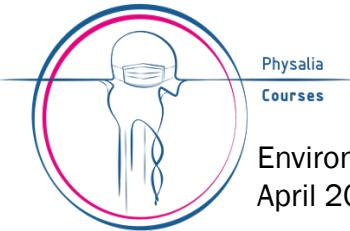


Environmental metagenomics

Introduction to metagenomics



Physalia
Courses

Environmental metagenomics
April 2021

Igor S. Pessi & Antti Karkman, University of Helsinki

The environment



“the complex of physical, chemical, and biotic factors (such as climate, soil, and living things) that act upon an organism or an ecological community and ultimately determine its form and survival”

- Merriam-Webster

The environment

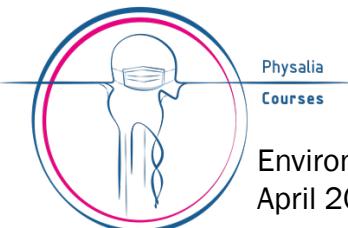


“the complex of physical, chemical, and biotic factors (such as climate, soil, and living things) that act upon an organism or an ecological community and ultimately determine its form and survival”

- Merriam-Webster

“The natural environment encompasses all living and non-living things occurring naturally, meaning in this case not artificial. The term is most often applied to the Earth or some parts of Earth. This environment encompasses the interaction of all living species, climate, weather and natural resources that affect human survival and economic activity. ”

- Wikipedia



Physalia
Courses

Environmental metagenomics
April 2021

Igor S. Pessi & Antti Karkman, University of Helsinki

Metagenomics

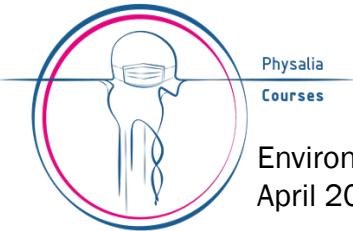
“The methodology has been made possible by advances in molecular biology and eukaryotic genomics, which have laid the groundwork for cloning and functional analysis of the collective genomes of soil microflora, which we term the metagenome of the soil.”

- Handelsman et al. 1998 (doi: /10.1016/S1074-5521(98)90108-9)

Jo Handelsman



<https://wid.wisc.edu/people/jo-handelsman/>

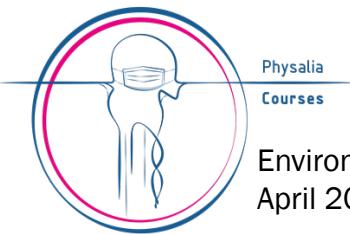


Metagenomics

“The methodology has been made possible by advances in molecular biology and eukaryotic genomics, which have laid the groundwork for cloning and functional analysis of the collective genomes of soil microflora, which we term the metagenome of the soil.”

- Handelsman et al. 1998 (doi: /10.1016/S1074-5521(98)90108-9)

Meta | gen | ome



Metagenomics

“The methodology has been made possible by advances in molecular biology and eukaryotic genomics, which have laid the groundwork for cloning and functional analysis of the collective genomes of soil microflora, which we term the metagenome of the soil.”

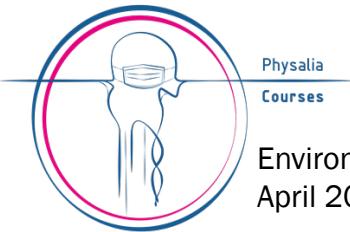
- Handelsman et al. 1998 (doi: /10.1016/S1074-5521(98)90108-9)

Meta | gen | ome

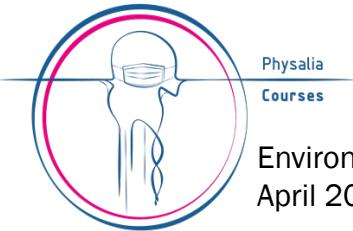
“Going beyond the genome”

- Matthew Schechter

(<https://merenlab.org/2020/07/27/history-of-metagenomics/>)



(The incomplete) history of metagenomics



Physalia
Courses

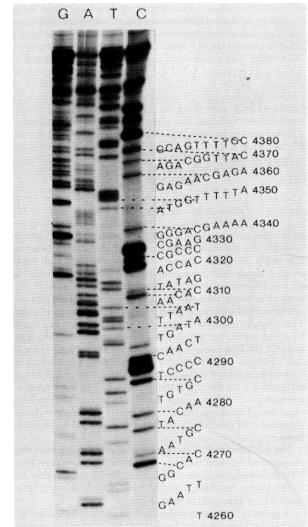
Environmental metagenomics
April 2021

Igor S. Pessi & Antti Karkman, University of Helsinki

(The incomplete) history of metagenomics

Chain termination:
Sanger sequencing

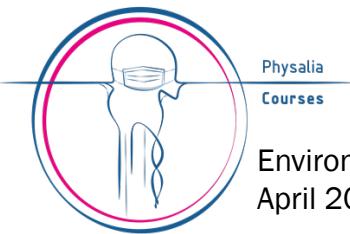
Proc. Natl. Acad. Sci. USA
Vol. 74, No. 12, pp. 5463–5467, December 1977
Biochemistry



DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage ϕ X174)

F. SANGER, S. NICKLEN, AND A. R. COULSON



Physalia
Courses

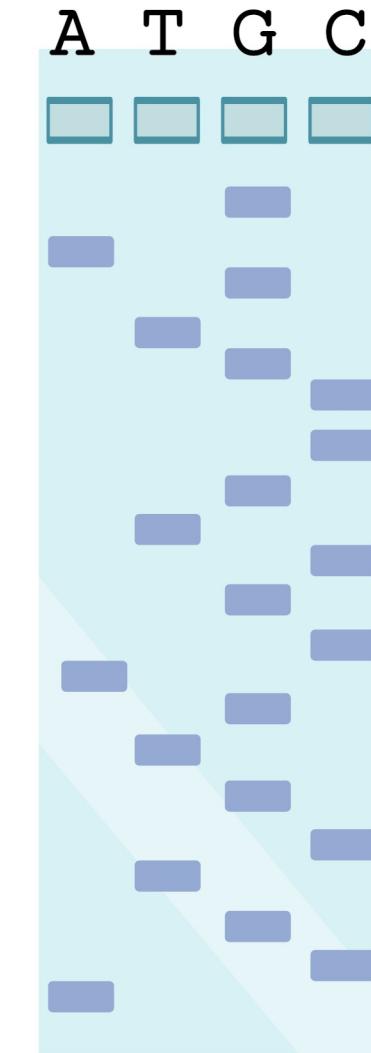
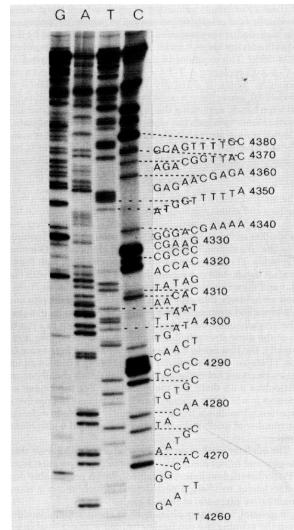
Environmental metagenomics
April 2021

Igor S. Pessi & Antti Karkman, University of Helsinki

(The incomplete) history of metagenomics

Chain termination:
Sanger sequencing

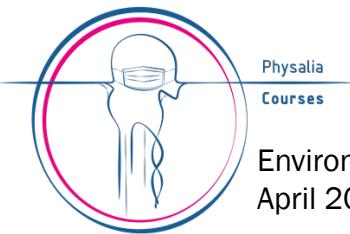
Proc. Natl. Acad. Sci. USA
Vol. 74, No. 12, pp. 5463–5467, December 1977
Biochemistry



DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage ϕ X174)

F. SANGER, S. NICKLEN, AND A. R. COULSON

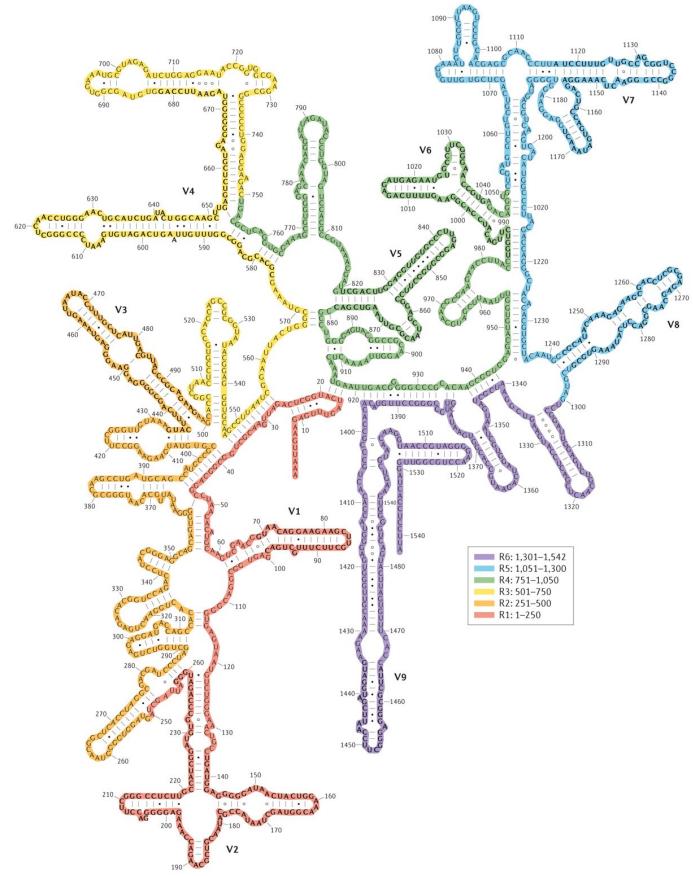


Physalia
Courses

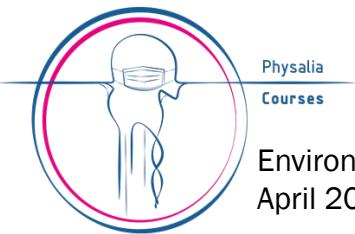
Environmental metagenomics
April 2021

Igor S. Pessi & Antti Karkman, University of Helsinki

(The incomplete) history of metagenomics



Nature Reviews | Microbiology



Environmental metagenomics

April 2021

Igor S. Pessi & Antti Karkman, University of Helsinki

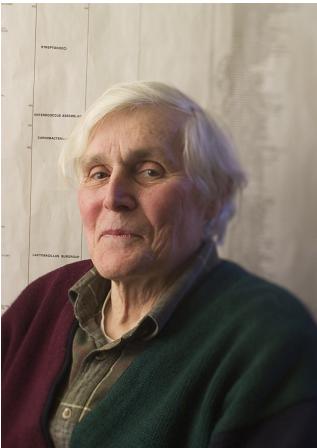
(The incomplete) history of metagenomics

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 11, pp. 5088–5090, November 1977
Evolution

