Metagenomics 101

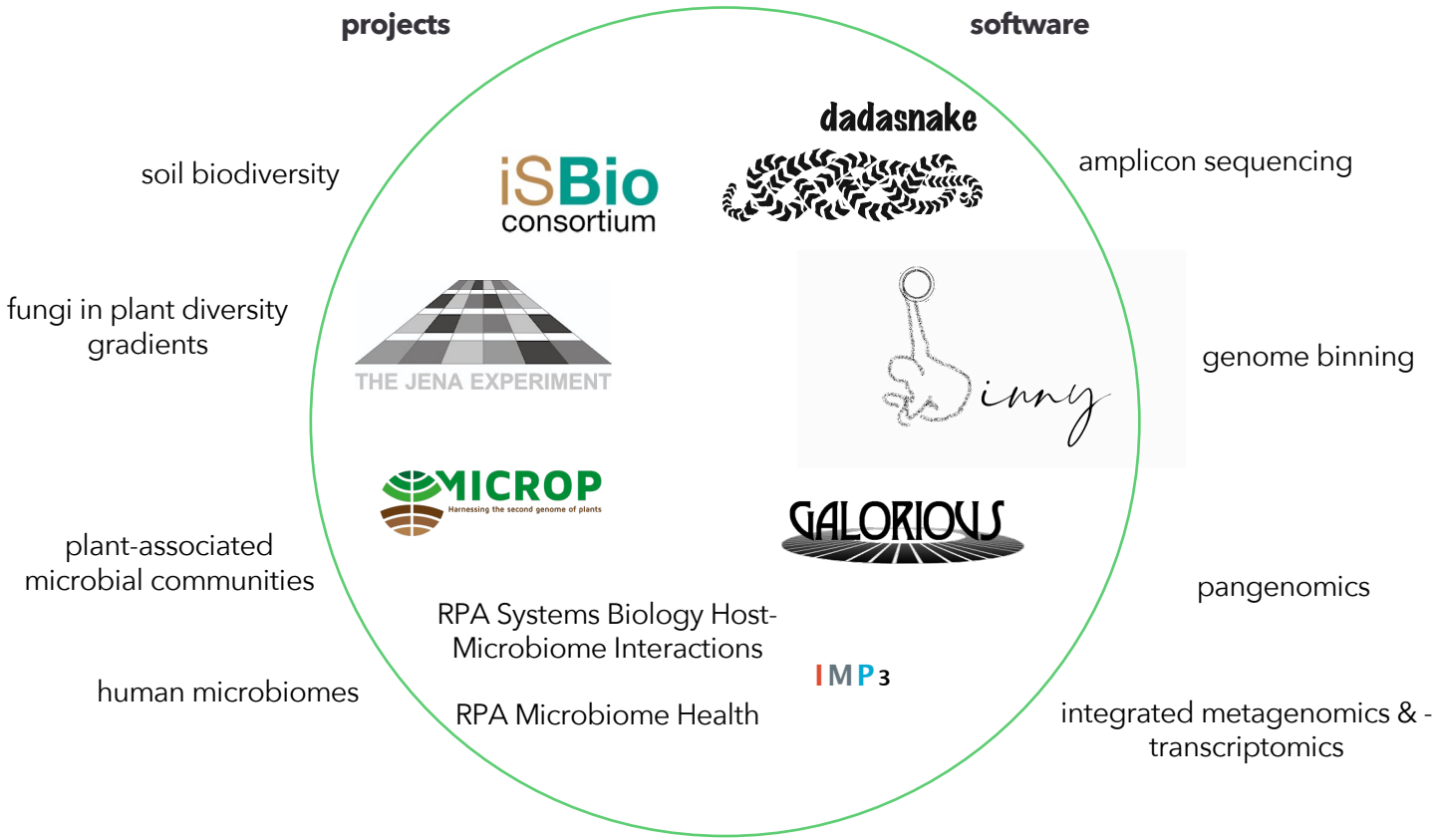
Anna Heintz-Buschart

February 2022

University of Amsterdam
Life Sciences

Who is here?

- A.H.–B.:



2008

MSc Biology (Microbiology, Botany, Molecular & Cell Biology)

PhD: Fungal human pathogen
- compound screening, mode-of action
- gene expression analysis

2011

Postdoc: Gene regulatory network modelling

2012

Postdoc: Integrated meta-omics
- human microbiome, wastewater treatment
- metagenomics, metatranscriptomics, metaproteomics
- lab automation
- bioinformatics pipelines

2017

Metagenomics support:
- biodiversity
- soil, plants, animal microbiomes
- bioinformatics pipelines
- data integration

2021

Assistant Prof Microbial Metagenomics
- meta-omics integration
- human and plant microbiomes

Who is here?

“Metagenomics”

- “**directly** accessing the **genomes** of […] organism**s** that cannot be, or have not been, cultured **by isolating their DNA**” (Handelsman *et al*. Chem Biol. 1998)

- “accessing” (nowadays): by sequencing
- uncultured organisms: usually more than one

