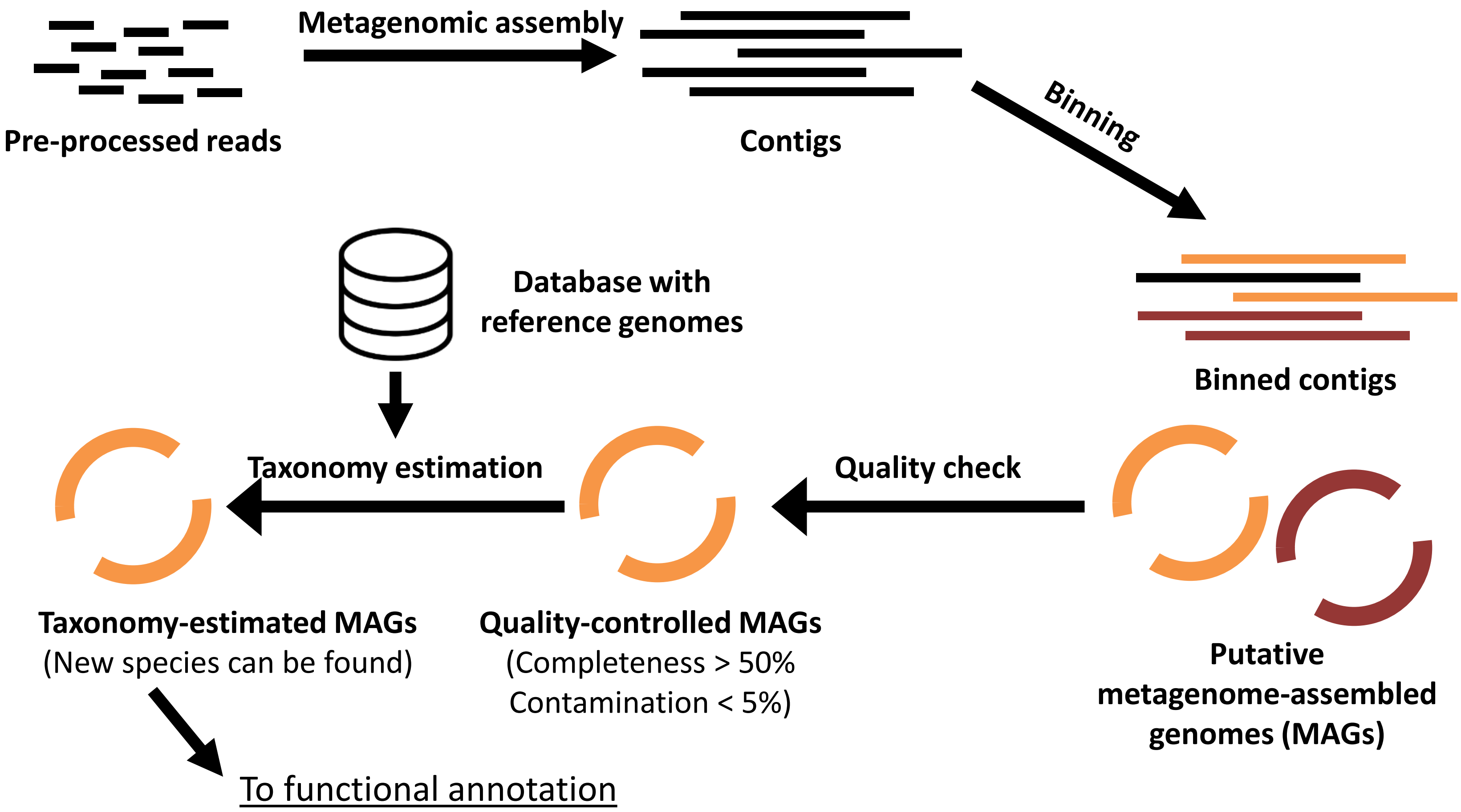
Andrew Thomas, later
«Colorectal cancer:
metagenomics in diagnostic pratice»

The shotgun metagenomics framework



Quince et al., **Shotgun metagenomics, from sampling to analysis**, *Nature Biotechnology*, 2017

Metagenomic assembly and genome reconstruction


Pre-processed reads

Metagenomic assembly

Contigs

Binning

Binned contigs

Database with reference genomes

Quality check

Taxonomy estimation

Putative metagenome-assembled genomes (MAGs)

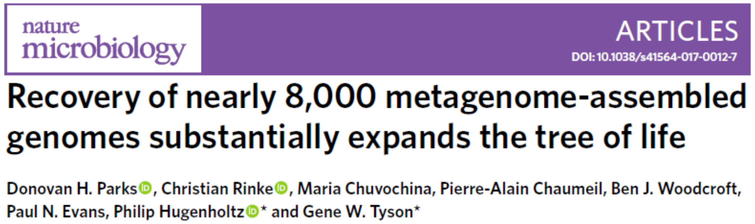
Quality-controlled MAGs
(Completeness > 50%
Contamination < 5%)

Taxonomy-estimated MAGs
(New species can be found)

To functional annotation

Some examples, different environments

Environmental

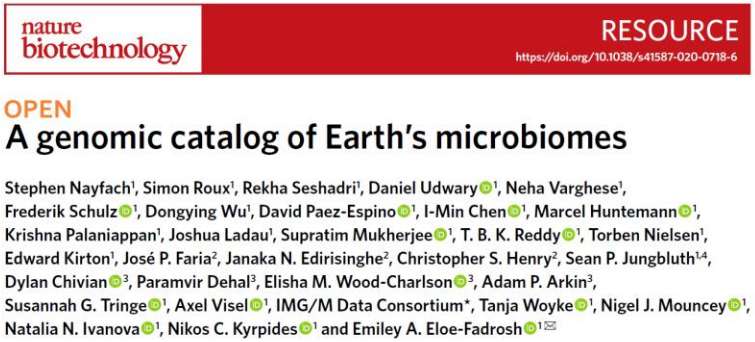

nature microbiology

ARTICLES

DOI: 10.1038/s41564-017-0012-7

Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life

Donovan H. Parks, Christian Rinke, Maria Chuvochina, Pierre-Alain Chaumeil, Ben J. Woodcroft, Paul N. Evans, Philip Hugenholtz* and Gene W. Tyson*


nature biotechnology

RESOURCE

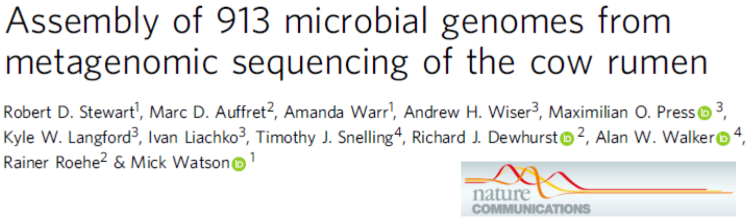
https://doi.org/10.1038/s41587-020-0718-6