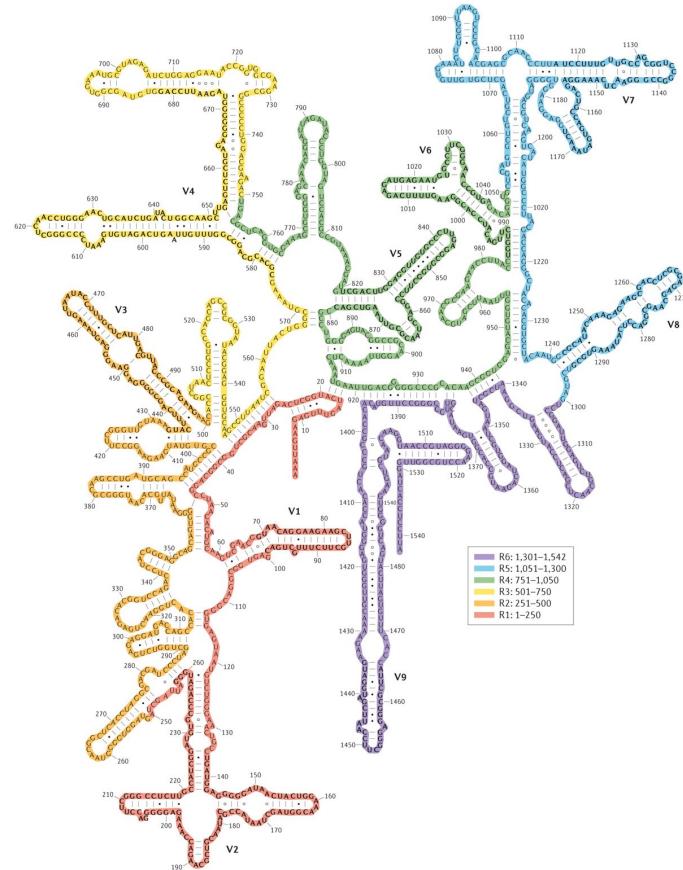
Phylogenetic structure of the prokaryotic domain: The primary kingdoms

(archaeabacteria/eubacteria/urkaryote/16S ribosomal RNA/molecular phylogeny)

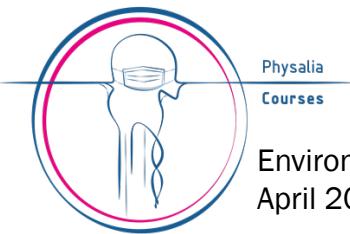
CARL R. WOESE AND GEORGE E. FOX*



By Don Hamerman - Institute for Genomic Biology, University of Illinois at Urbana-Champaign, CC BY 3.0,



Nature Reviews | Microbiology



(The incomplete) history of metagenomics

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 11, pp. 5088–5090, November 1977
Evolution

Phylogenetic structure of the prokaryotic domain: The primary kingdoms

(archaeabacteria/eubacteria/urkaryote/16S ribosomal RNA/molecular phylogeny)

CARL R. WOEESE AND GEORGE E. FOX*

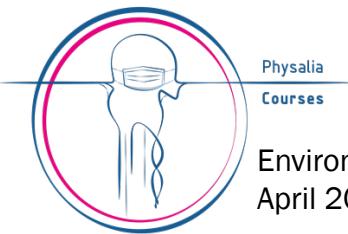


By Don Hamerman - Institute for Genomic Biology, University of Illinois at Urbana-Champaign, CC BY 3.0,

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. <i>Saccharomyces cerevisiae</i> , 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. <i>Lemna minor</i> , 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. <i>Escherichia coli</i>	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. <i>Chlorobium vibrioforme</i>	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. <i>Bacillus firmus</i>	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. <i>Corynebacterium diphtheriae</i>	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. <i>Aphanocapsa</i> 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (<i>Lemna</i>)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. <i>Methanobacterium thermoautotrophicum</i>	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. <i>M. ruminantium</i> strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. <i>Methanobacterium</i> sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. <i>Methanosaarcina barkeri</i>	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A , N_B , and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).



(The incomplete) history of metagenomics

RESEARCH ARTICLE

Science, 1985

Enzymatic Amplification of β -Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia

Randall K. Saiki, Stephen Scharf, Fred Faloona, Kary B. Mullis

Glenn T. Horn, Henry A. Erlich, Norman Arnheim

MICROBIOLOGICAL REVIEWS, June 1987, p. 221–271
0146-0749/87/020221-51\$02.00/0
Copyright © 1987, American Society for Microbiology

Bacterial Evolution

CARL R. WOESE

Department of Microbiology, University of Illinois, Urbana, Illinois 61801

Vol. 51, No. 2

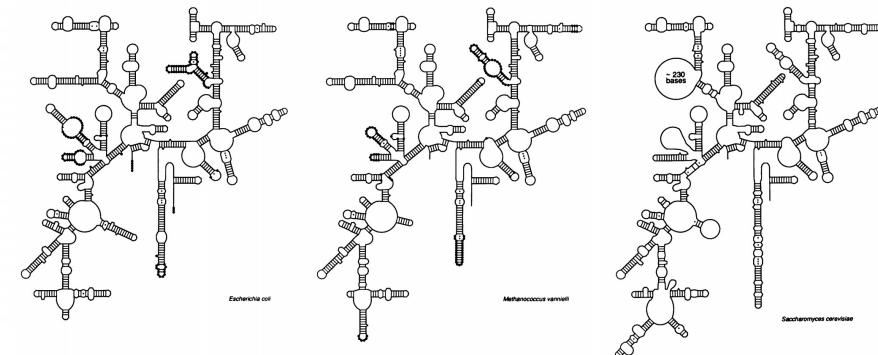
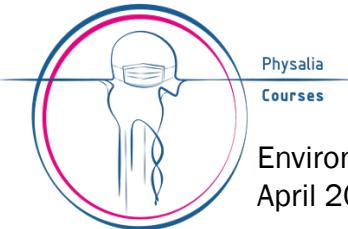


FIG. 6. Representative small-subunit rRNA secondary structures for the three primary kingdoms (75). Dots in the *Escherichia coli* or *Methanococcus vannielii* structures identify the positions or locales where eubacteria and archaeabacteria characteristically differ.



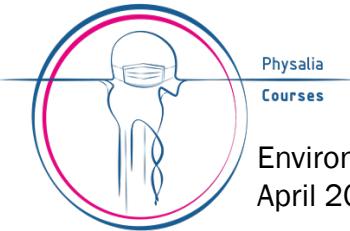
Environmental metagenomics
April 2021

Igor S. Pessi & Antti Karkman, University of Helsinki

(The incomplete) history of metagenomics

High-throughput Sanger sequencing

Dye-terminators and capillary electrophoresis



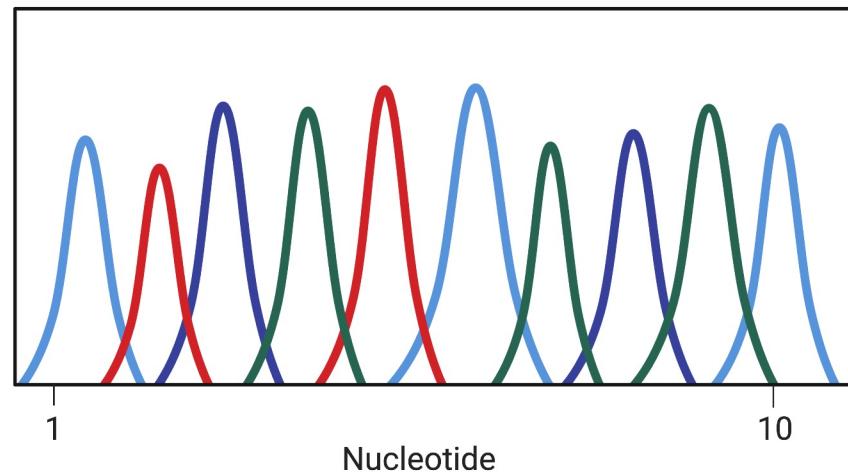
Environmental metagenomics
April 2021

Igor S. Pessi & Antti Karkman, University of Helsinki

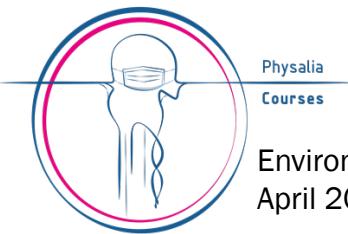
(The incomplete) history of metagenomics

High-throughput Sanger sequencing

Dye-terminators and capillary electrophoresis



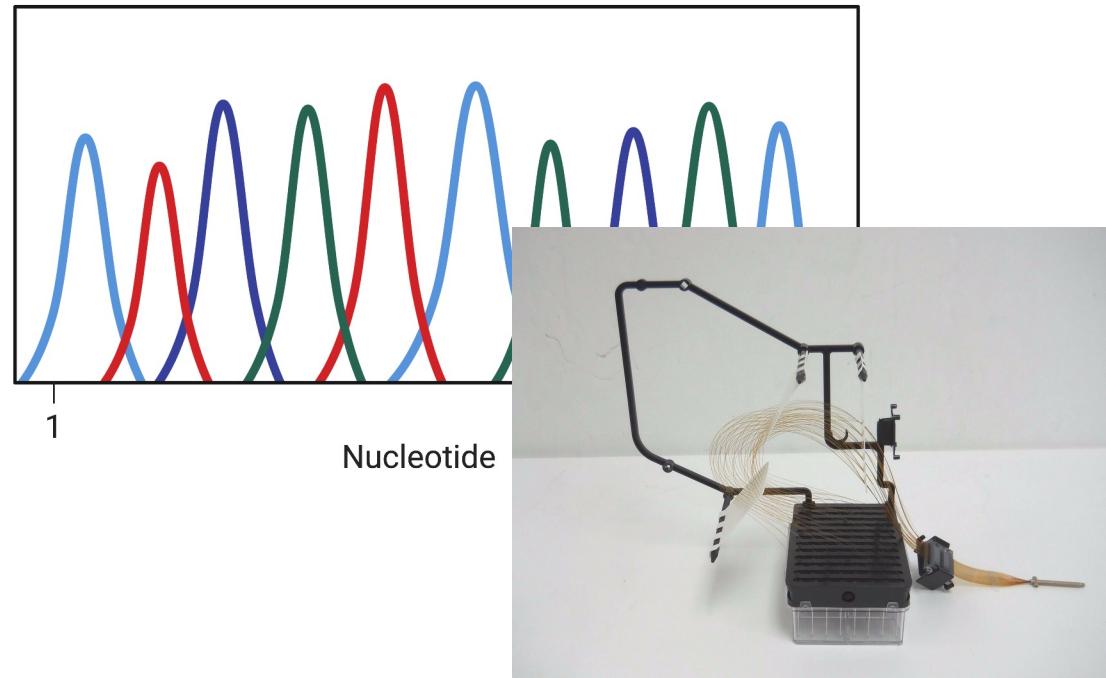
life
technologies™



(The incomplete) history of metagenomics

High-throughput Sanger sequencing

Dye-terminators and capillary electrophoresis



48/96 -capillars
Read length < 900 bp
< 90 Kbp / run

life
technologies™

(The incomplete) history of metagenomics

JOURNAL OF BACTERIOLOGY, Feb. 1996, p. 591–599
0021-9193/96/\$04.00+0
Copyright © 1996, American Society for Microbiology