Tweet

Willem van Schaik
@WvSchaik

Replying to @WvSchaik

And 16S is not metagenomics

6:31 PM · Nov 9, 2015 · Twitter for Android

Tweet

Irene Newton
@chicaScientific

16S rRNA gene sequencing is NOT metagenomics, people

4:50 PM · Dec 29, 2020 · Twitter W

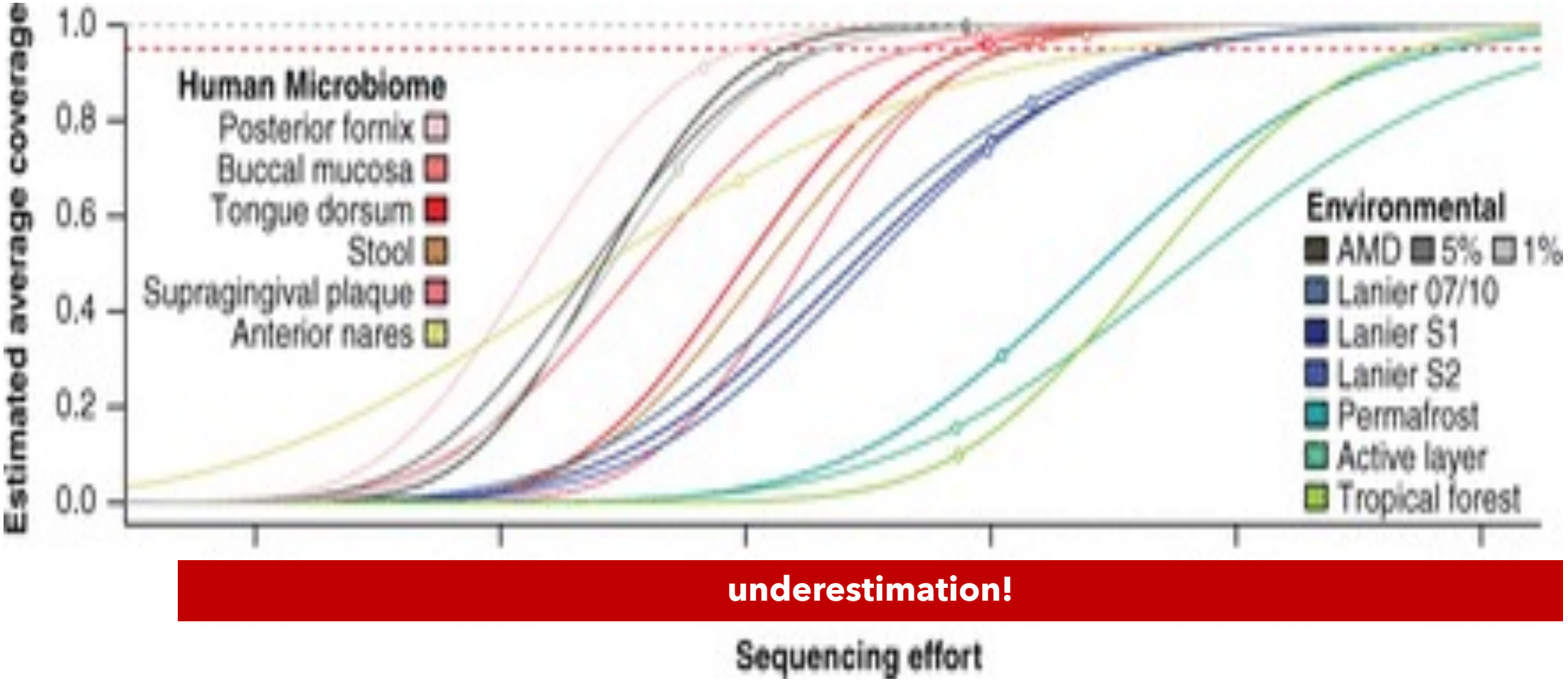
Tweet

Ken McGrath
@DrKenMcGrath

Hey @illumina - 16S profiling is not metagenomics. Clear distinction here should avoid confusion for people starting out in the field.

This course is about shotgun metagenomics

WHOLE metagenome shotgun sequencing?

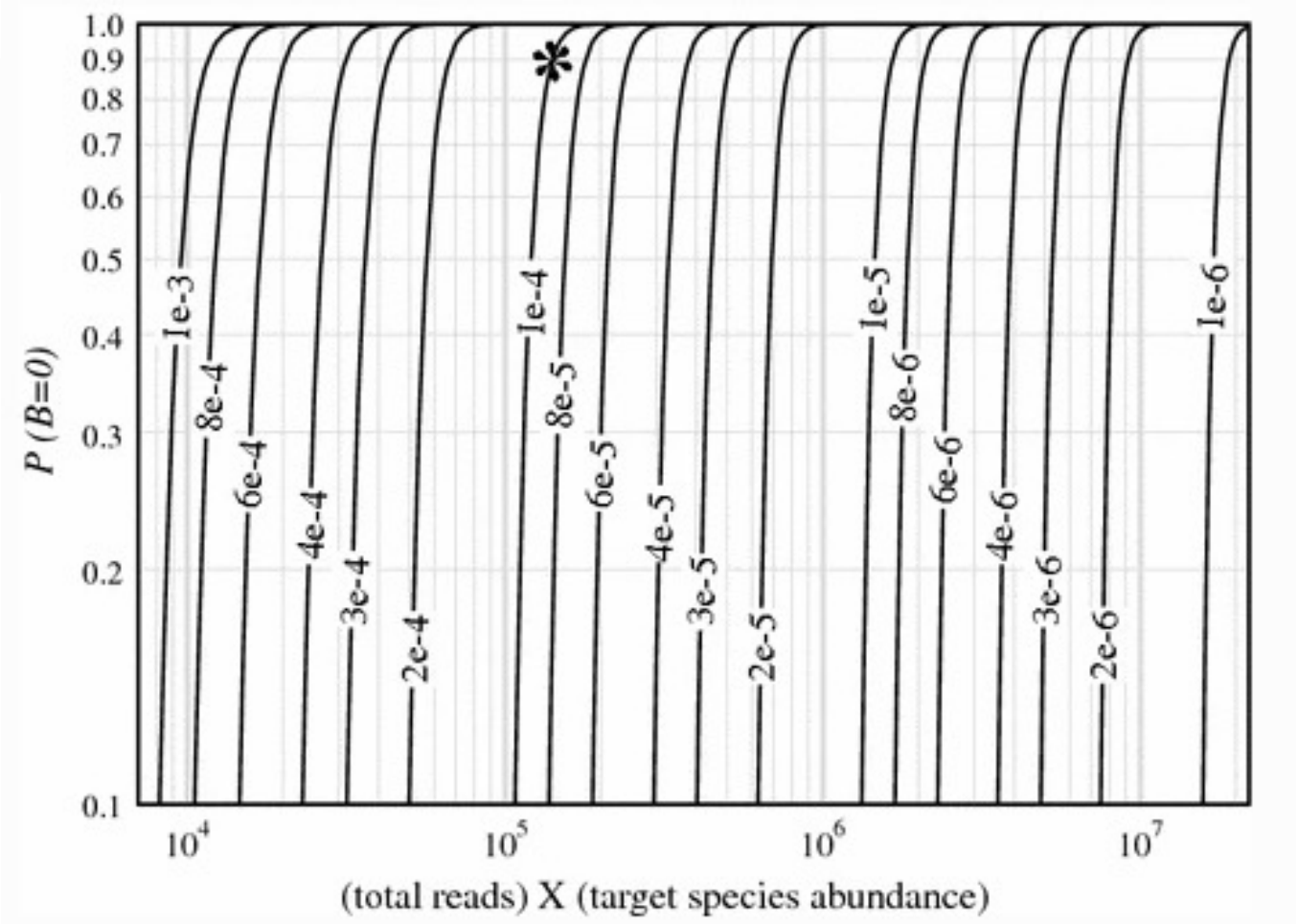


Rodriguez-R & Konstantinidis. 2014 Bioinformatics

Sequencing depth and metagenomic coverage

$$P(B = k) = \binom{R}{k} \sum_{\beta=k}^{\eta} \binom{R-k}{\beta-k} (-1)^{\beta-k} \alpha^{\beta} (1-\beta\varphi)^{\beta-1} (1-\beta\varphi\alpha)^{R-\beta}$$

P(B=k): probability of k gaps
alpha: relative abundance
R: number of reads
phi: probability of a position being covered = L/gamma
L: read length
gamma: genome size
eta: the smaller of R or the maximum number of non-overlapping reads on the genome



Wendl et al. 2013 J Math Biol

Sequencing depth and metagenomic coverage

- **beware**: these calculations are still often underestimations, because of:

- sequencing biases
- uneven copy numbers during replication
- regions with high inter-species similarity and horizontal gene transfer
- repeats
- micro-diversity / strains