OPEN

A genomic catalog of Earth's microbiomes

Stephen Nayfach[1], Simon Roux[1], Rekha Seshadri[1], Daniel Udwary[1], Neha Varghese[1], Frederik Schulz[1], Dongying Wu[1], David Paez-Espino[1], I-Min Chen[1], Marcel Huntemann[1], Krishna Palaniappan[1], Joshua Ladau[1], Supratim Mukherjee[1], T. B. K. Reddy[1], Torben Nielsen[1], Edward Kirton[1], José P. Faria[2], Janaka N. Edirisinghe[2], Christopher S. Henry[2], Sean P. Jungbluth[1,4], Dylan Chivian[3], Paramvir Dehal[3], Elisha M. Wood-Charlson[3], Adam P. Arkin[3], Susannah G. Tringe[1], Axel Visel[1], IMG/M Data Consortium*, Tanja Woyke[1], Nigel J. Mouncey[1], Natalia N. Ivanova[1], Nikos C. Kyrpides[1] and Emiley A. Eloe-Fadrosh[1] ✉

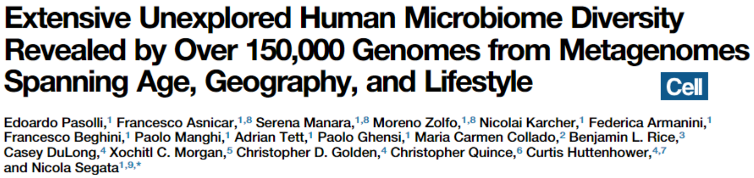
Rumen


Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen

Robert D. Stewart[1], Marc D. Auffret[2], Amanda Warr[1], Andrew H. Wiser[3], Maximilian O. Press[3], Kyle W. Langford[3], Ivan Liachko[3], Timothy J. Snelling[4], Richard J. Dewhurst[2], Alan W. Walker[4], Rainer Roehe[2] & Mick Watson[1]

nature COMMUNICATIONS

Human


Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle

Cell

Edoardo Pasolli,[1] Francesco Asnicar,[1,8] Serena Manara,[1,8] Moreno Zolfo,[1,8] Nicolai Karcher,[1] Federica Armanini,[1] Francesco Beghini,[1] Paolo Manghi,[1] Adrian Tett,[1] Paolo Ghensi,[1] Maria Carmen Collado,[2] Benjamin L. Rice,[3] Casey DuLong,[4] Xochitl C. Morgan,[5] Christopher D. Golden,[4] Christopher Quince,[6] Curtis Huttenhower,[4,7] and Nicola Segata[1,9,*]


A new genomic blueprint of the human gut microbiota

nature

Alexandre Almeida[1,2*], Alex L. Mitchell[1], Miguel Boland[1], Samuel C. Forster[2,3,4], Gregory B. Gloor[5], Aleksandra Tarkowska[1], Trevor D. Lawley[2] & Robert D. Finn[1*]


New insights from uncultivated genomes of the global human gut microbiome

Stephen Nayfach[1,2*], Zhou Jason Shi[3,4], Rekha Seshadri[1,2], Katherine S. Pollard[3,4,5,6,7,8] & Nikos C. Kyrpides[1,2*]

nature

Food


ARTICLES

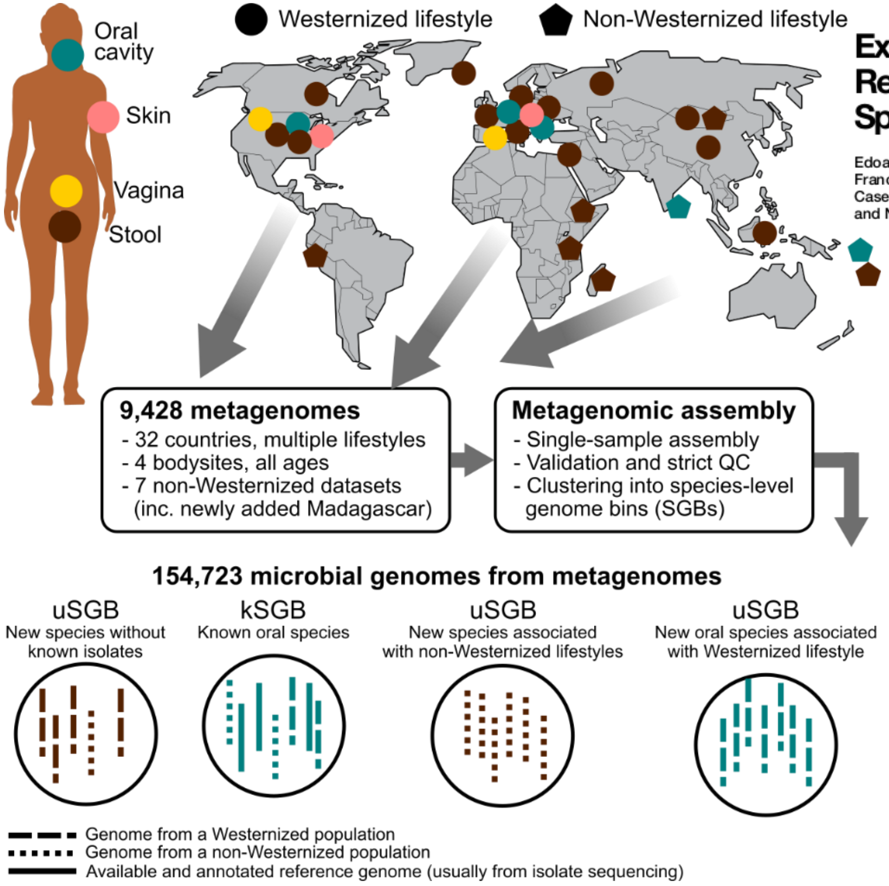
https://doi.org/10.1038/s43016-020-0129-3

nature food

Meta-analysis of cheese microbiomes highlights contributions to multiple aspects of quality

Aaron M. Walsh[1,2,3], Guerrino Macori[1,2,3], Kieran N. Kilcawley[1] and Paul D. Cotter[1,2] ✉

The largest genome catalogue



9,428 metagenomes
- 32 countries, multiple lifestyles
- 4 bodysites, all ages
- 7 non-Westernized datasets (inc. newly added Madagascar)

Metagenomic assembly
- Single-sample assembly
- Validation and strict QC
- Clustering into species-level genome bins (SGBs)

154,723 microbial genomes from metagenomes

uSGB	kSGB	uSGB	uSGB
New species without known isolates	Known oral species	New species associated with non-Westernized lifestyles	New oral species associated with Westernized lifestyle