Characterization of Uncultivated Prokaryotes: Isolation and Analysis of a 40-Kilobase-Pair Genome Fragment from a Planktonic Marine Archaeon

JEFFEREY L. STEIN,^{1*} TERENCE L. MARSH,² KE YING WU,³ HIROAKI SHIZUYA,⁴ AND EDWARD F. DELONG^{3*}

Recombinant BioCatalysis, Inc., La Jolla, California 92037¹; Microbiology Department, University of Illinois, Urbana, Illinois 61801²; Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, California 93106³; and Division of Biology, California Institute of Technology, Pasadena, California 91125⁴

Vol. 178, No. 3

Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products

Jo Handelsman¹, Michelle R Rondon¹, Sean F Brady², Jon Clardy² and Robert M Goodman¹

RESEARCH ARTICLE

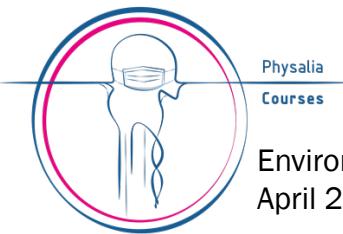
Environmental Genome Shotgun Sequencing of the Sargasso Sea

J. Craig Venter^{1,*}, Karin Remington¹, John F. Heidelberg³, Aaron L. Halpern², Doug Rusch², Jonathan A. Eisen³, Dongying W...

* See all authors and affiliations

Science 02 Apr 2004:
Vol. 304, Issue 5667, pp. 66-74
DOI: 10.1126/science.1093857

100 Mbp – 1 Gbp



Environmental metagenomics
April 2021

OPEN ACCESS Freely available online

PLOS BIOLOGY

The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific

Douglas B. Rusch^{1*}, Aaron L. Halpern¹, Granger Sutton¹, Karla B. Heidelberg^{1,2}, Shannon Williamson¹, Shibu Yooseph¹, Dongying Wu^{1,3}, Jonathan A. Eisen^{1,3}, Jeff M. Hoffman¹, Karin Remington^{1,4}, Karen Beeson¹, Bao Tran¹, Hamilton Smith¹, Holly Baden-Tillson¹, Clare Stewart¹, Joyce Thorpe¹, Jason Freeman¹, Cynthia Andrews-Pfannkoch¹, Joseph E. Venter¹, Kelvin Li¹, Saul Kravitz¹, John F. Heidelberg^{1,2}, Terry Utterback¹, Yu-Hui Rogers¹, Luisa I. Falcón⁵, Valeria Souza⁵, Germán Bonilla-Rosso⁵, Luis E. Eguiarte⁵, David M. Karl⁶, Shubha Sathyendranath⁷, Trevor Platt⁷, Eldredge Bermingham⁸, Victor Gallardo⁹, Giselle Tamayo-Castillo¹⁰, Michael R. Ferrari¹¹, Robert L. Strausberg¹, Kenneth Nealson^{1,12}, Robert Friedman¹, Marvin Frazier¹, J. Craig Venter¹

1 Gbp – 10 Gbp

Igor S. Pessi & Antti Karkman, University of Helsinki

17

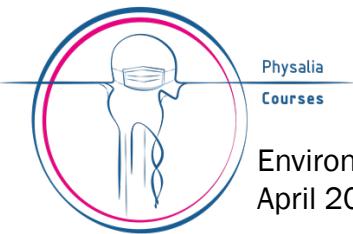
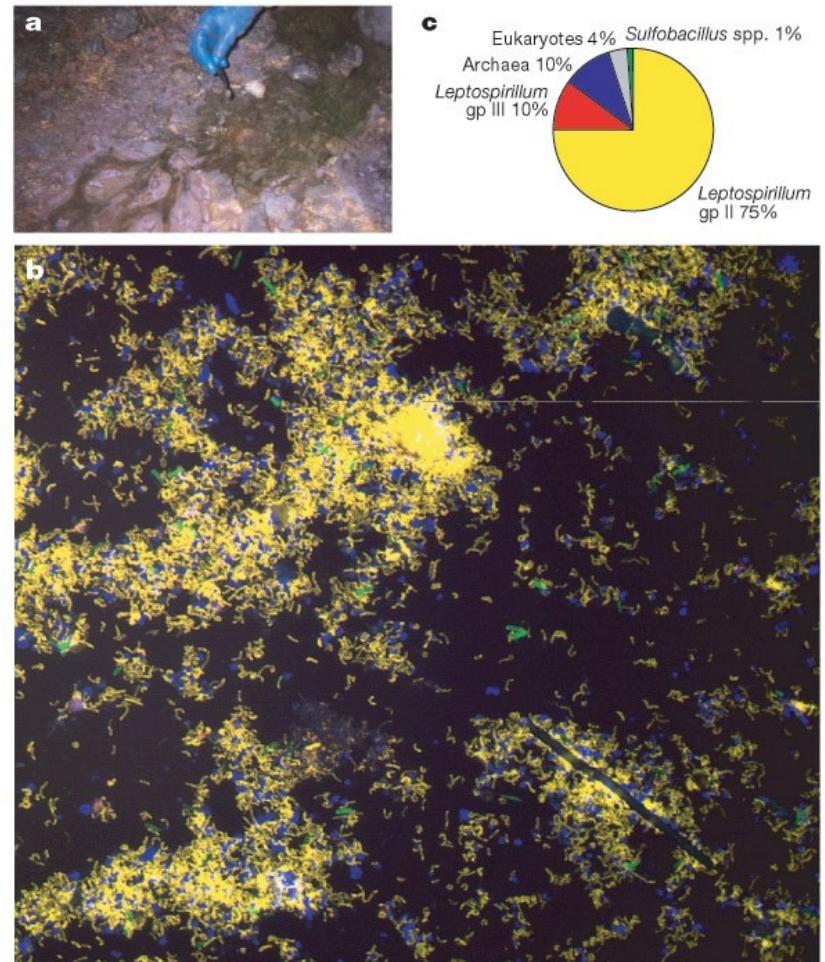
(The incomplete) history of metagenomics

Community structure and metabolism through reconstruction of microbial genomes from the environment

Gene W. Tyson, Jarrod Chapman, Philip Hugenholtz, Eric E. Allen, Rachna J. Ram, Paul M. Richardson, Victor V. Solovyev, Edward M. Rubin, Daniel S. Rokhsar & Jillian F. Banfield 

Nature 428, 37–43(2004) | [Cite this article](#)

Acid mine drainage
76.2 Mbp
2 MAGs



(The incomplete) history of metagenomics

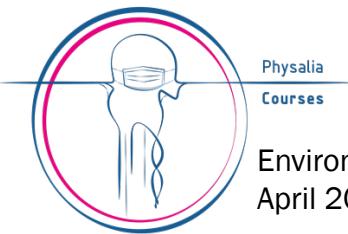
Sequencing by synthesis:

Pyrosequencing

Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies, Michael Egholm, [...] Jonathan M. Rothberg 

Nature **437**, 376–380(2005) | [Cite this article](#)



Physalia
Courses

Environmental metagenomics
April 2021

Igor S. Pessi & Antti Karkman, University of Helsinki

(The incomplete) history of metagenomics

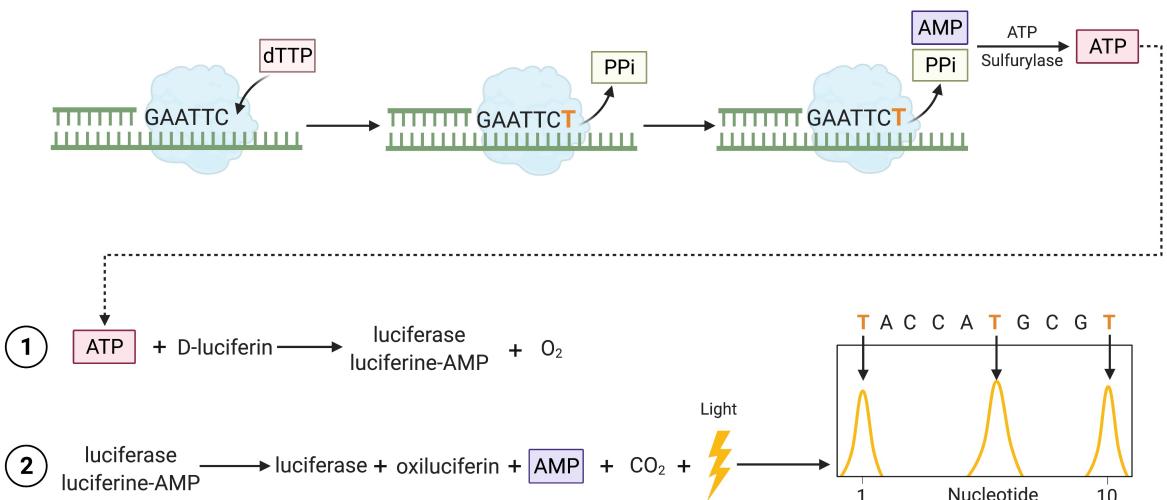
Sequencing by synthesis:

Pyrosequencing

Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies, Michael Egholm, [...] Jonathan M. Rothberg 

Nature 437, 376–380(2005) | Cite this article



Created with BioRender.com



(The incomplete) history of metagenomics

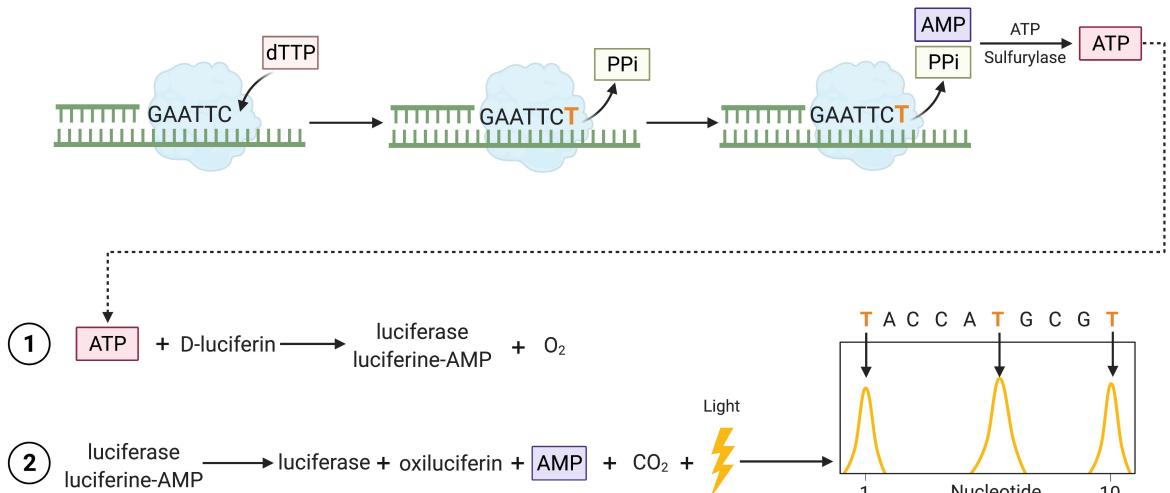
Sequencing by synthesis:

Pyrosequencing

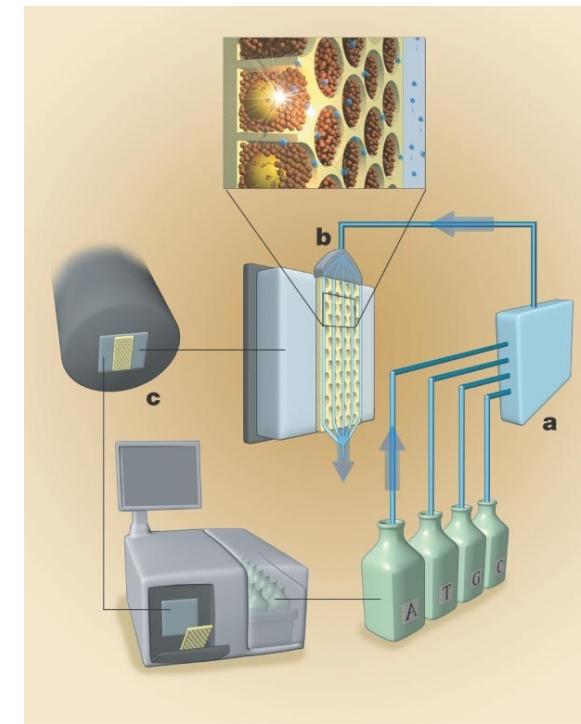
Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies, Michael Egholm, [...] Jonathan M. Rothberg 

Nature 437, 376–380(2005) | Cite this article



Created with BioRender.com



Margulies et al., 2005. Nature

Read length ~ 400–1000 bp
20 – 600 Mbp / run

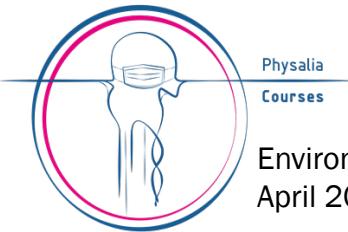
(The incomplete) history of metagenomics

Sequencing by synthesis:

Short-read sequencing



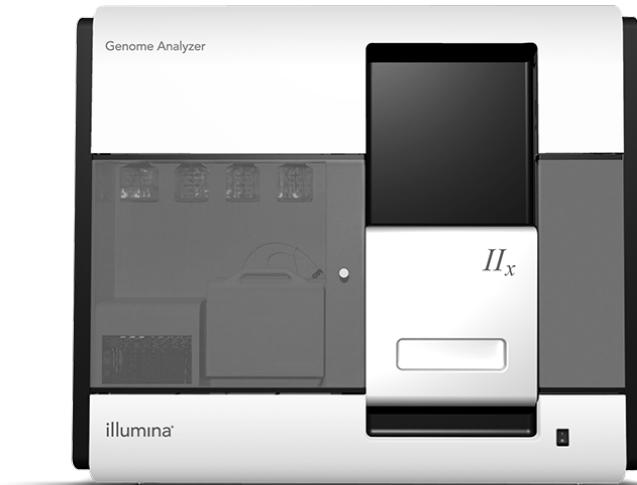
Illumina (Solexa)
Genome Analyzer, 2006
1 Gbp / run



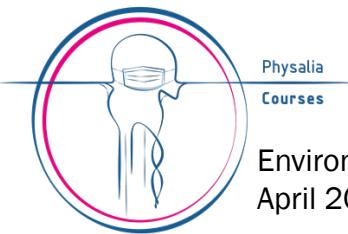
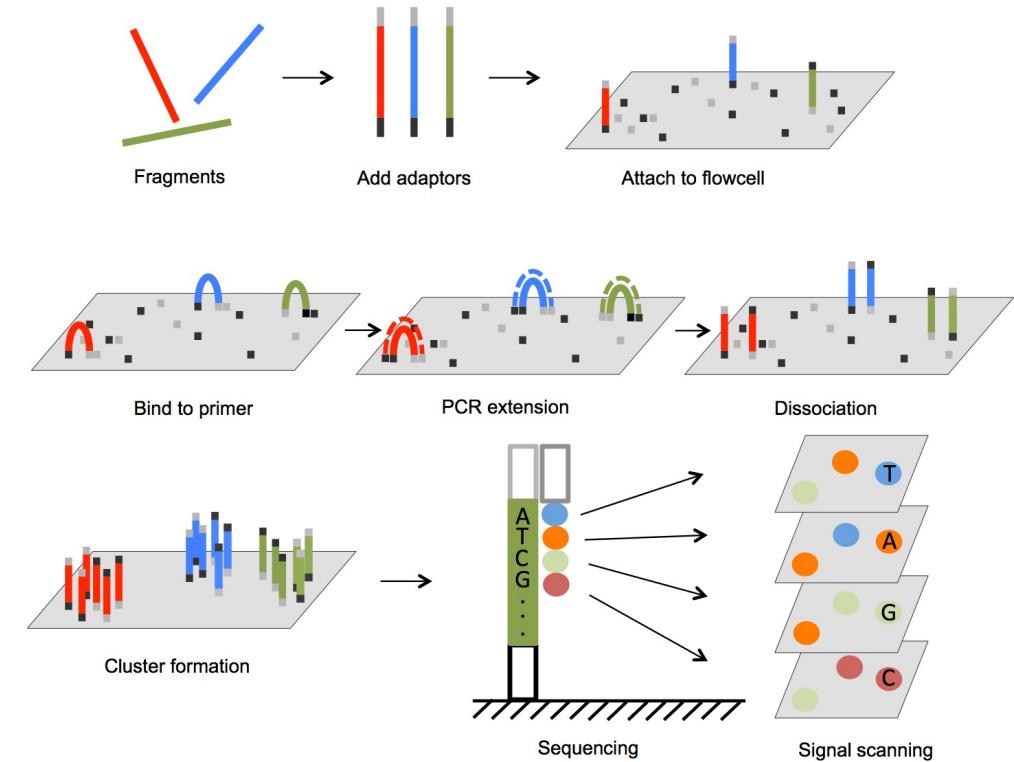
(The incomplete) history of metagenomics

Sequencing by synthesis:

Short-read sequencing



Illumina
Genome Analyzer IIx, 2008
20 Gbp / run (9.5 days)



Physalia
Courses

Environmental metagenomics
April 2021

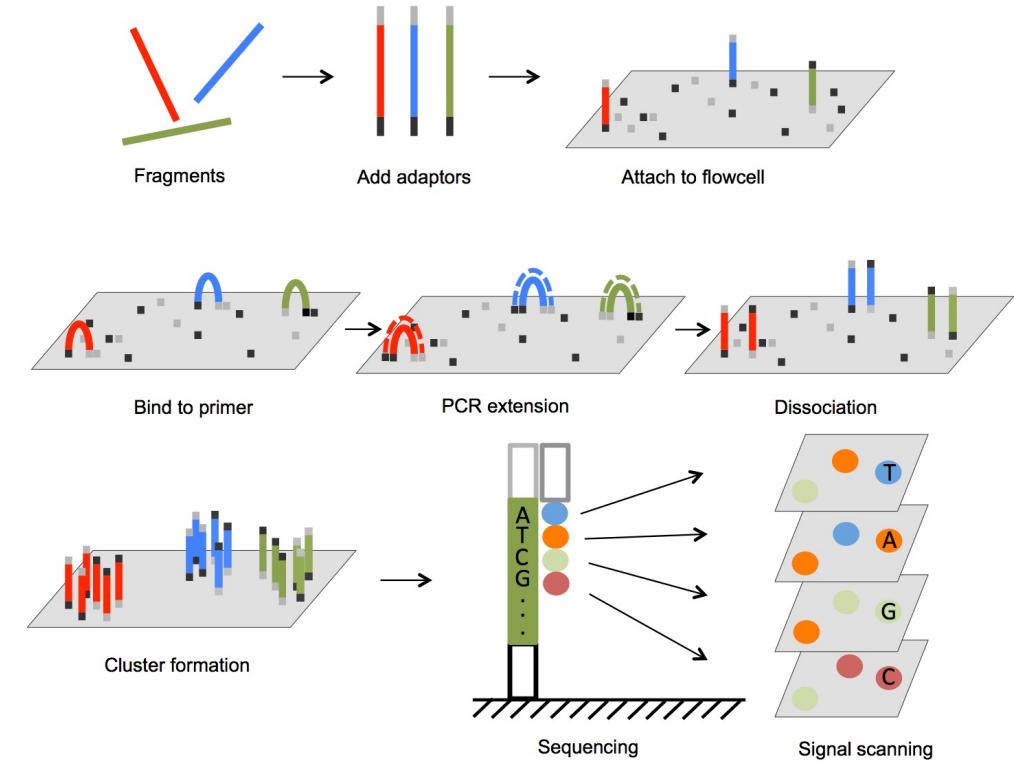
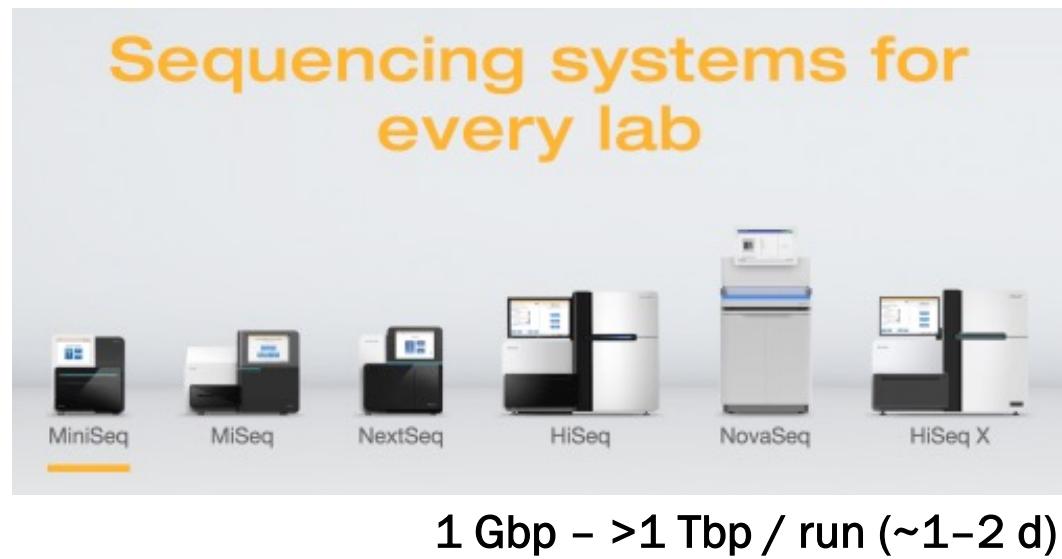
Igor S. Pessi & Antti Karkman, University of Helsinki