Short & biased history of metagenomics

- **2004 beginnings**

articles

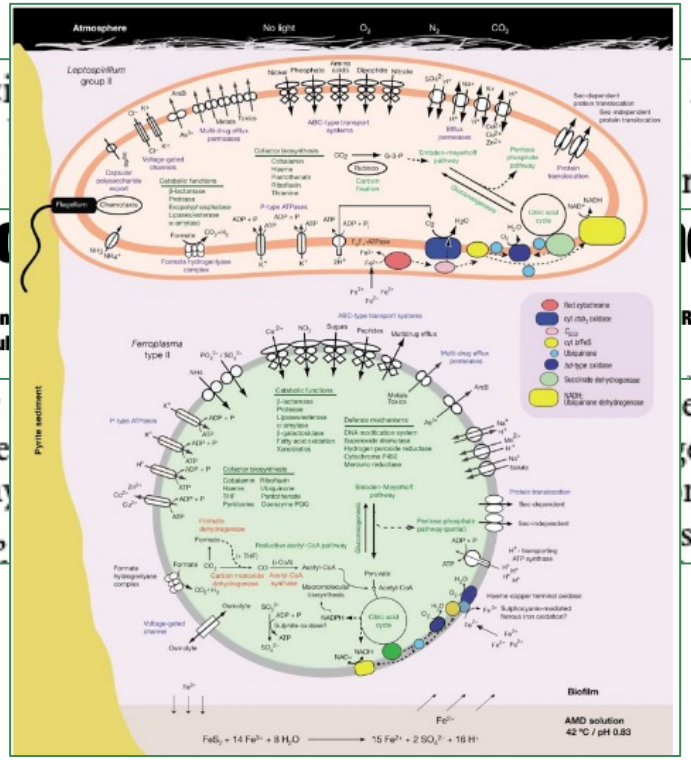
sequenci … A total of 76.2
million … generated from
103,462 … r read). Analysis

gen… …ent

Gene W. Tyson … Richardson[4], Victor V. Solovyev[1],
Edward M. Rub…

…sics, University of California,

straints. Over … ed into
scaffolds longe… genomic
region that may… mbined
length of the 1,… ssembly



TIGR role category	Total genes
Amino acid biosynthesis	37,118
Biosynthesis of cofactors, prosthetic groups, and carriers	25,905
Cell envelope	27,883
Cellular processes	17,260
Central intermediary metabolism	13,639
DNA metabolism	25,346
Energy metabolism	69,718
Fatty acid and phospholipid metabolism	18,558
Mobile and extrachromosomal element functions	1,061
Protein fate	28,768
Protein synthesis	48,012
Purines, pyrimidines, nucleosides, and nucleotides	19,912
Regulatory functions	8,392
Signal transduction	4,817
Transcription	12,756
Transport and binding proteins	49,185
Unknown function	38,067
Miscellaneous	1,864
Conserved hypothetical	794,061
Total number of roles assigned	1,242,230
Total number of genes	1,214,207

collec… A total of 1.045 billion base pairs
of nor… …tated, and analyzed to elucidate

…emblies. Our
…l-sampled ge-
…s with at least
…333 scaffolds
…panning 30.9
…S3), account-
ing for roughly 410,000 reads, or 25% of the

Short & biased history of metagenomics

- **2007-2012 'large-scale' projects**

OPEN ACCESS Freely available online

PLOS | BIOLOGY

Community Page

The Human Microbiome Project: A Community Resource for the Healthy Human Microbiome

Dirk Gevers[1], Rob Knight[2,3], Joseph F. Petrosino[4,5,6], Katherine Huang[1], Amy L. McGuire[7], Bruce W. Birren[1], Karen E. Nelson[8], Owen White[9], Barbara A. Methé[8]*, Curtis Huttenhower[1,10]*

1 The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America, **2** Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado, United States of America, **3** Howard Hughes Medical Institute, Boulder, Colorado, United States of America, **4** Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, United States of America, **5** Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas, United States of America, **6** Alkek Center for Metagenomics and Microbiome Research, Baylor College of Medicine, Houston, Texas, United States of America, **7** Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, Texas, United States of America, **8** J. Craig Venter Institute, Rockville, Maryland, United States of America, **9** Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, United States of America, **10** Biostatistics, Harvard School of Public Health, Boston, Massachusetts, United States of America

Short & biased history of metagenomics

- **around 2014 tailored tools**

Bioinformatics, 31(10), 2015, 1674–1676
doi: 10.1093/bioinformatics/btv033
Advance Access Publication Date: 20 January 2015
Applications Note

OXFORD

Sequence analysis

MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph

Dinghua Li[1,†], Chi-Man Liu[2,†], Ruibang Luo[2,†], Kunihiko Sadakane[3] and Tak-Wah Lam[1,2,*]

BRIEF COMMUNICATIONS

Binning metagenomic contigs by coverage and composition

Johannes Alneberg[1,8], Brynjar Smári Bjarnason[1,8], Ino de Bruijn[1,2], Melanie Schirmer[3], Joshua Quick[4,5], Umer Z Ijaz[3], Leo Lahti[6,7], Nicholas J Loman[4], Anders F Andersson[1,9] & Christopher Quince[3,9]

PeerJ

MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities

Dongwan D. Kang[1,2], Jeff Froula[1,2], Rob Egan[1,2] and Zhong Wang[1,2,3]

[1] Department of Energy Joint Genome Institute, Walnut Creek, CA, USA
[2] Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[3] School of Natural Sciences, University of California at Merced, Merced, CA, USA

Short & biased history of metagenomics

- **since 2015 pipelines & platforms**

Bioinformatics, 32(16), 2016, 2520–2523
doi: 10.1093/bioinformatics/btw183
Advance Access Publication Date: 8 April 2016
Applications Note

Genome analysis

MOCAT2: a metagenomic assembly, annotation and profiling framework

Jens Roat Kultima[1], Luis Pedro Coelho[1], Kristoffer Forslund[1], Jaime Huerta-Cepas[1], Simone S. Li[1,2], Marja Driessen[1], Anita Yvonne Voigt[1,3], Georg Zeller[1], Shinichi Sunagawa[1] and Peer Bork[1,3,4,5,*]

frontiers in Microbiology
TECHNOLOGY REPORT
published: 24 January 2019
doi: 10.3389/fmicb.2018.03349

SqueezeMeta, A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline

Javier Tamames and Fernando Puente-Sánchez*

eLife
TOOLS AND RESOURCES

Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3

Francesco Beghini[1†], Lauren J McIver[2†], Aitor Blanco-Míguez[1], Leonard Dubois[1], Francesco Asnicar[1], Sagun Maharjan[2,3], Ana Mailyan[2,3], Paolo Manghi[1], Matthias Scholz[4], Andrew Maltez Thomas[1], Mireia Valles-Colomer[1], George Weingart[2,3], Yancong Zhang[2,3], Moreno Zolfo[1], Curtis Huttenhower[2,3*], Eric A Franzosa[2,3*], Nicola Segata[1,5*]

PeerJ

Anvi'o: an advanced analysis and visualization platform for 'omics data

A. Murat Eren[1,2], Özcan C. Esen[1], Christopher Quince[3], Joseph H. Vineis[1], Hilary G. Morrison[1], Mitchell L. Sogin[1] and Tom O. Delmont[1]

[1] Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, MA, United States
[2] Department of Medicine, The University of Chicago, Chicago, IL, United States
[3] Warwick Medical School, University of Warwick, Coventry, United Kingdom

Uritskiy *et al. Microbiome* (2018) 6:158
https://doi.org/10.1186/s40168-018-0541-1