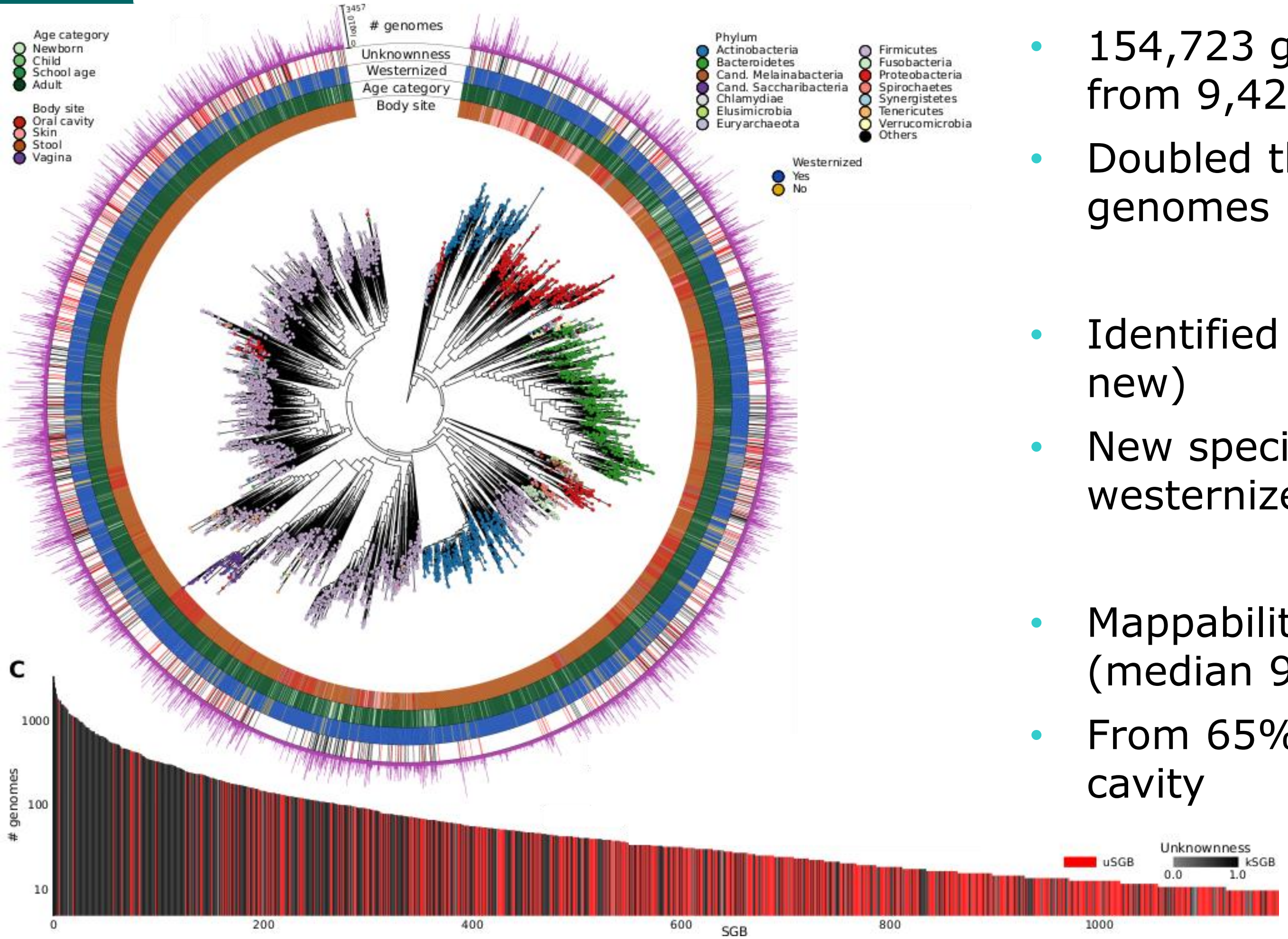
▬ ▬ Genome from a Westernized population
▪▪▪▪▪▪ Genome from a non-Westernized population
▬▬ Available and annotated reference genome (usually from isolate sequencing)

Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle

Edoardo Pasolli,[1] Francesco Asnicar,[1,8] Serena Manara,[1,8] Moreno Zolfo,[1,8] Nicolai Karcher,[1] Federica Armanini,[1] Francesco Beghini,[1] Paolo Manghi,[1] Adrian Tett,[1] Paolo Ghensi,[1] Maria Carmen Collado,[2] Benjamin L. Rice,[3] Casey DuLong,[4] Xochitl C. Morgan,[5] Christopher D. Golden,[4] Christopher Quince,[6] Curtis Huttenhower,[4,7] and Nicola Segata[1,9,*]

Cell

Human MAGs associated with almost 5,000 species

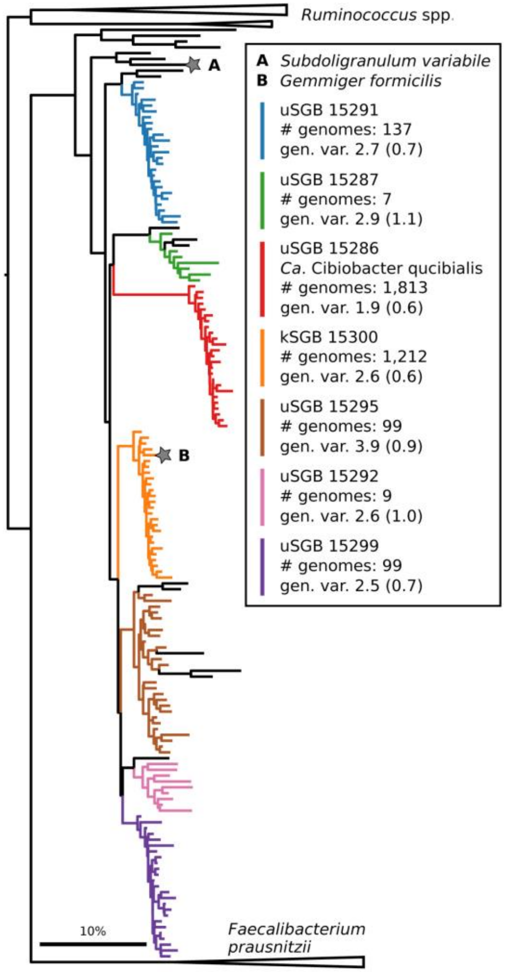


- 154,723 genomes reconstructed from 9,428 metagenomes
- Doubled the number of available genomes

- Identified 4,930 species (77% new)
- New species prevalent in non-westernized populations

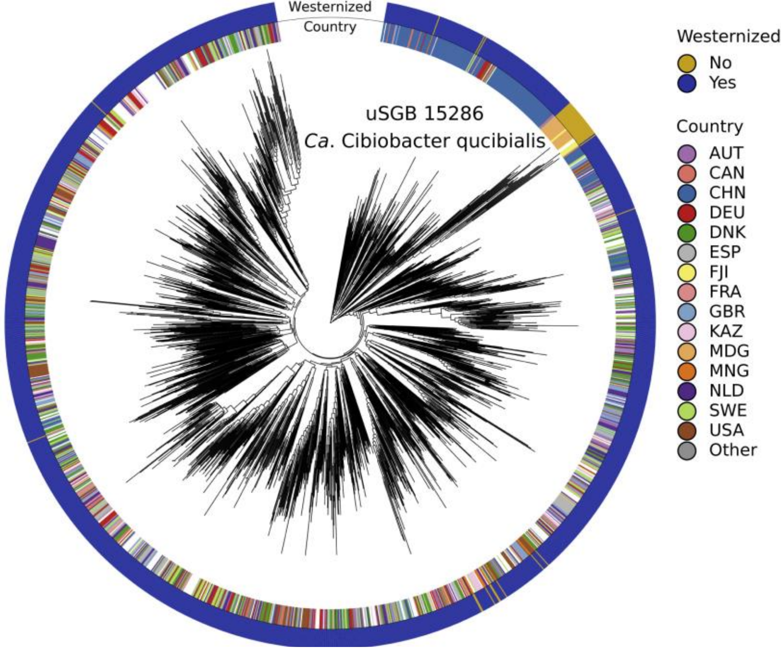
- Mappability from 68% to 88% (median 94%) in the gut
- From 65% to 82% in the oral cavity

