PLEASE SIT ON EVERY OTHER ROW

BACTERIAL GENOMICS
MBDP-105

Jenni Hultman, Antti Karkman, David Fewer, Tania Shishido, Endrews Delbaje

LEARNING GOALS

- Foundational skills to work with bacterial genome data
- Familiarity and practice with bioinformatics tools
- Perspective and confidence to apply these skills in your own work
- Empower you to ask and answer the questions you have of your own data

THIS COURSE

- Hands-on
- Materials available during and after the course
  - Github
- Mix of lectures, tutorials and practice. Schedule might and will change
- Ask questions
- Learn from each other as well as instructors

IF YOU ARE SICK

- Stay at home
- To get the credits, you need to write a report on the tasks of the day you were away

GENOMICS

- Area within genetics that focuses in sequencing and analysis of an organism's genome

- Bacterial genomics is **a scientific discipline that concerns the genome, encompassing the entire hereditary information, of bacteria**.
- First (microbial) genome 1977
  - Bacteriophage ϕX174

- Today (27.3.22) 393 904 curated bacterial and archeal genomes available at NCBI

WHY BACTETERIAL GENOMICS?

- Insight into bacterial evolution and diversity beyond single gene or protein
- New application for biotechnology
- New approaches to treatment and control of pathogenic or otherwise harmful bacteria
  - outbreaks of bacterial infections
- Focus of this course in bacterial genomes
  - What is different in fungi and eukaryotes overall?

nature microbiology

ANALYSIS

https://doi.org/10.1038/s41564-022-01079-y

OPEN

Populat

livestoc

nature COMMUNICATIONS

Check for updates

ulosis

nature COMMUNICATIONS

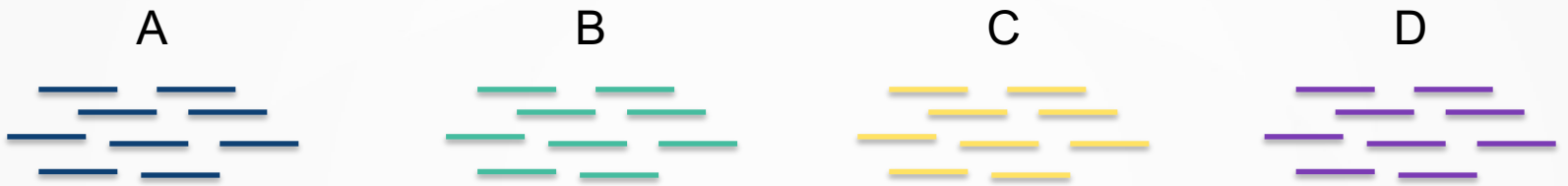
ARTICLE

https://doi.org/10.1038/s41467-021-25462-1 OPEN

Check for updates

Dynamics of the compartmentalized *Streptomyces* chromosome during metabolic differentiation