Microbiome

SOFTWARE Open Access

MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis

Gherman V. Uritskiy, Jocelyne DiRuggiero* and James Taylor*

Narayanasamy *et al. Genome Biology* (2016) 17:260
DOI 10.1186/s13059-016-1116-8

Genome Biology

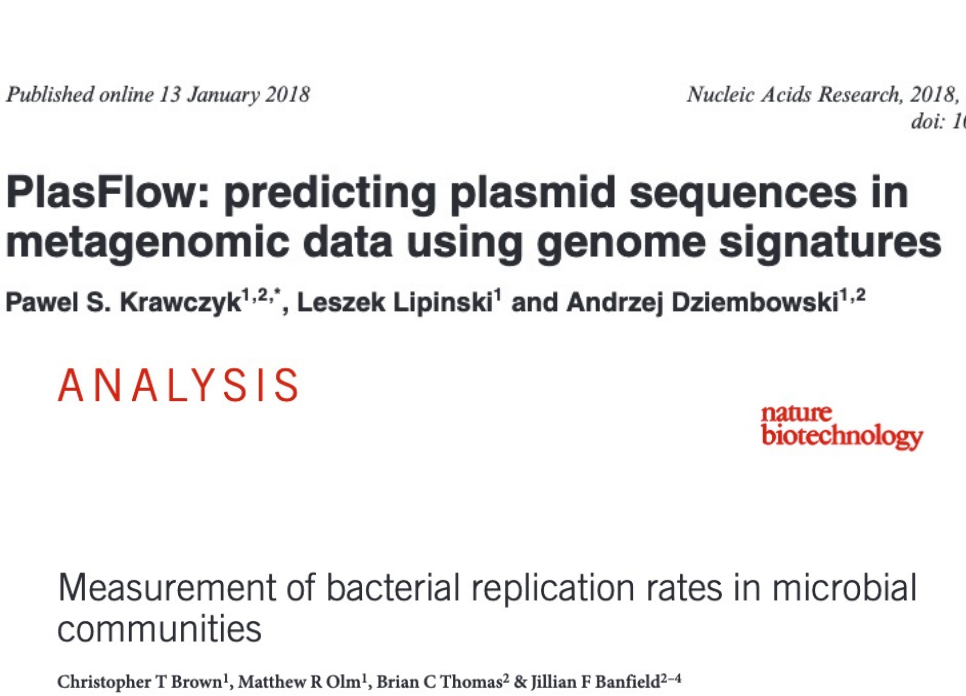
SOFTWARE Open Access

IMP: a pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses

Shaman Narayanasamy[1†], Yohan Jarosz[1†], Emilie E. L. Muller[1,2], Anna Heintz-Buschart[1], Malte Herold[1], Anne Kaysen[1], Cédric C. Laczny[1,3], Nicolás Pinel[4,5], Patrick May[1] and Paul Wilmes[1*]

Short & biased history of metagenomics

- **since 2017 diversification of tools**



Published online 13 January 2018

Nucleic Acids Research, 2018, Vol. 46, No. 6 e35
doi: 10.1093/nar/gkx1321

PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures

Pawel S. Krawczyk[1,2,*], Leszek Lipinski[1] and Andrzej Dziembowski[1,2]

ANALYSIS

nature biotechnology

Measurement of bacterial replication rates in microbial communities

Christopher T Brown[1], Matthew R Olm[1], Brian C Thomas[2] & Jillian F Banfield[2–4]

MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies

Dongwan D. Kang[1], Feng Li[2], Edward Kirton[1], Ashleigh Thomas[1], Rob Egan[1], Hong An[2] and Zhong Wang[1,3,4]

Lind and Pollard *Microbiome* (2021) 9:58
https://doi.org/10.1186/s40168-021-01015-y

Microbiome

METHODOLOGY Open Access

Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing

Abigail L. Lind[1] and Katherine S. Pollard[1,2,3,4,5*]

Check for updates

Short & biased history of metagenomics

- **discoveries!**

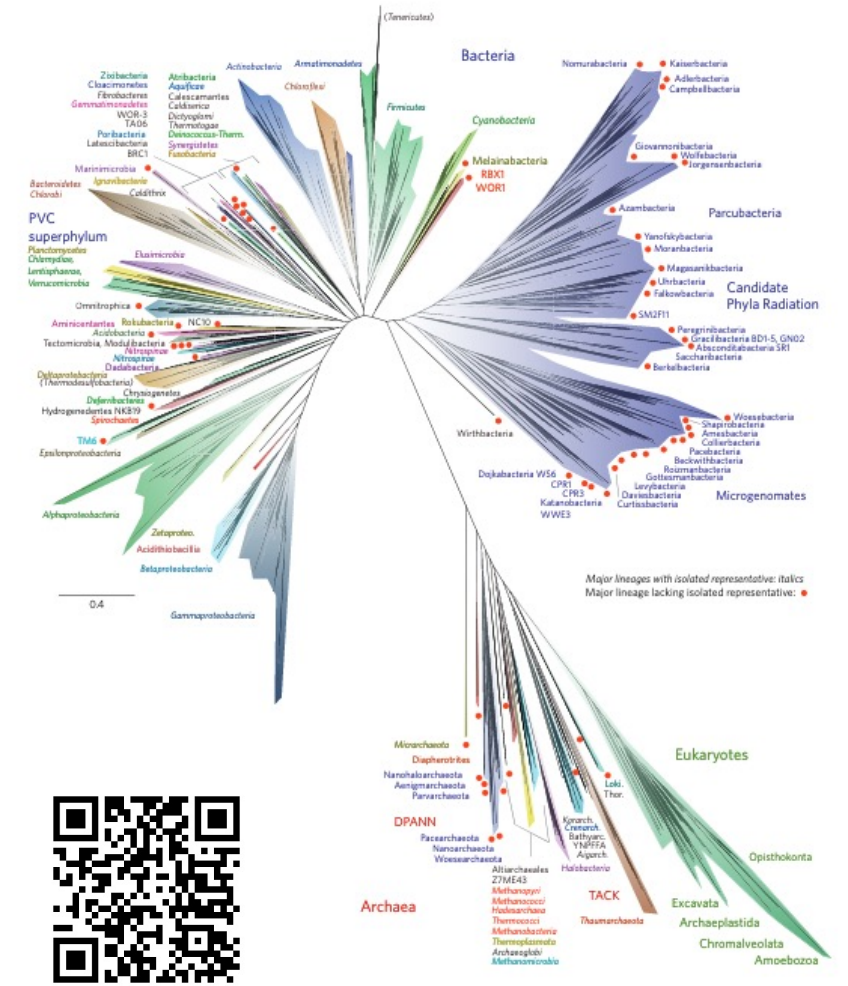
CellPress

Cell Host & Microbe
Article

The *Prevotella copri* Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations

Adrian Tett,[1,*] Kun D. Huang,[1,2] Francesco Asnicar,[1] Hannah Fehlner-Peach,[3] Edoardo Pasolli,[1,19] Nicolai Karcher,[1] Federica Armanini,[1] Paolo Manghi,[1] Kevin Bonham,[4,6] Moreno Zolfo,[1] Francesca De Filippis,[6,7] Cara Magnabosco,[8] Richard Bonneau,[8,9] John Lusingu,[10] John Amuasi,[11] Karl Reinhard,[12] Thomas Rattei,[13] Fredrik Boulund,[14] Lars Engstrand,[15] Albert Zink,[15] Maria Carmen Collado,[16] Dan R. Littman,[3] Daniel Eibach,[17,18] Danilo Ercolini,[6,7] Omar Rota-Stabelli,[2] Curtis Huttenhower,[4,5] Frank Maixner,[15] and Nicola Segata[1,20,*]