More insights from uncharacterized species

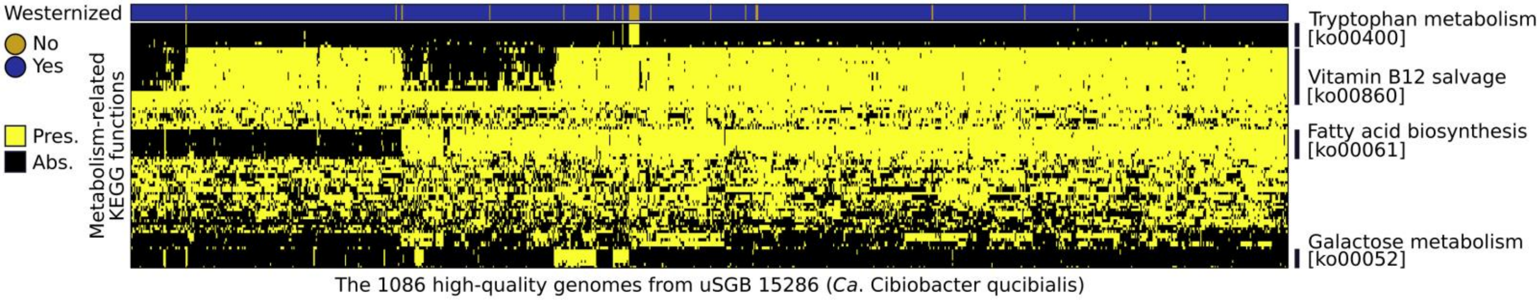


- Several uncharacterized taxa fall into the Clostridiales order
- As example, six prevalent species are between *Ruminococcus* and *Faecalibacterium* spp.
- *Candidatus* Cibiobacter qucibialis is the most prevalent uSGB

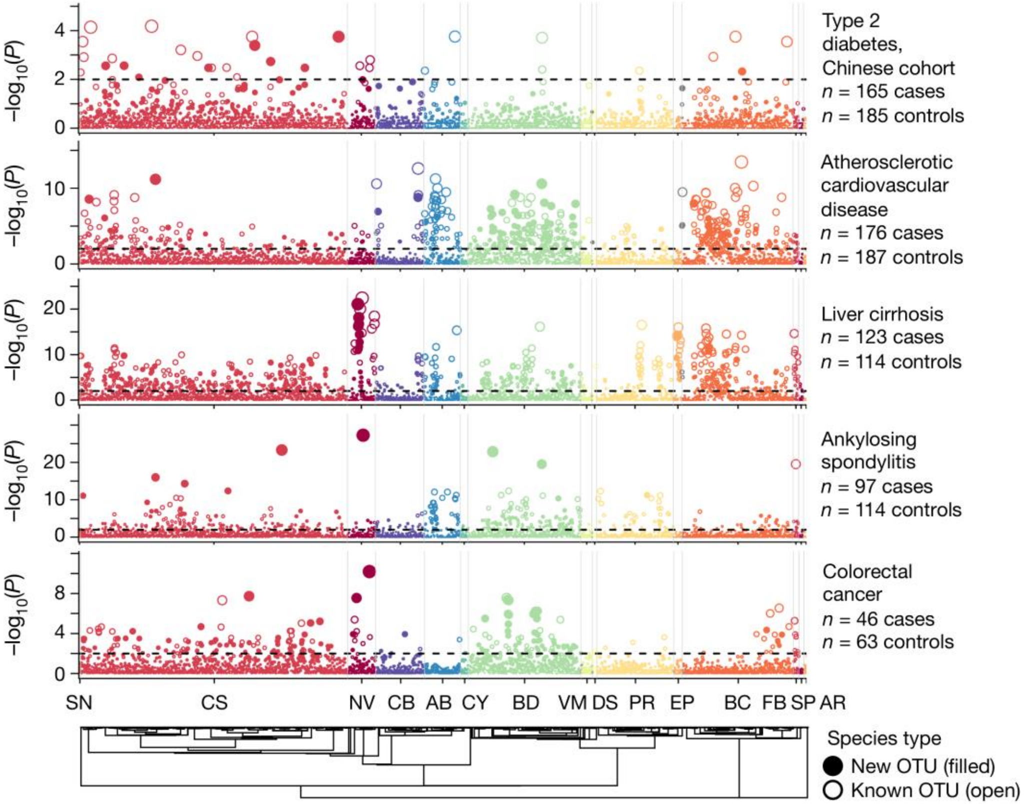
More insights from uncharacterized species



- Prevalent and uncharacterized species may be associated to different lifestyles
- The operon trp for tryptophan metabolism found only in non-westernized populations