TWO APPROACHES TO BACTERIAL GENOMICS



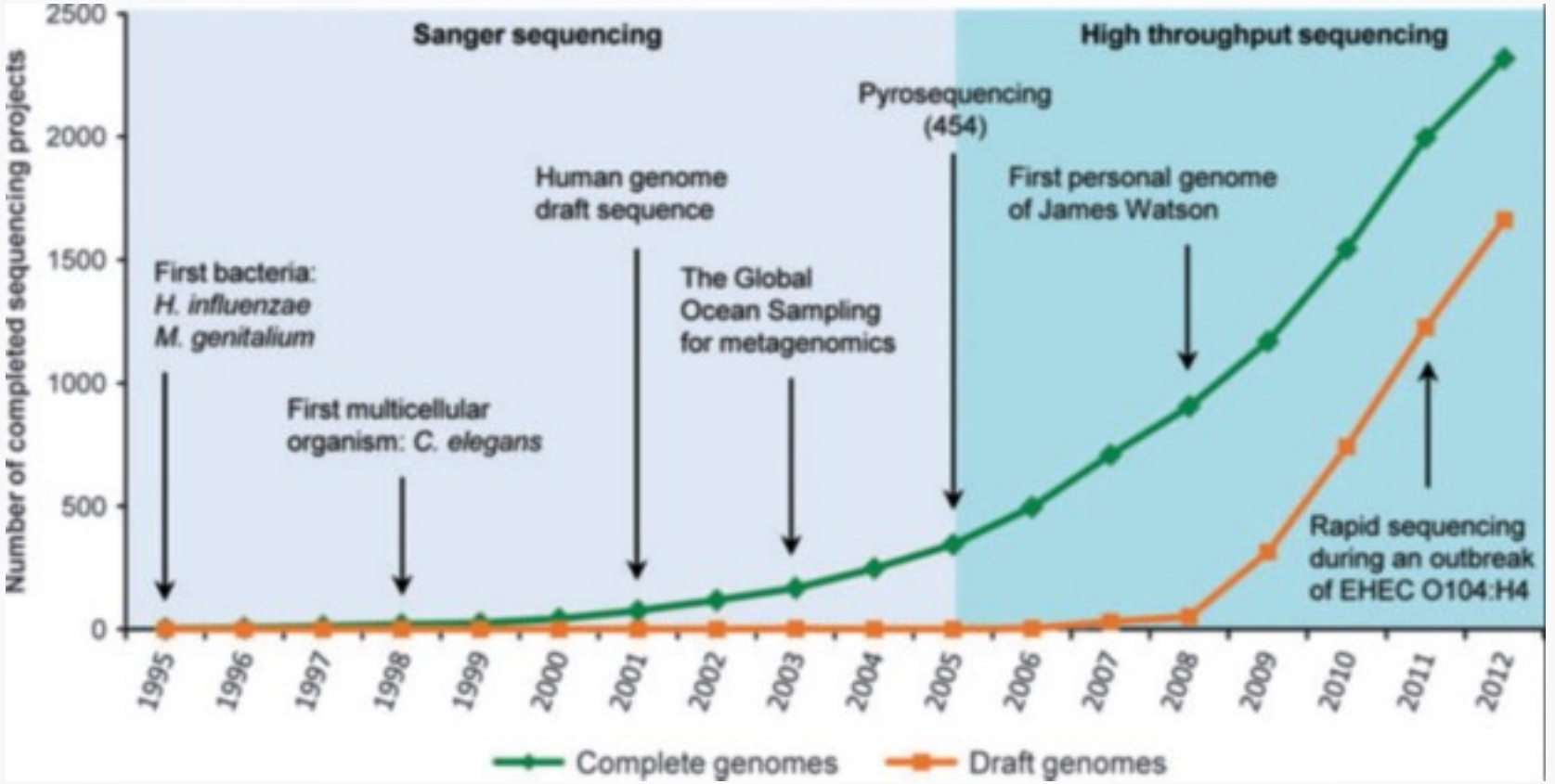
Assembly-based

1. De-novo assemble reads into a genome sequence
2. Annotate genome
3. Cluster genes and compare between each genome

Variant-based

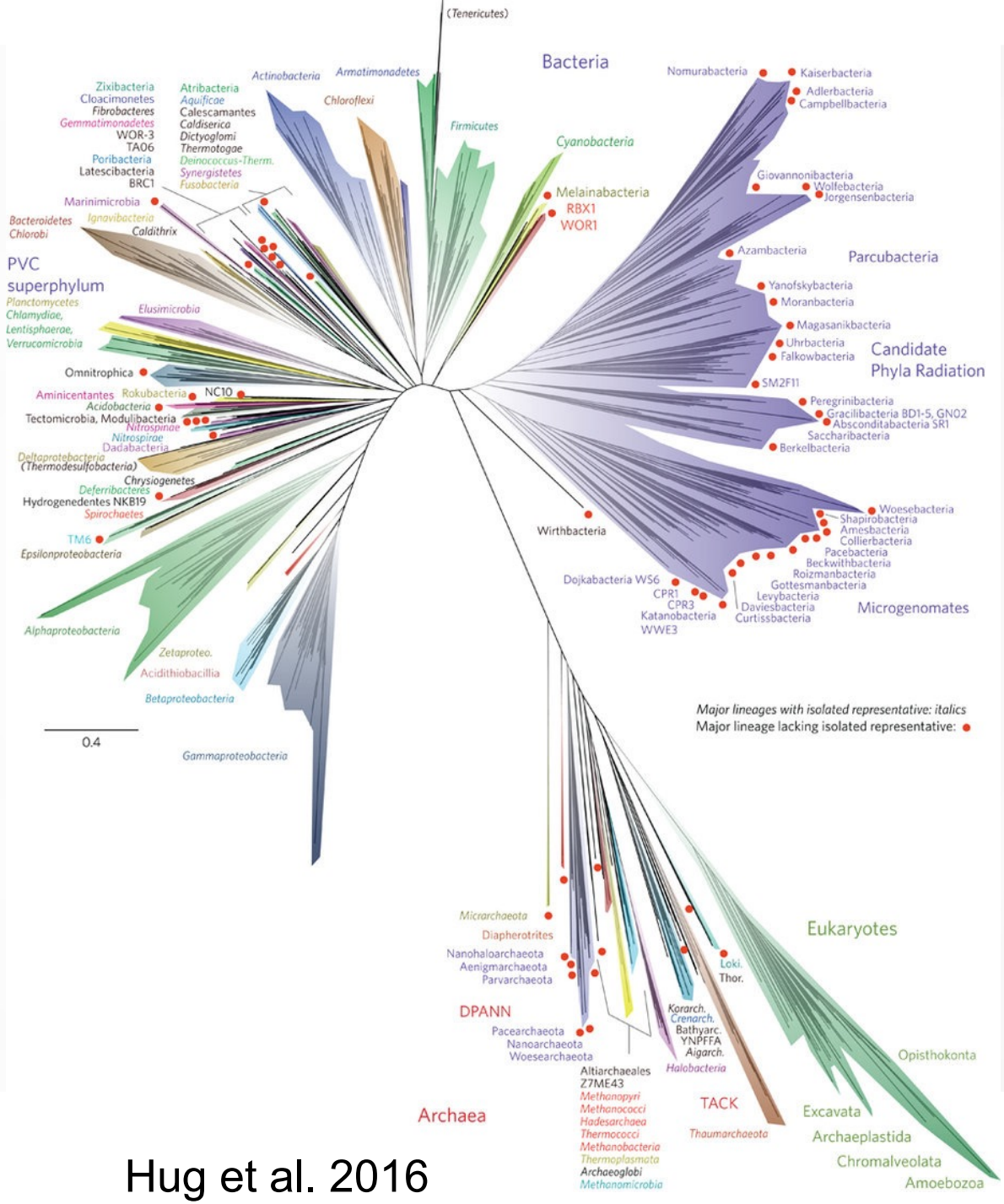
1. Compare read to a reference genome assembly
2. Directly compare variants between each genome

COMPLETE GENOME VS DRAFT GENOME


Sanger sequencing
High throughput sequencing
Number of completed sequencing projects
2500
2000
1500
1000
500
0
First bacteria:
H. influenzae
M. genitalium
First multicellular
organism: C. elegans
Human genome
draft sequence
The Global
Ocean Sampling
for metagenomics
Pyrosequencing
(454)
First personal genome
of James Watson
Rapid sequencing
during an outbreak
of EHEC O104:H4
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
Complete genomes
Draft genomes

OMICS

- Genomics
- Transcriptomics
- Proteomics
- Metabolomics



Hug et al. 2016

OMICS

- Genomics
- Transcriptomics
- Proteomics
- Metabolomics

All omics need good reference genomes/databases

Hug et al. 2016

GROUP WORK

Group work will be done in 4 groups

- Written report, return by 19.4.2022
- Presentation on Friday

1. Describe what was done
2. Compare different approaches and the results (assembly approaches)
3. Present pangenome
4. Present secondary metabolism

Report: Article format, best approaches, not comparison

Data analyzed during the course will be published in peer reviewed journal and all students will be authors.

HOUSEKEEPING

- Unicafe Viikuna is still closed, lunch at Ladonlukko or elsewhere
- MBDP will provide coffee and pulla on Mon, Wed and Fri
- No eating in the lecture rooms
- Masks
- Extension cords

CSC GREETINGS

- You all have CSC account with 1000 billing units
  - But not project where to do more intensive computing
  - You can run out of billing units
    - `saldo`, should not be negative
- For this course we have a project MBDP_genomics, Jenni has added you
  - Accept rules at mycsc.fi (bell sign)
- Make sure that when you work with **real data** you have **a PI who has a project** with enough billing units and you are member of that project