23

(The incomplete) history of metagenomics

Sequencing by synthesis:

Short-read sequencing



(The incomplete) history of metagenomics

Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes

Mads Albertsen, Philip Hugenholtz, Adam Skarshewski, Kåre L Nielsen, Gene W Tyson & Per H Nielsen



Nature Biotechnology 31, 533–538(2013) | [Cite this article](#)

Activated sludge
~90 Gbp
13 MAGs

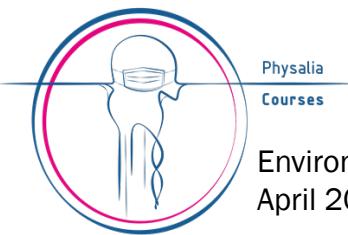
Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes

Tom O. Delmont , Christopher Quince, Alon Shaiber, Özcan C. Esen, Sonny TM Lee, Michael S. Rappé,
Sandra L. McLellan, Sebastian Lütke & A. Murat Eren

Nature Microbiology 3, 804–813(2018) | [Cite this article](#)

957 MAGs!
(using Anvi'o)

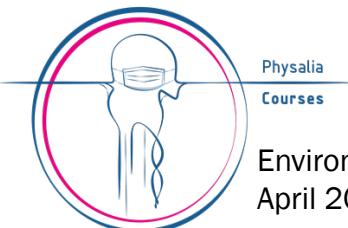
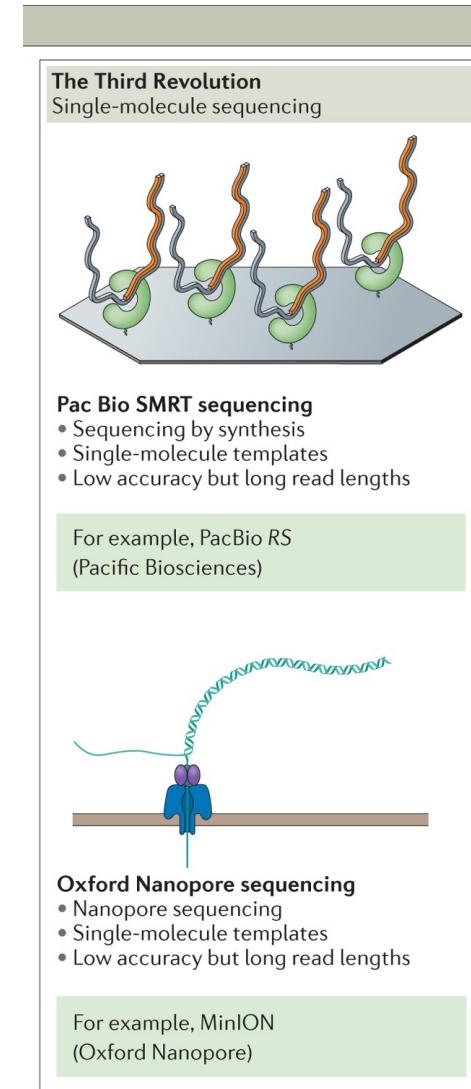
Igor S. Pessi & Antti Karkman, University of Helsinki



Environmental metagenomics
April 2021

(The incomplete) history of metagenomics

Third revolution:
Long-read sequencing



Environmental metagenomics
April 2021

Igor S. Pessi & Antti Karkman, University of Helsinki

(The incomplete) history of metagenomics

Third revolution:
Long-read sequencing



Igor S. Pessi & Antti Karkman, University of Helsinki

The Third Revolution
Single-molecule sequencing

Pac Bio SMRT sequencing

- Sequencing by synthesis
- Single-molecule templates
- Low accuracy but long read lengths

For example, PacBio RS
(Pacific Biosciences)

Oxford Nanopore sequencing

- Nanopore sequencing
- Single-molecule templates
- Low accuracy but long read lengths

For example, MinION
(Oxford Nanopore)

(The incomplete) history of metagenomics

Third revolution:
Long-read sequencing



Igor S. Pessi & Antti Karkman, University of Helsinki

The Third Revolution
Single-molecule sequencing

Pac Bio SMRT sequencing

- Sequencing by synthesis
- Single-molecule templates
- Low accuracy but long read lengths

For example, PacBio RS
(Pacific Biosciences)

Oxford Nanopore sequencing

- Nanopore sequencing
- Single-molecule templates
- Low accuracy but long read lengths

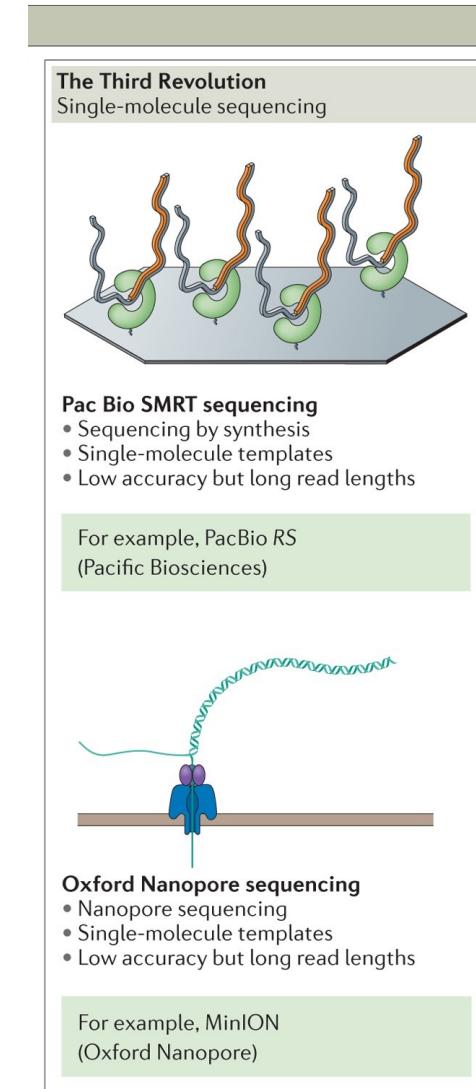
For example, MinION
(Oxford Nanopore)

(The incomplete) history of metagenomics

Third revolution:
Long-read sequencing



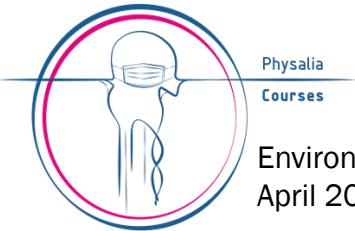
Igor S. Pessi & Antti Karkman, University of Helsinki



Sequel II
HiFi reads:
Read length:
20 kb
30 Gbp /run
(99.92 % acc.)

Minion:
Read length:
> 4 Mb
1–50 Gbp /run
(~ 97 % acc.)
Real-time!

From samples to sequences



Physalia
Courses

Environmental metagenomics
April 2021

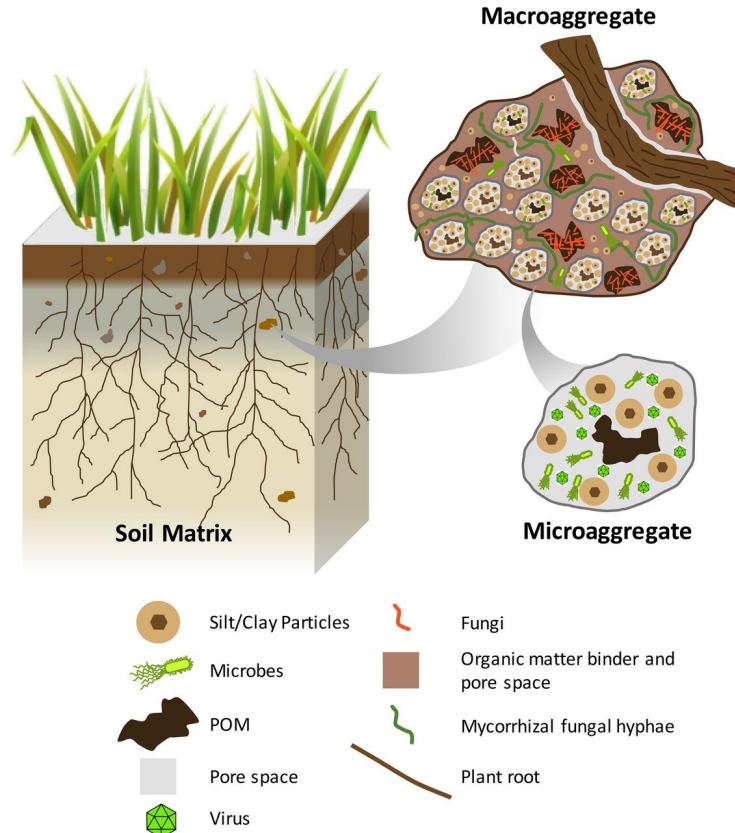
Igor S. Pessi & Antti Karkman, University of Helsinki

From samples to sequences

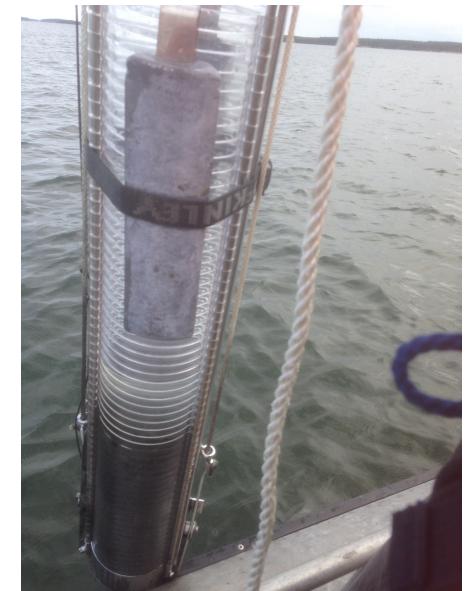
Sampling



PHOTO: JOCHEN TACK/ALAMY STOCK PHOTO

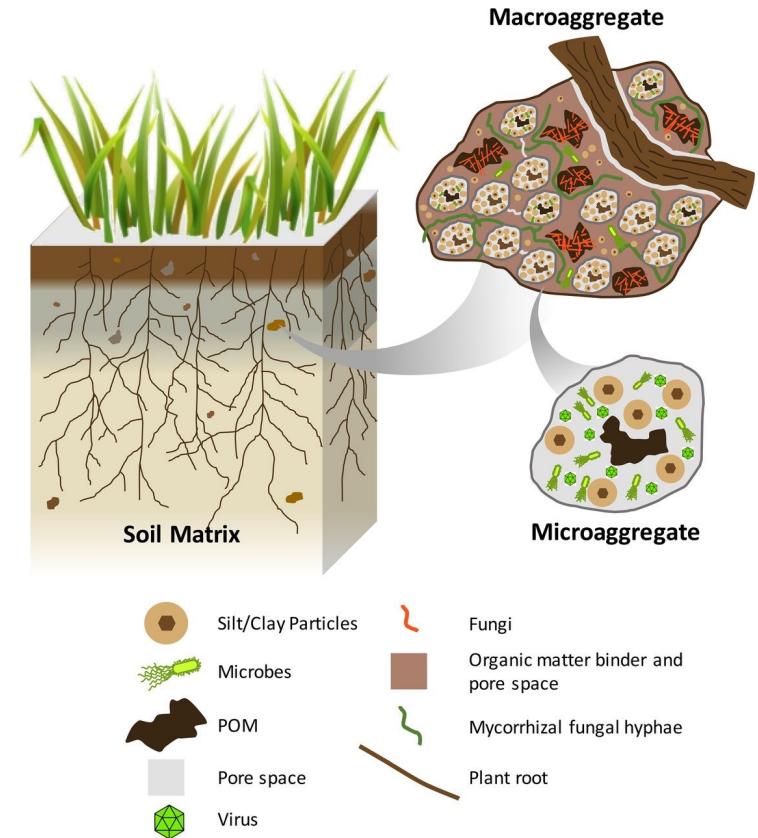


Regina L. Wilpszeski et al. *Appl. Environ. Microbiol.* 2019; doi:10.1128/AEM.00324-19