Association of gut taxa with human diseases

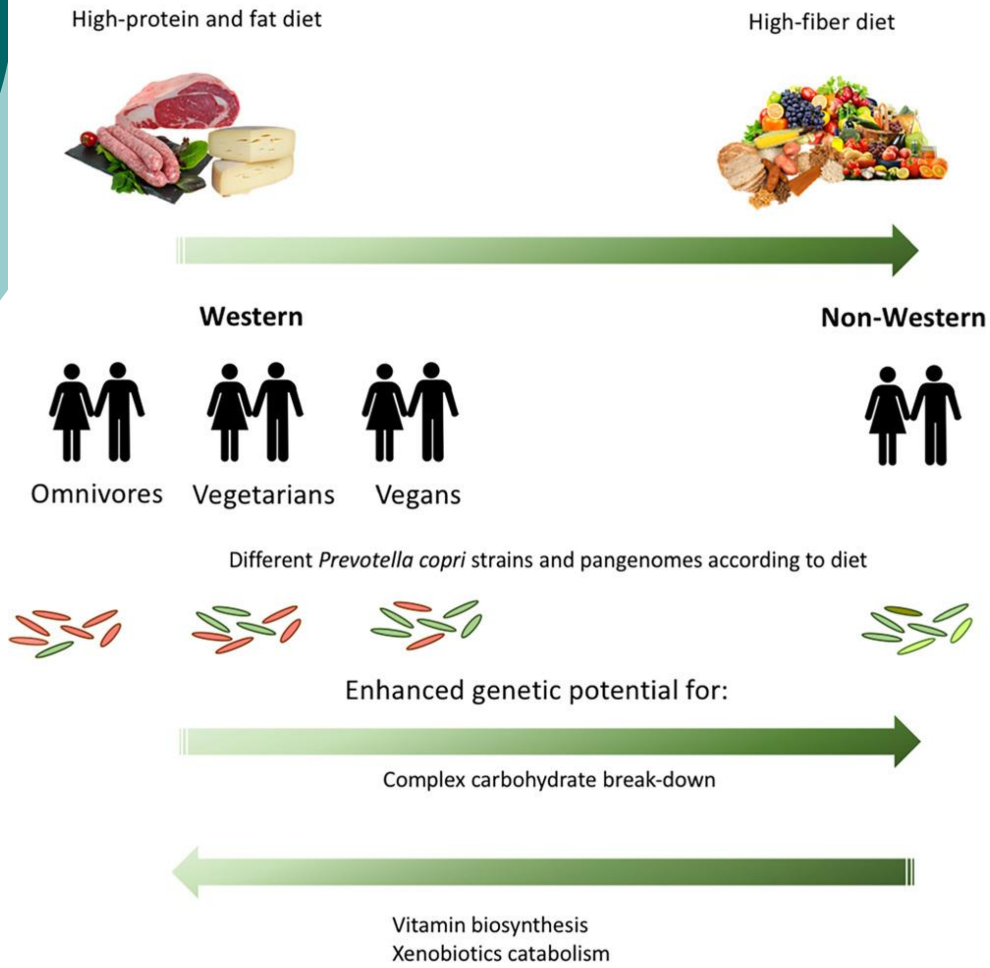


nature

New insights from uncultivated genomes of the global human gut microbiome

Stephen Nayfach[1,2*], Zhou Jason Shi[3,4], Rekha Seshadri[1,2], Katherine S. Pollard[3,4,5,6,7,8] & Nikos C. Kyrpides[1,2*]

Prevotella copri & habitual diet



High-protein and fat diet

High-fiber diet

Western

Non-Western

Omnivores

Vegetarians

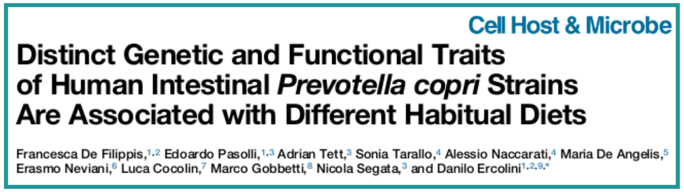
Vegans

Different *Prevotella copri* strains and pangenomes according to diet

Enhanced genetic potential for:

Complex carbohydrate break-down

Vitamin biosynthesis
Xenobiotics catabolism



Cell Host & Microbe

Distinct Genetic and Functional Traits of Human Intestinal *Prevotella copri* Strains Are Associated with Different Habitual Diets

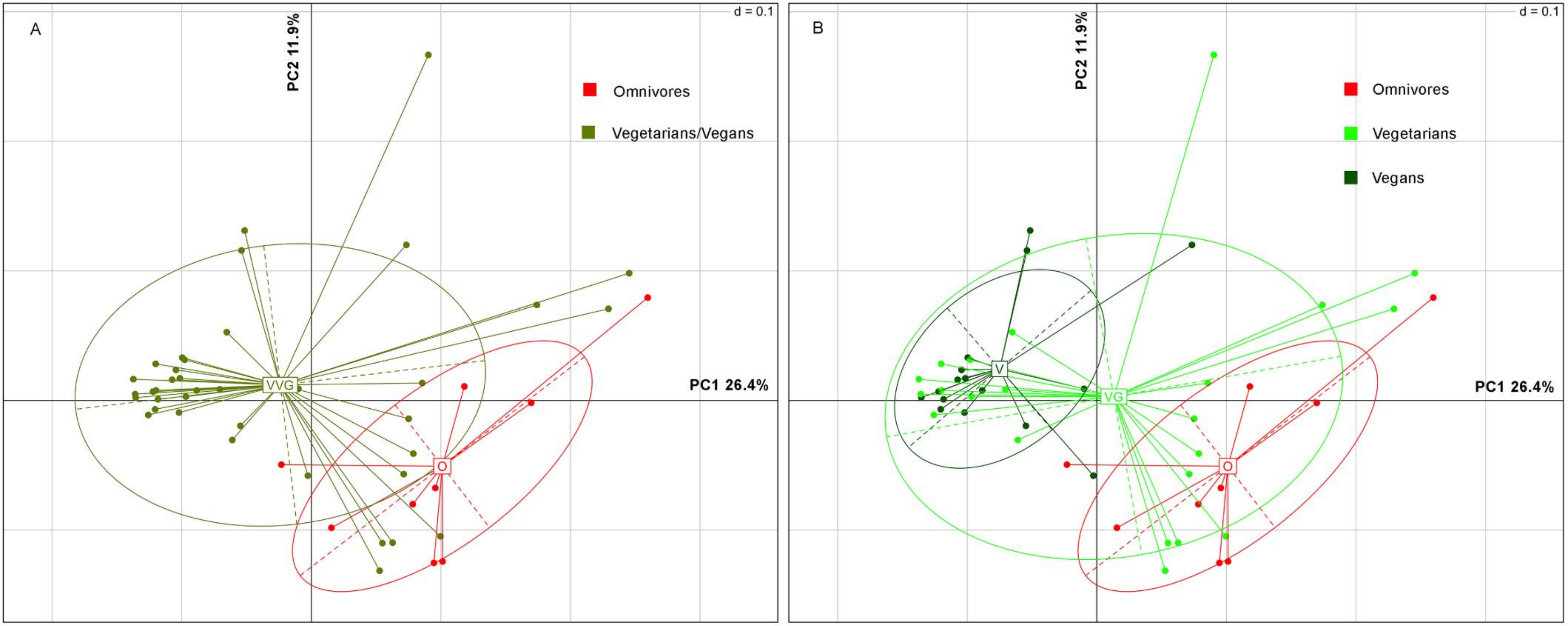
Francesca De Filippis,[1,2] Edoardo Pasolli,[1,3] Adrian Tett,[3] Sonia Tarallo,[4] Alessio Naccarati,[4] Maria De Angelis,[5] Erasmo Neviani,[6] Luca Cocolin,[7] Marco Gobbetti,[8] Nicola Segata,[3] and Danilo Ercolini[1,2,9,*]

Prevotella copri & habitual diet

Cell Host & Microbe

Distinct Genetic and Functional Traits of Human Intestinal *Prevotella copri* Strains Are Associated with Different Habitual Diets

Francesca De Filippis,[1,2] Edoardo Pasolli,[1,3] Adrian Tett,[3] Sonia Tarallo,[4] Alessio Naccarati,[4] Maria De Angelis,[5] Erasmo Neviani,[6] Luca Cocolin,[7] Marco Gobbetti,[8] Nicola Segata,[3] and Danilo Ercolini[1,2,9,*]