LETTERS NATURE MICROBIOLOGY DOI: 10.1038/NMICROBIOL.2016.48



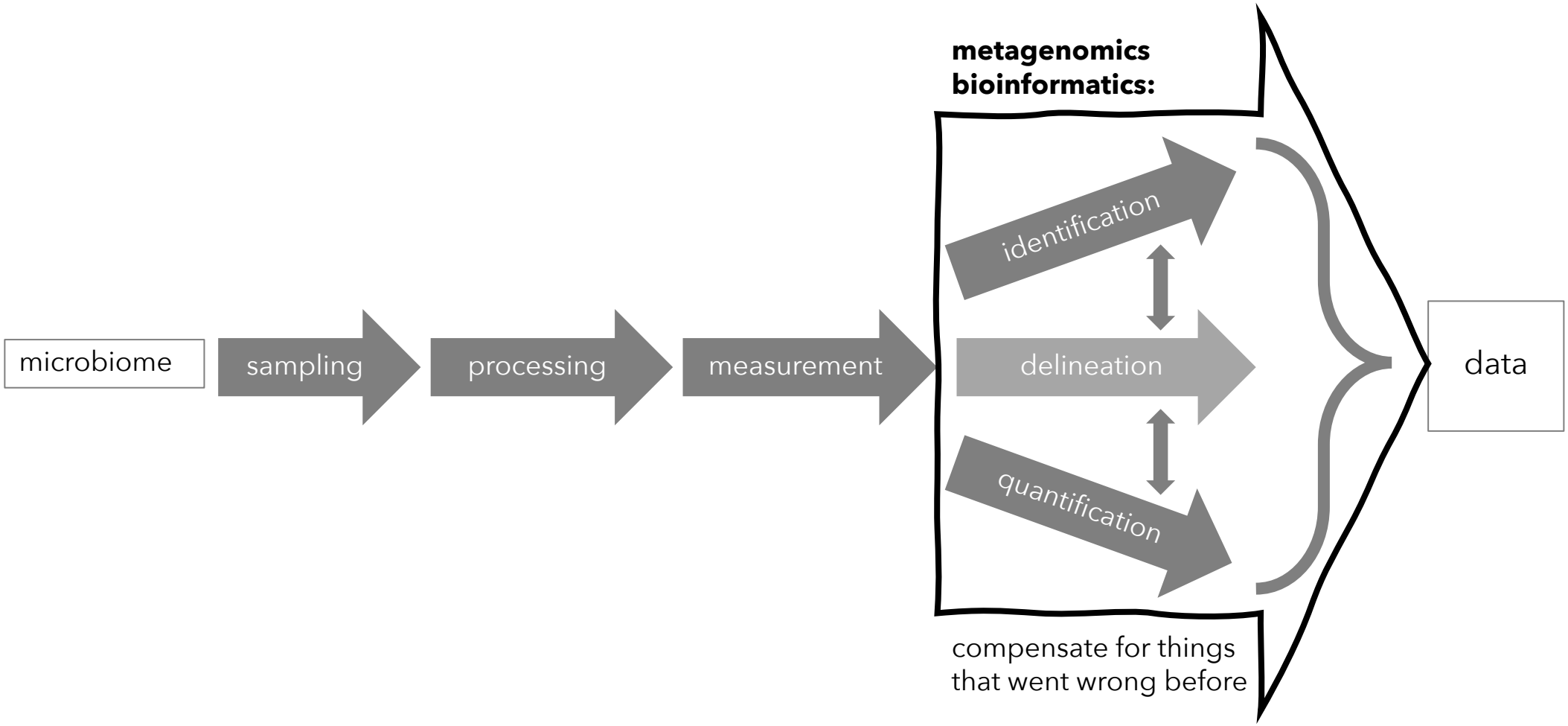
ARTICLES
https://doi.org/10.1038/s41564-020-00840-5
nature microbiology
Check for updates
OPEN

Genome-resolved metagenomics reveals site-specific diversity of episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems

Christine He[1], Ray Keren[2], Michael L. Whittaker[3,4], Ibrahim F. Farag[1], Jennifer A. Doudna[1,5,6,7,8], Jamie H. D. Cate[1,5,6,7] and Jillian F. Banfield[1,4,9]

Why do you study microbiomes?

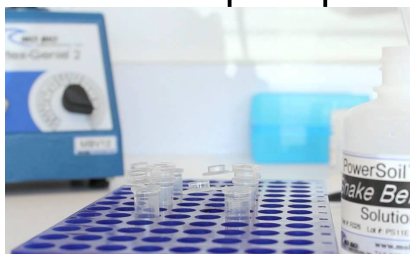
Omics paradigm



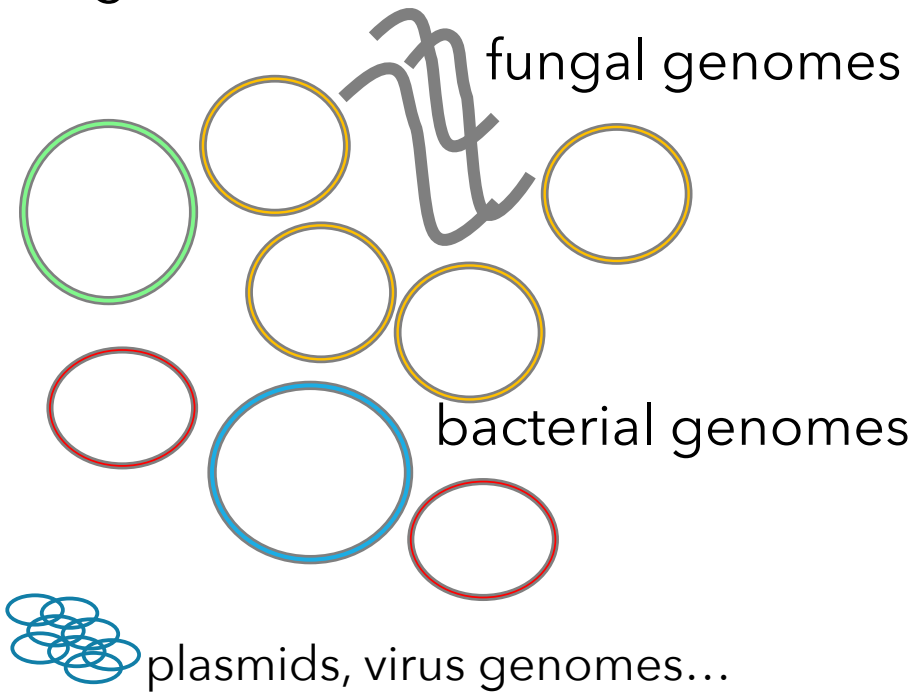
Measuring microbiomes: DNA based methods