From samples to sequences

Sampling

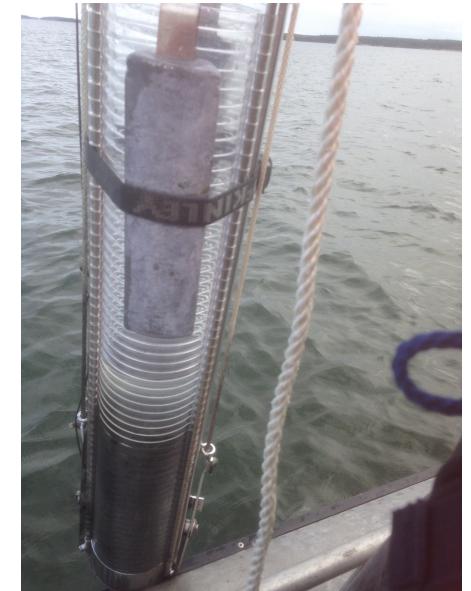


Sample size?

How many samples?

Research questions!

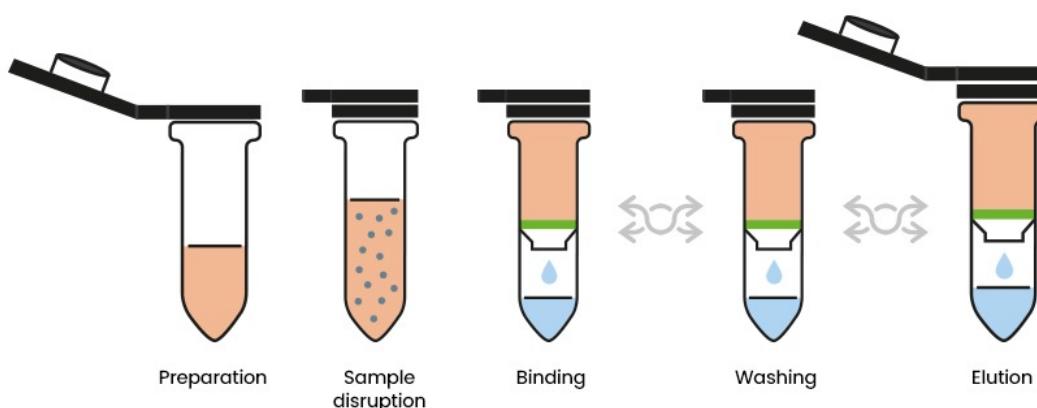
Metadata!



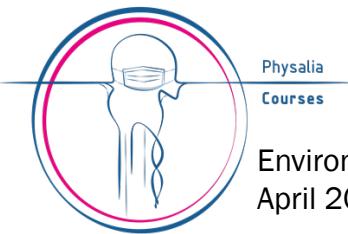
From samples to sequences

DNA extraction

Fragmented or HMW DNA?



<https://reallaboratory.com/product/real-microbial-dna-kit/>



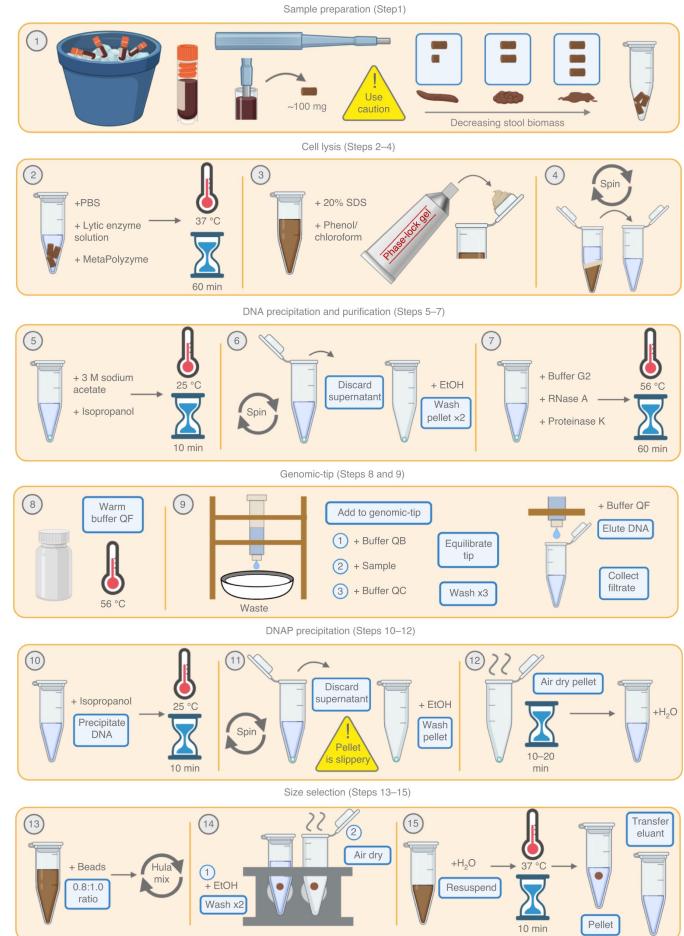
Environmental metagenomics
April 2021

Igor S. Pessi & Antti Karkman, University of Helsinki

Improved high-molecular-weight DNA extraction,
nanopore sequencing and metagenomic assembly
from the human gut microbiome

Dylan G. Maghini, Eli L. Moss, Summer E. Vance & Ami S. Bhatt

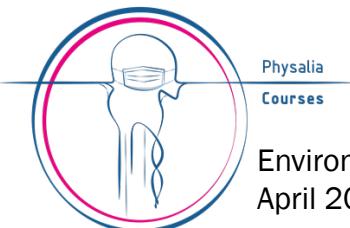
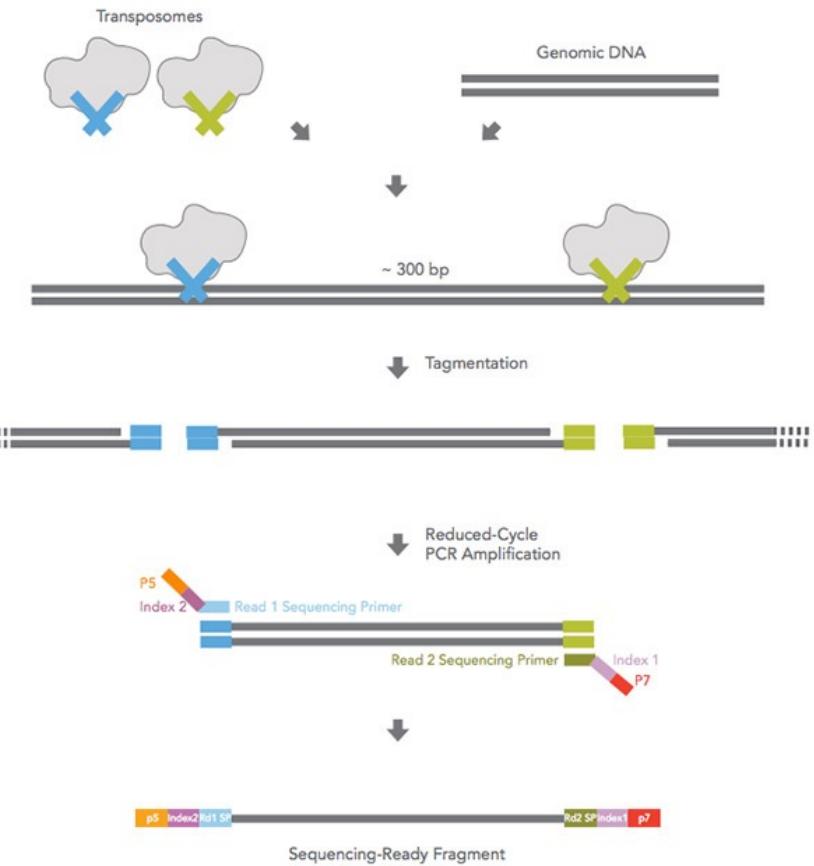
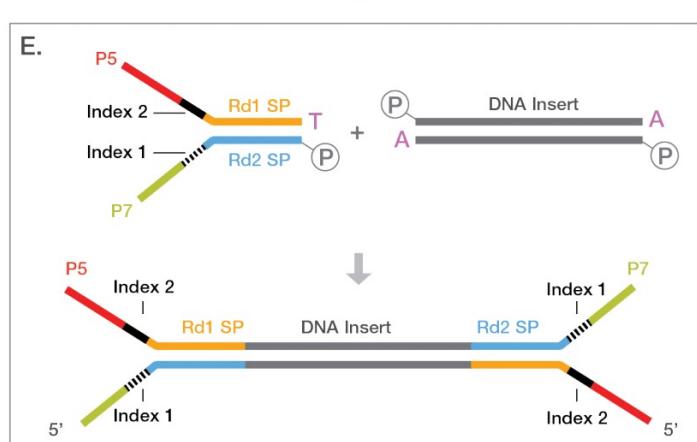
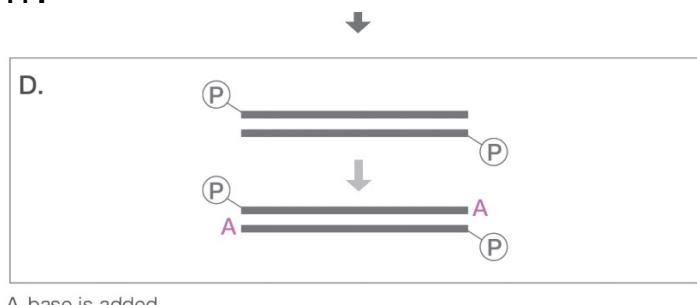
Nature Protocols 16, 458–471(2021) | Cite this article



From samples to sequences

Library preparation

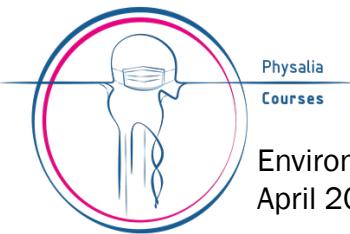
DNA concentration?