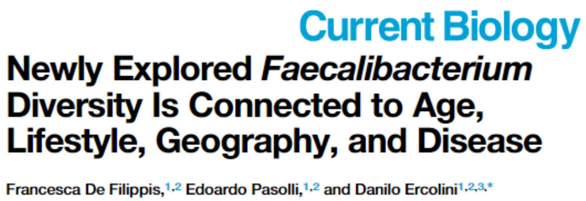


Trieste, 11 dicembre 2020

Docente: Edoardo Pasolli
edoardo.pasolli@unina.it

Newly explored *Faecalibacterium* diversity

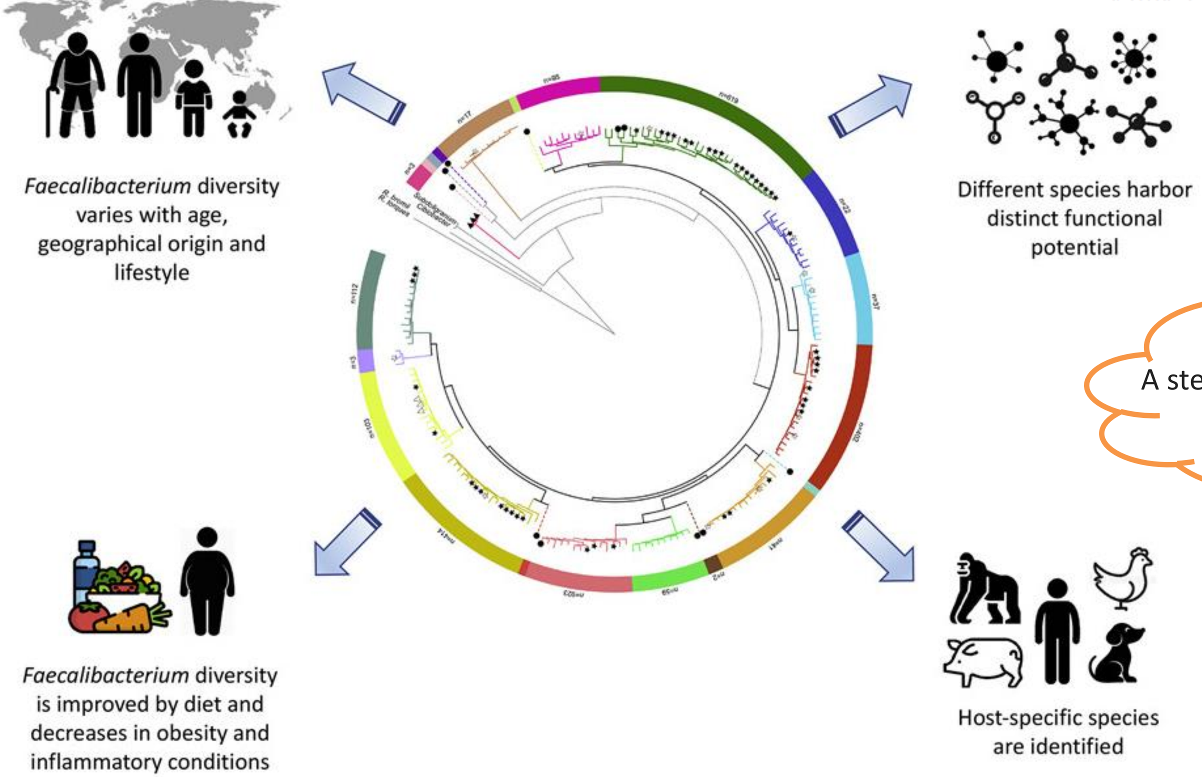
- ✓ 22 genetically distinct *Faecalibacterium* species exist
- ✓ 12 species are prevalent in human gut
- ✓ On average, 5 species co-occur in the same subject



Current Biology

Newly Explored *Faecalibacterium* Diversity Is Connected to Age, Lifestyle, Geography, and Disease

Francesca De Filippis,[1,2] Edoardo Pasolli,[1,2] and Danilo Ercolini[1,2,3,*]



Faecalibacterium diversity varies with age, geographical origin and lifestyle

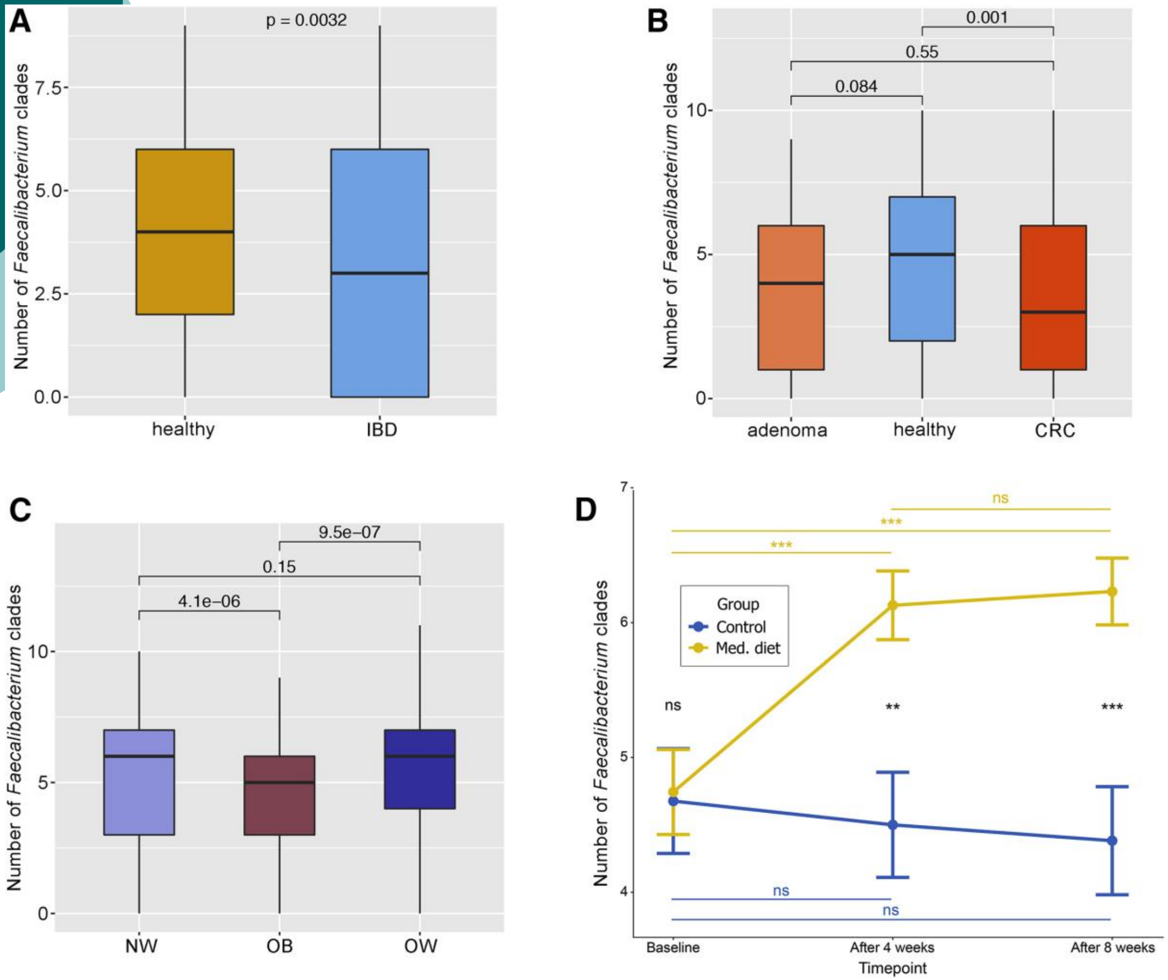
Different species harbor distinct functional potential