sampling + preservation

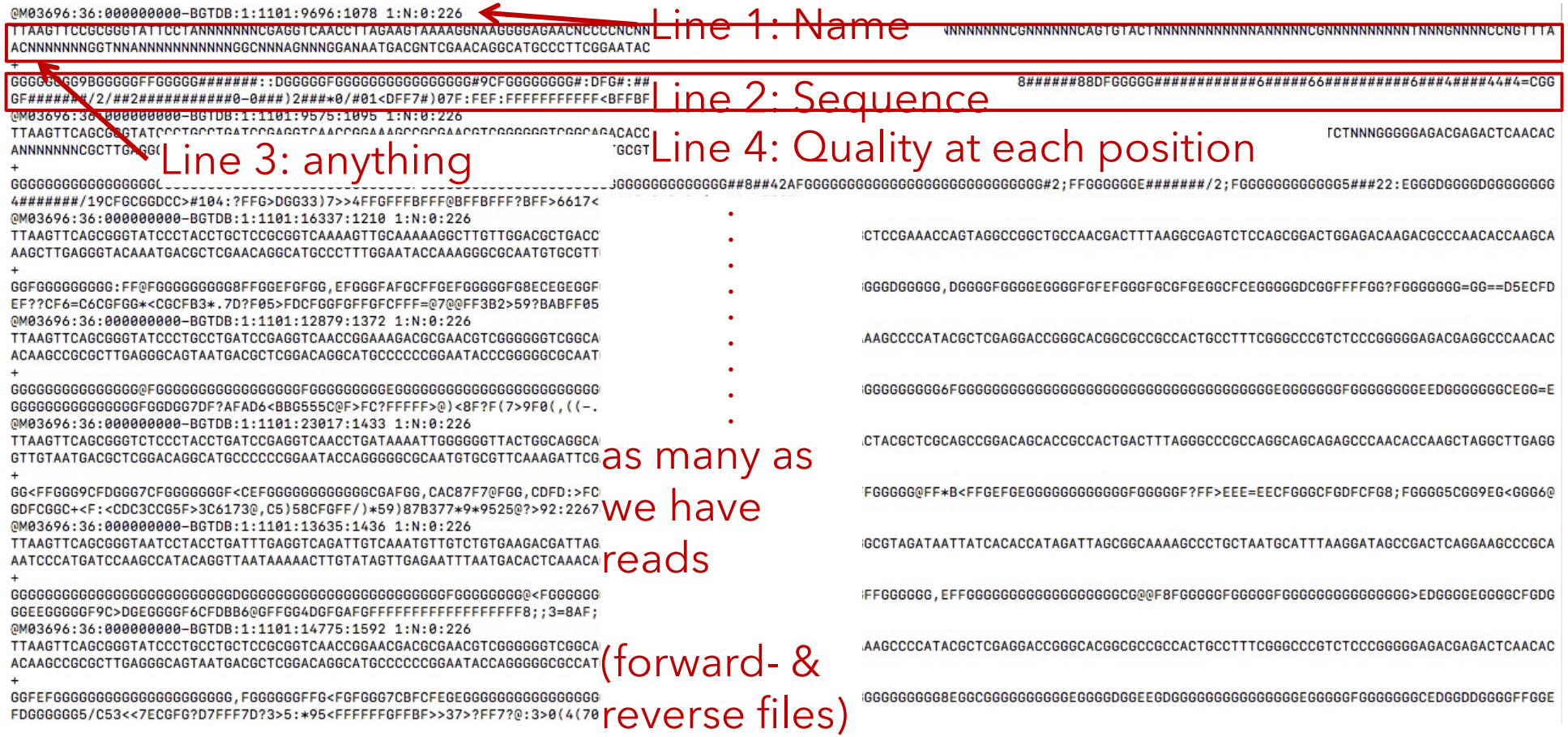
sample processing



the 'metagenome':



What does sequencing data look like?



Metagenomics overview

metagenomic DNA

sequencing

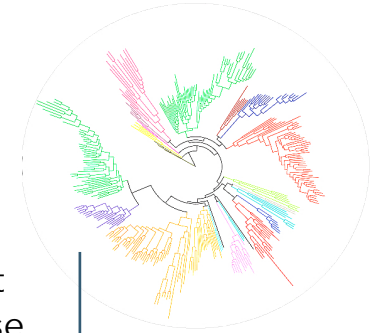
reads

2 quality filtering

high quality reads

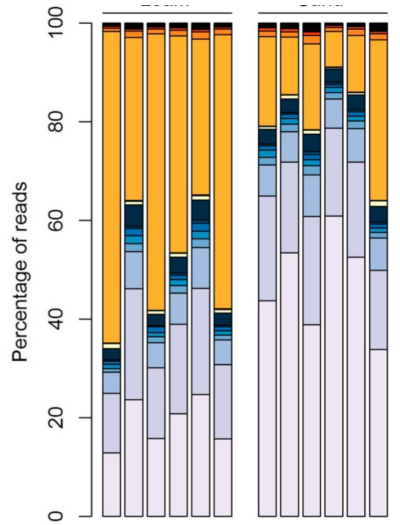
3 alignment against taxonomic database

4



taxonomic profiles

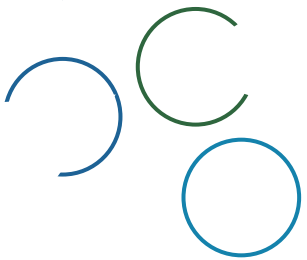
9



assembly of (partial) genomes

5

8 genome reconstruction



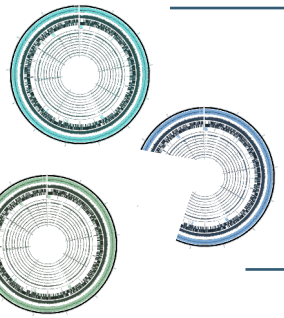
6 annotate genes

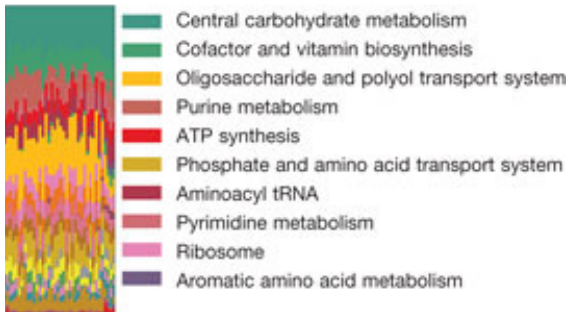
7 gene functions databases

count reads that align with genomes/genes

7

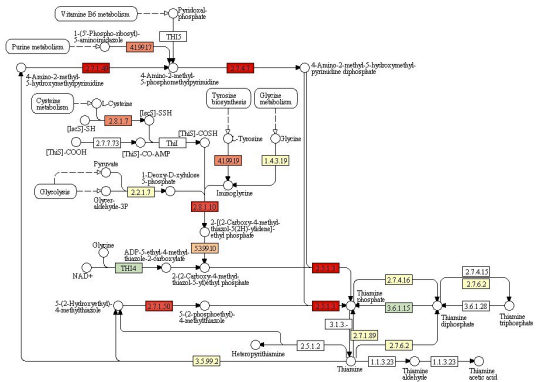
11





functional profiles

e.g. metabolism

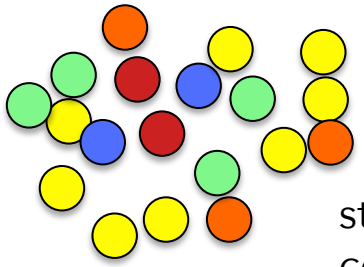


contextualize information

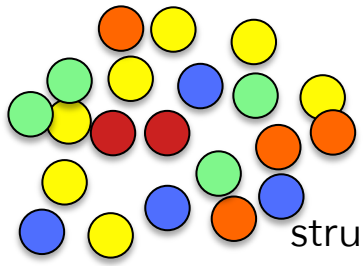
10

Metagenomics (+ other omics) pipeline

metabarcoding:



structure of community A



structure of community B

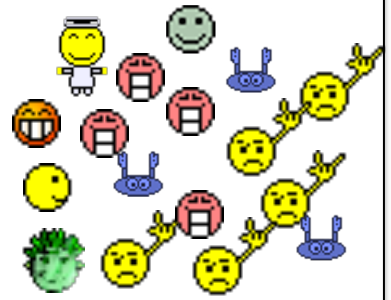
metagenomics:

potential of community A

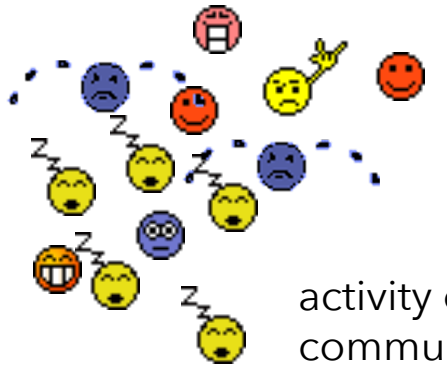
potential of community B

functional omics:

activity of community A

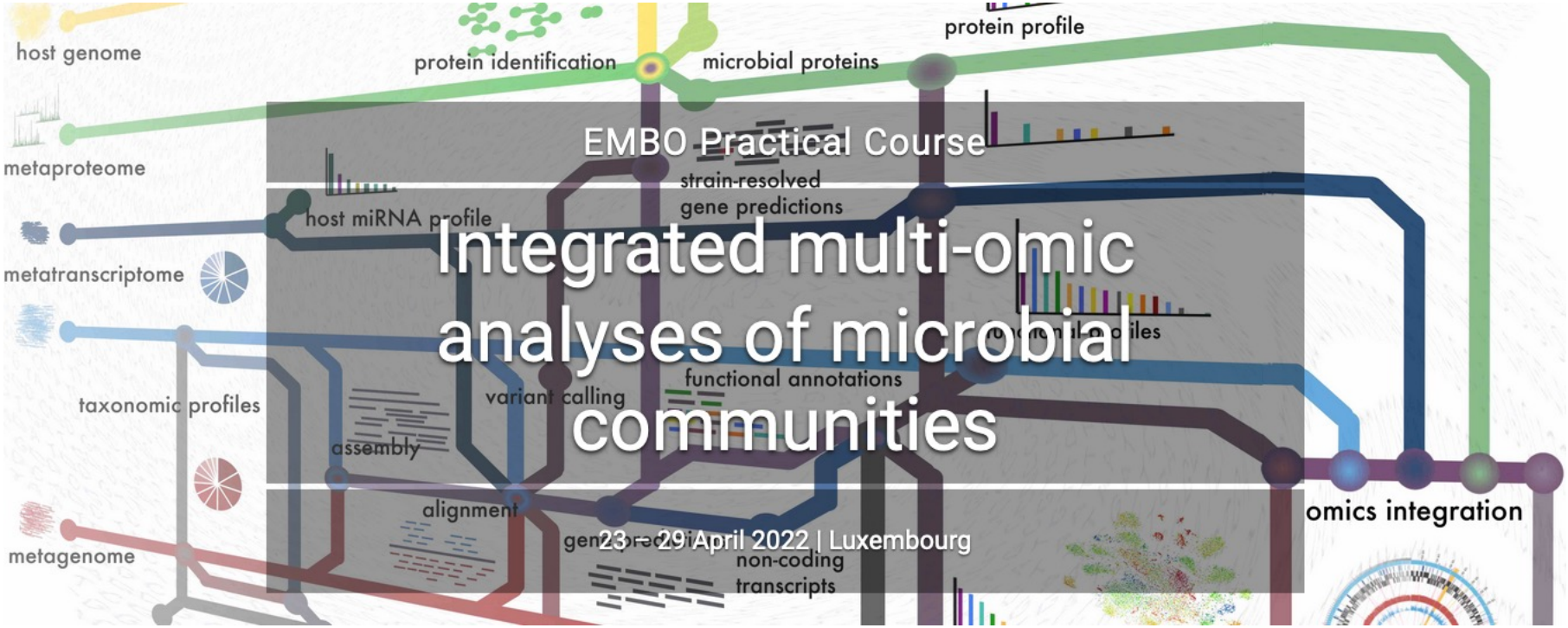


activity of populations within communities



activity of community B

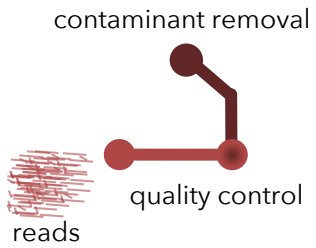
Advertisement Break



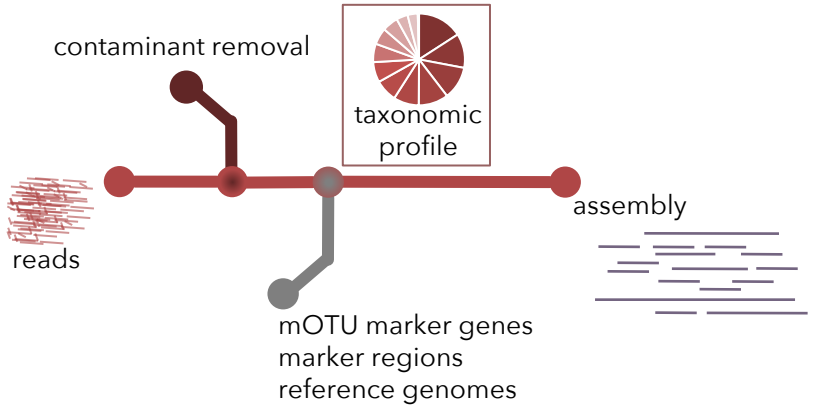
- 23rd-29th April, Luxembourg
- application deadline: 24th February