From samples to sequences

Sequencing

Long and/or short reads?



From samples to sequences

Sequencing

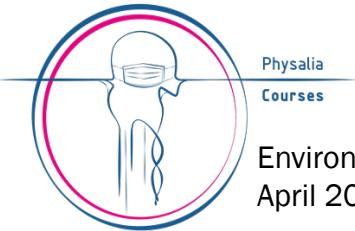
Long and/or short reads?



	NextSeq 550* †	NextSeq 1000 & 2000*	NovaSeq 6000* ‡‡
Output Range	20–120 Gb	40–330 Gb***	65–6000 Gb
Run Time	11–29 hours	11–48 hours	13–44 hours
Reads Per Run	130–400 million	0.4–1.1 billion***	Up to 20 billion
Max Read Length	2 × 150 bp	2 × 150 bp	2 × 250 bp†††
Samples Per Run§	8	8–20	12–400
Relative Price Per Sample§	Higher Cost	Mid Cost	Low Cost

Igor S. Pessi & Antti Karkman, University of Helsinki

Bioinformatics



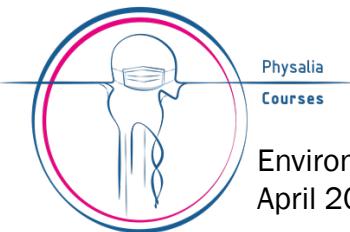
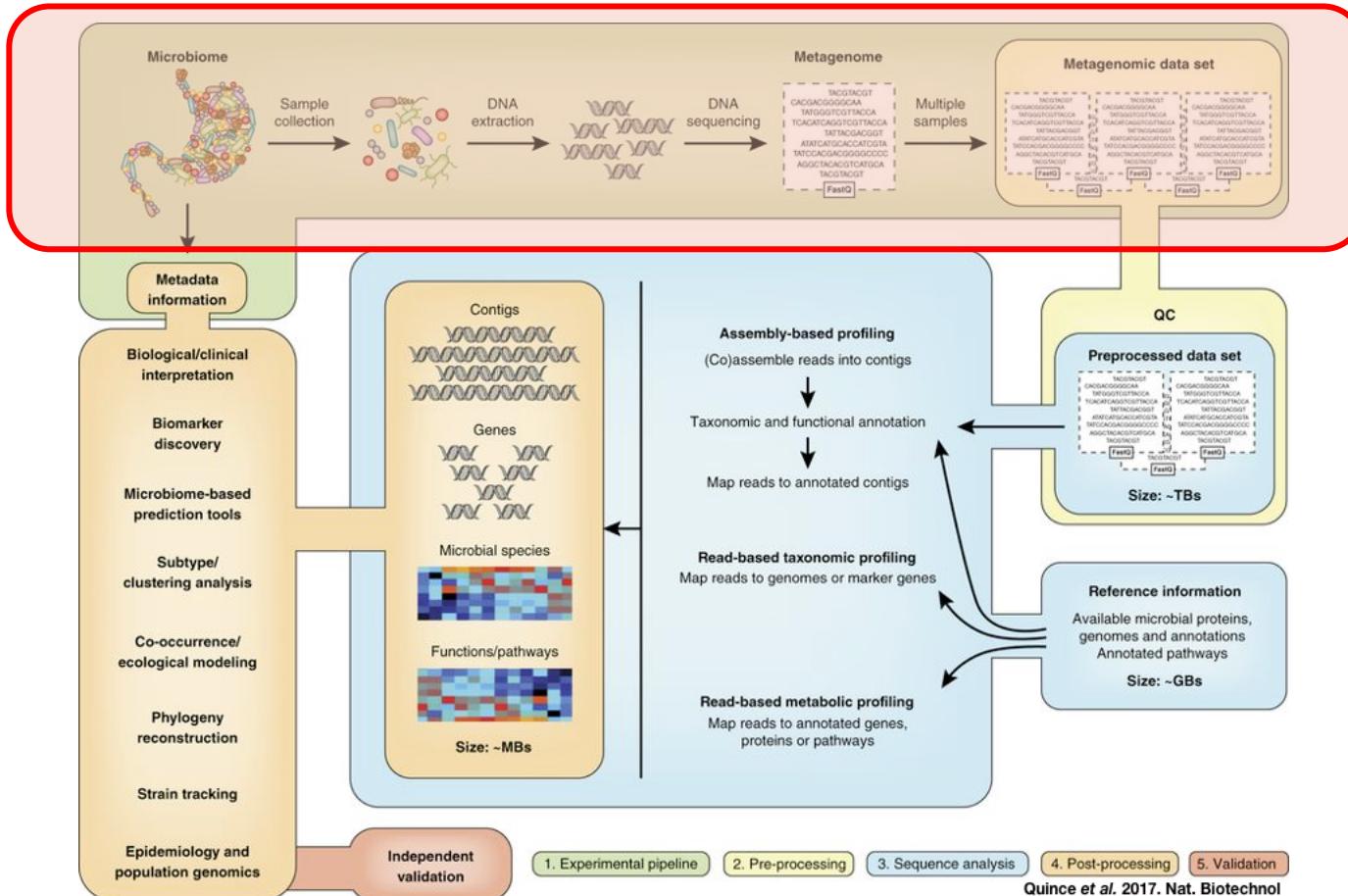
Physalia
Courses

Environmental metagenomics
April 2021

Igor S. Pessi & Antti Karkman, University of Helsinki

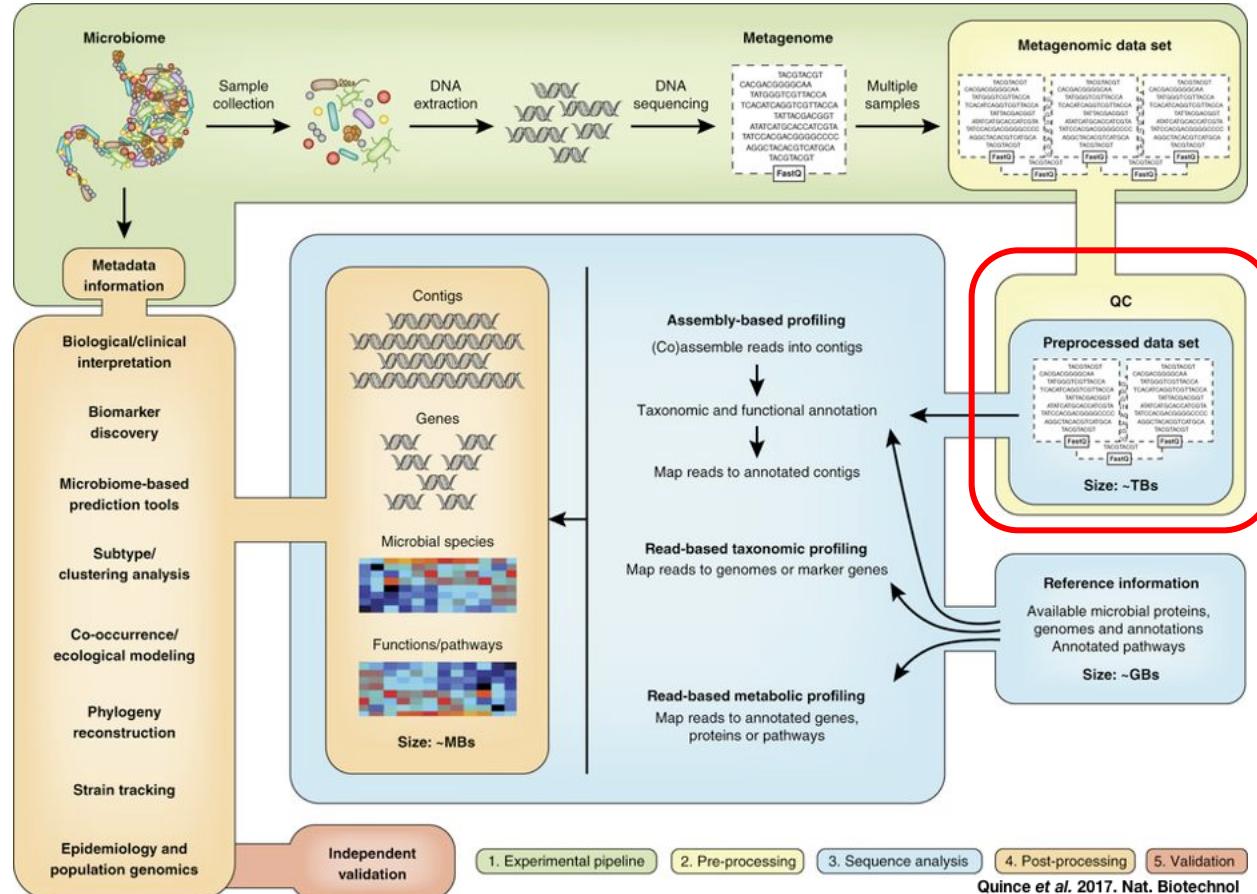
Bioinformatics

Whole workflow



Bioinformatics

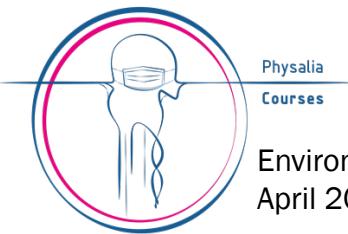
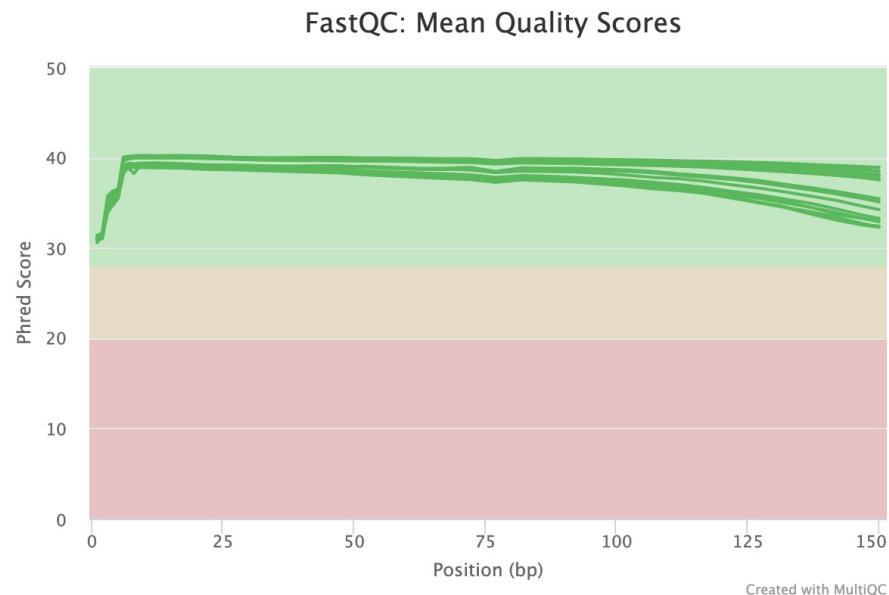
Quality control (QC)