Faecalibacterium diversity is improved by diet and decreases in obesity and inflammatory conditions

Host-specific species are identified

A step forward in the use of *F. prausnitzii* as a next-generation probiotic

Newly explored *Faecalibacterium* diversity



Current Biology

Newly Explored *Faecalibacterium* Diversity Is Connected to Age, Lifestyle, Geography, and Disease

Francesca De Filippis,[1,2] Edoardo Pasolli,[1,2] and Danilo Ercolini[1,2,3,*]

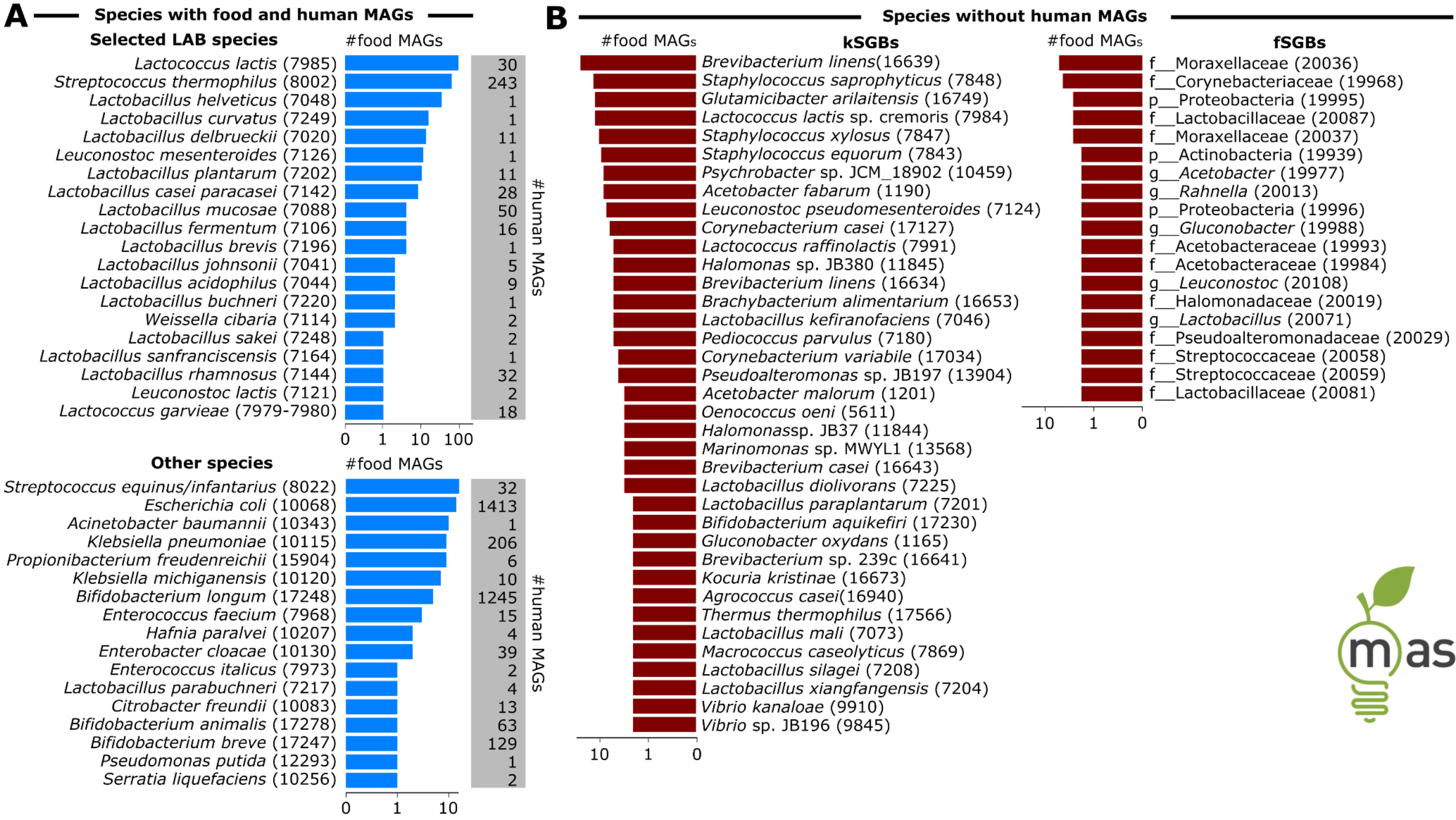
Linking food and human microbiomes

Large-scale genome-wide analysis links lactic acid bacteria from food with the gut microbiome

Edoardo Pasolli[1,2], Francesca De Filippis[1,2], Italia E. Mauriello[1], Fabio Cumbo[3], Aaron M. Walsh[4,5], John Leech[4,5], Paul D. Cotter[4,5], Nicola Segata[3] & Danilo Ercolini[1,2]✉

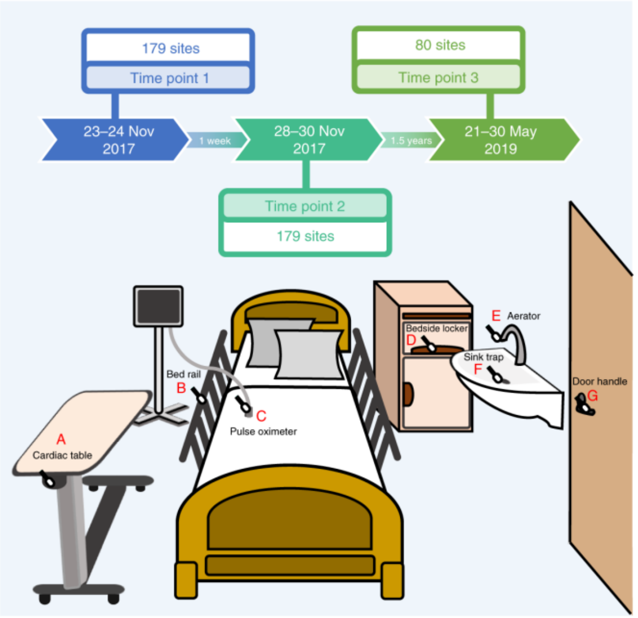
N = 303 food metagenomes (from 11 public available datasets)