EMBO

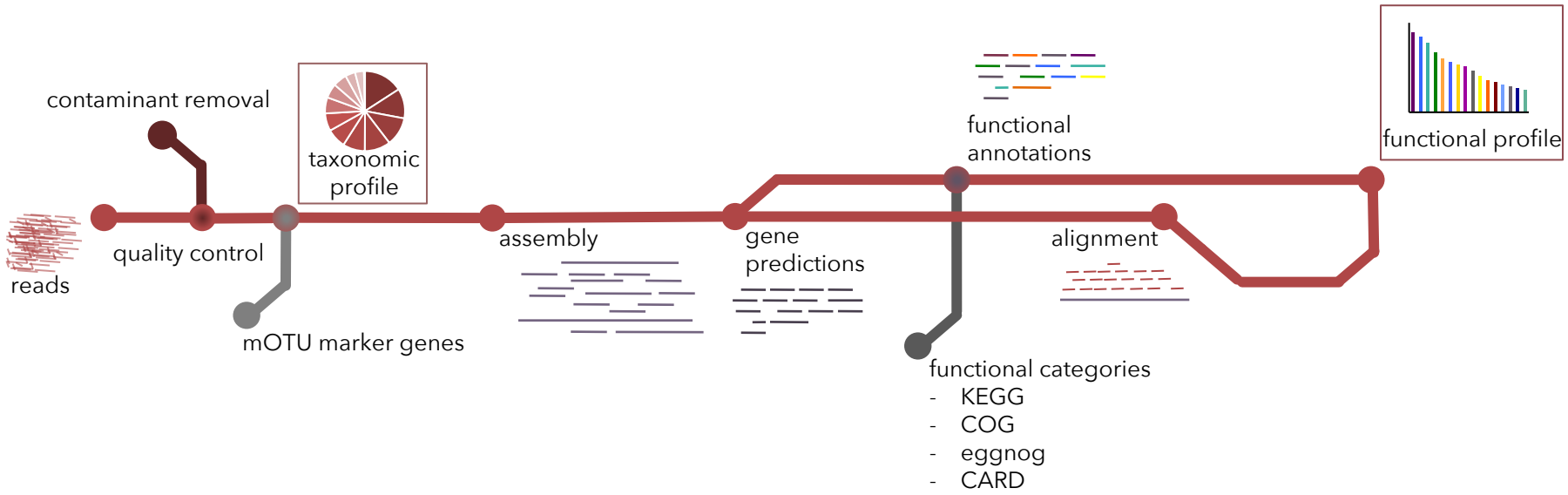
Metagenomics (+ other omics) pipeline



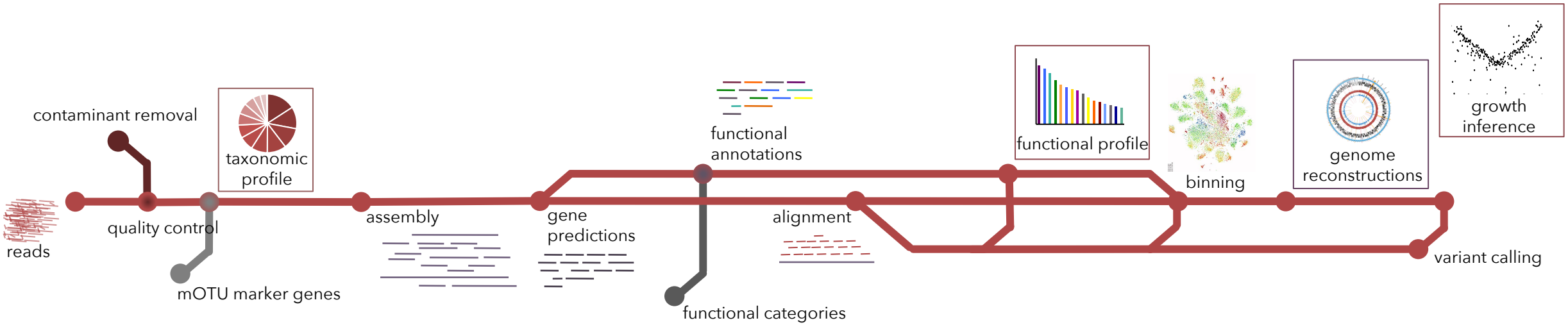
Metagenomics (+ other omics) pipeline



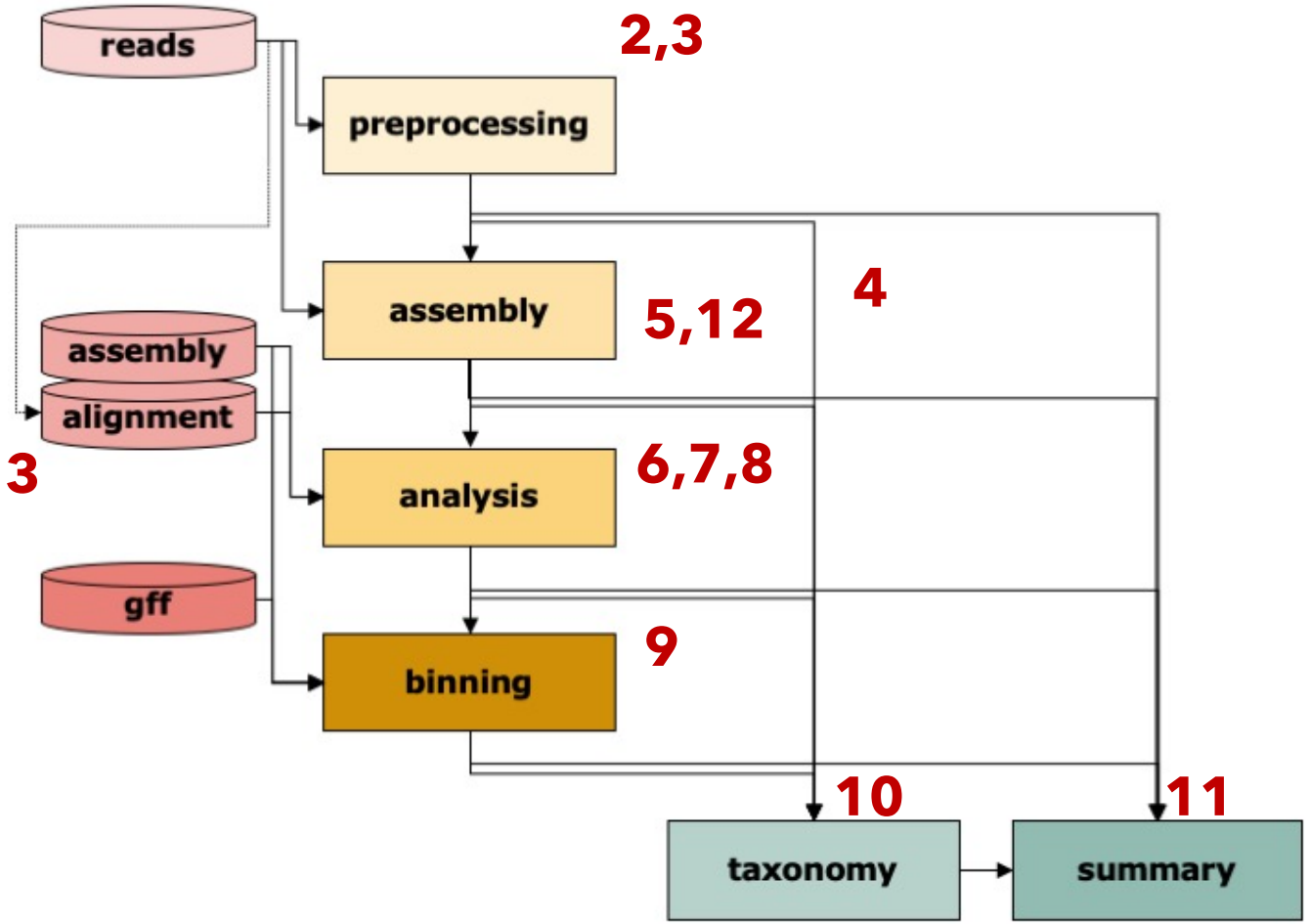
Metagenomics (+ other omics) pipeline



Metagenomics (+ other omics) pipeline



Metagenomics (+ other omics) pipeline



Thanks for your attention!

a.u.s.heintzbuschart@uva.nl

SP C2.205

github.com/a-h-b

twitter.com/_a_h_b_

University of Amsterdam
Life Sciences