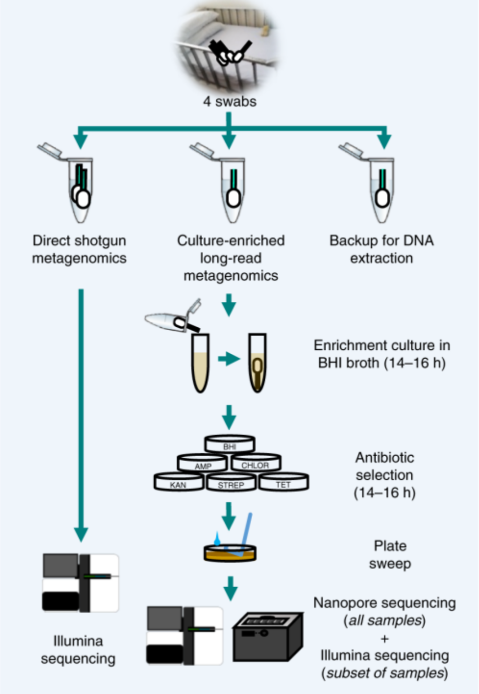
N = 9,445 human metagenomes



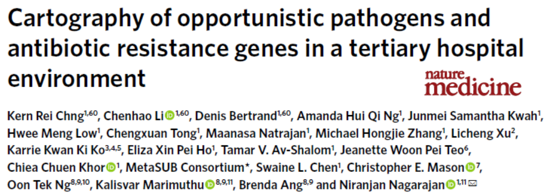
master

Large-scale analysis in a tertiary-care hospital in Singapore





- Most extensive genomic characterization of microbiomes, pathogens and antibiotic resistance cassettes in a tertiary-care hospital
- Repeated sampling (up to 1.5 years apart) of 179 sites associated with 45 beds
- Shotgun + nanopore sequencing provided thousands of MAGs + phage and plasmid sequences (>60% novel)
- Enabling characterization of resistome diversity and spatio-temporal patterns
- Identification of widely distributed and stably colonizing multidrug-resistant strains



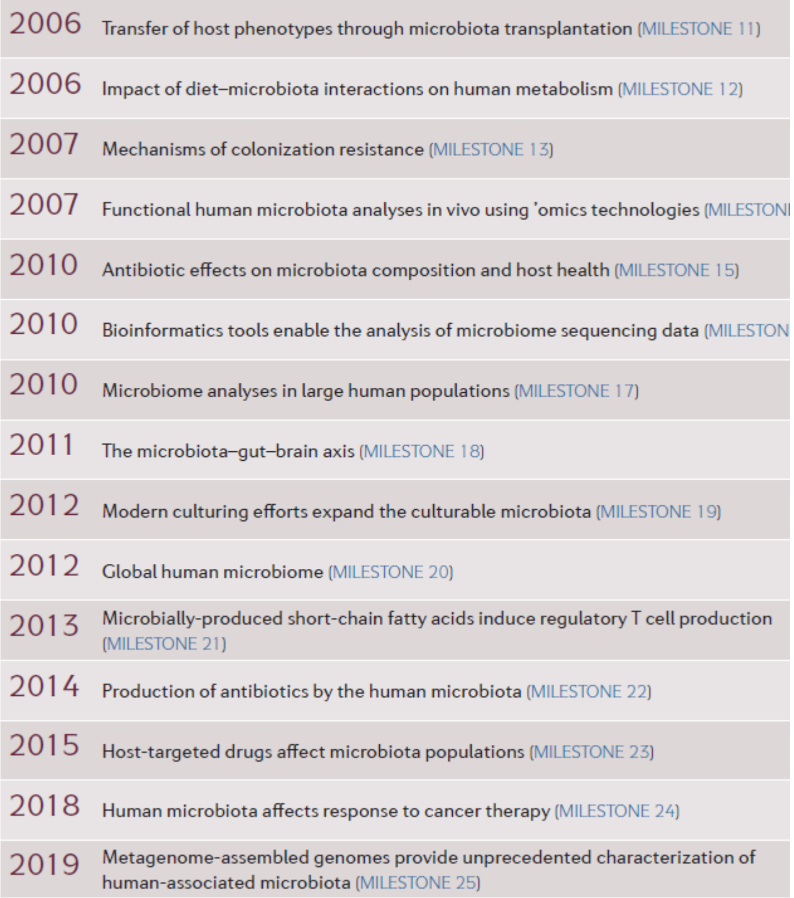
Cartography of opportunistic pathogens and antibiotic resistance genes in a tertiary hospital environment

nature medicine

Kern Rei Chng[1,60], Chenhao Li[1,60], Denis Bertrand[1,60], Amanda Hui Qi Ng[1], Junmei Samantha Kwah[1], Hwee Meng Low[1], Chengxuan Tong[1], Maanasa Natrajan[1], Michael Hongjie Zhang[1], Licheng Xu[2], Karrie Kwan Ki Ko[3,4,5], Eliza Xin Pei Ho[1], Tamar V. Av-Shalom[1], Jeanette Woon Pei Teo[6], Chiea Chuen Khor[1], MetaSUB Consortium*, Swaine L. Chen[1], Christopher E. Mason[7], Oon Tek Ng[8,9,10], Kalisvar Marimuthu[8,9,11], Brenda Ang[8,9] and Niranjan Nagarajan[1,11] ✉

Further reading

https://www.nature.com/immersive/d42859-019-00041-z/pdf/d42859-019-00041-z.pdf



2006	Transfer of host phenotypes through microbiota transplantation (MILESTONE 11)
2006	Impact of diet–microbiota interactions on human metabolism (MILESTONE 12)
2007	Mechanisms of colonization resistance (MILESTONE 13)
2007	Functional human microbiota analyses in vivo using 'omics technologies (MILESTONI
2010	Antibiotic effects on microbiota composition and host health (MILESTONE 15)
2010	Bioinformatics tools enable the analysis of microbiome sequencing data (MILESTON
2010	Microbiome analyses in large human populations (MILESTONE 17)
2011	The microbiota–gut–brain axis (MILESTONE 18)
2012	Modern culturing efforts expand the culturable microbiota (MILESTONE 19)
2012	Global human microbiome (MILESTONE 20)
2013	Microbially-produced short-chain fatty acids induce regulatory T cell production (MILESTONE 21)
2014	Production of antibiotics by the human microbiota (MILESTONE 22)
2015	Host-targeted drugs affect microbiota populations (MILESTONE 23)
2018	Human microbiota affects response to cancer therapy (MILESTONE 24)
2019	Metagenome-assembled genomes provide unprecedented characterization of human-associated microbiota (MILESTONE 25)

nature MILESTONES

June 2019
www.nature.com/collections/microbiota-milestone

Human microbiota research

REGIONE AUTONOMA FRIULI VENEZIA GIULIA

ISTITUTO DI RICOVERO E CURA
A CARATTERE SCIENTIFICO
Burlo Garofolo di Trieste

BURLO

E.C.M.

IRCBG_20081
“Metagenomica e culturomica: aspetti complementari della microbiologia omica"

Metagenome-assembled genomes (MAGs) to uncover the hidden human microbiome diversity

Trieste, 11 dicembre 2020

Docente: Edoardo Pasolli
edoardo.pasolli@unina.it