Bioinformatics

Quality control (QC)

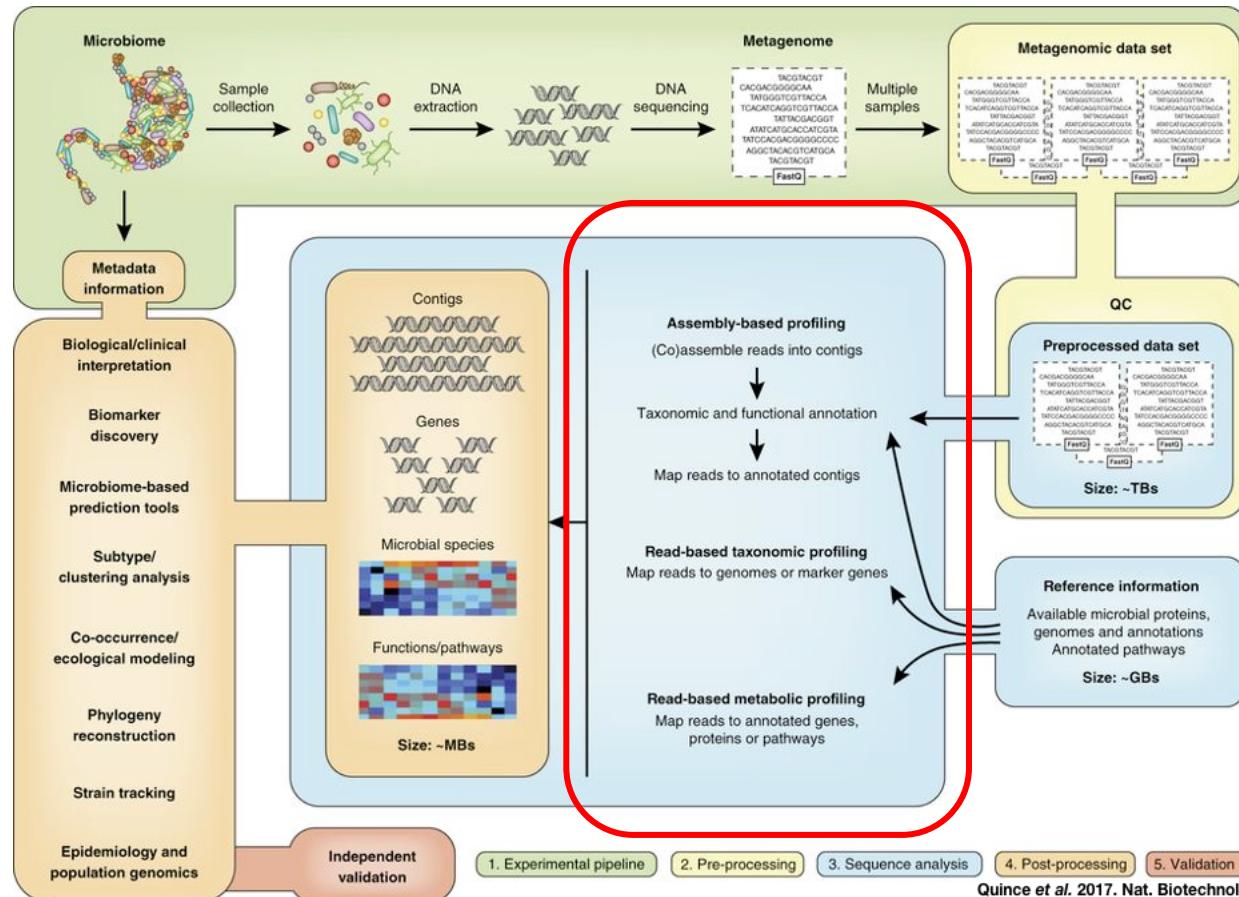
Phred score:
 $Q = -10 \log E$
 $E = 10^{-Q/10}$



Bioinformatics

Profiling:

Read- vs. assembly-based

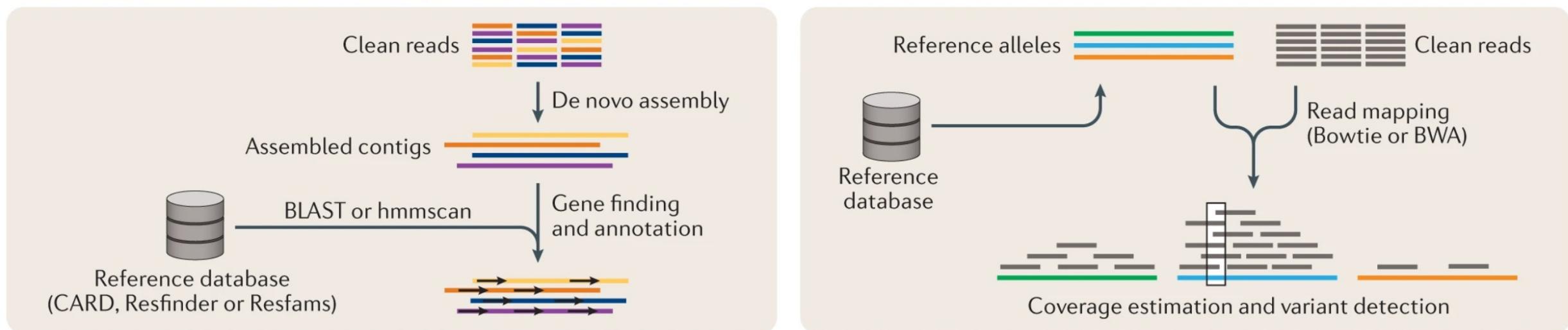


Bioinformatics

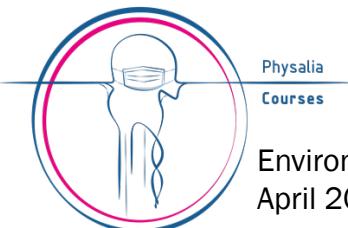
Profiling:

Read- vs. assembly-based

b Assembly-based versus read-based approach



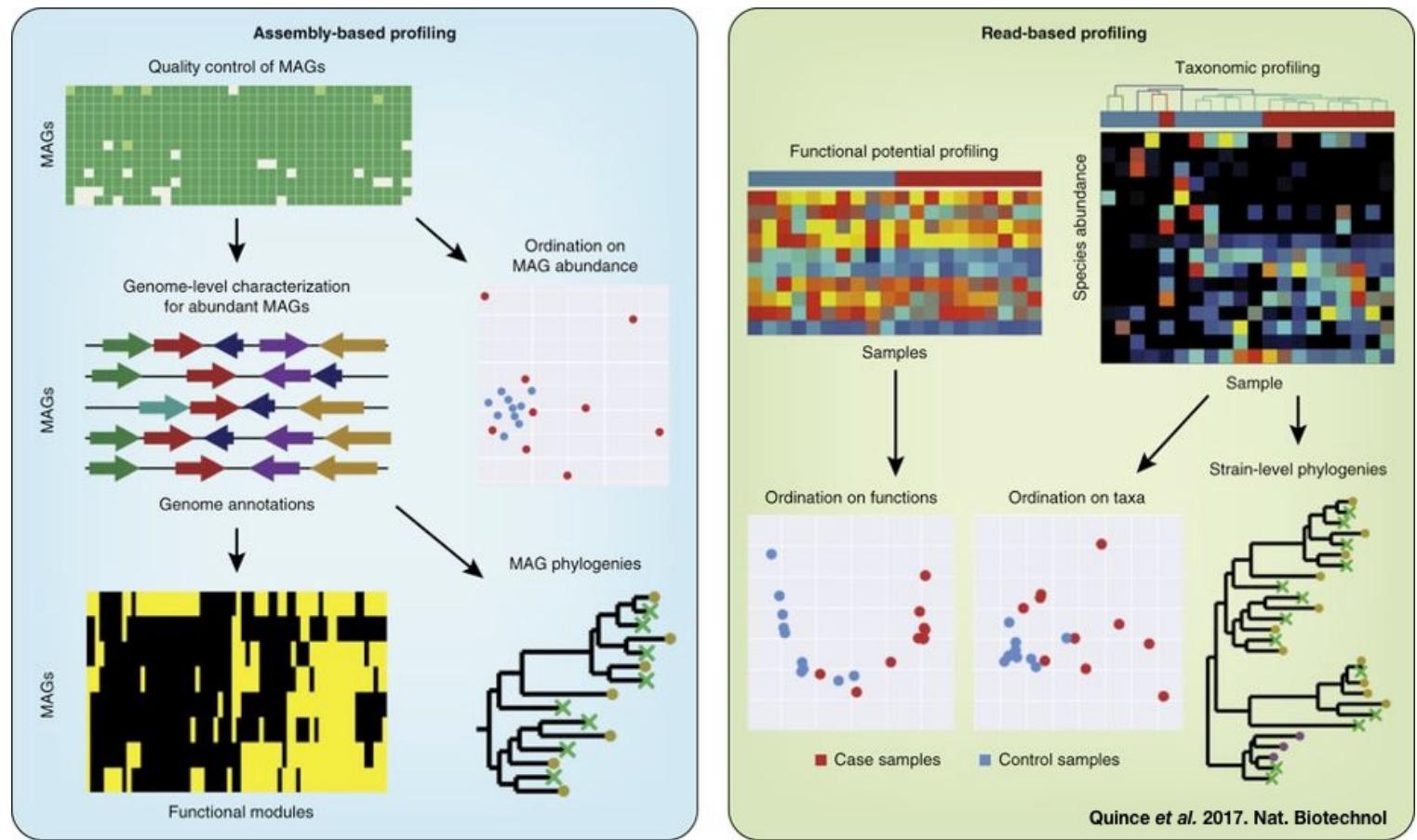
Boolchandani et al. 2019. Nat.Rev.Gen.



Bioinformatics

Profiling:

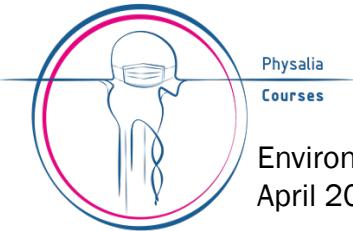
Read- vs. assembly-based



Bioinformatics

Profiling:

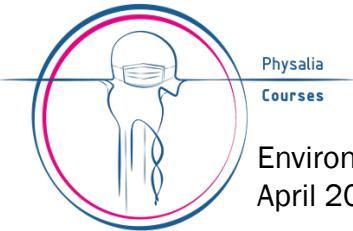
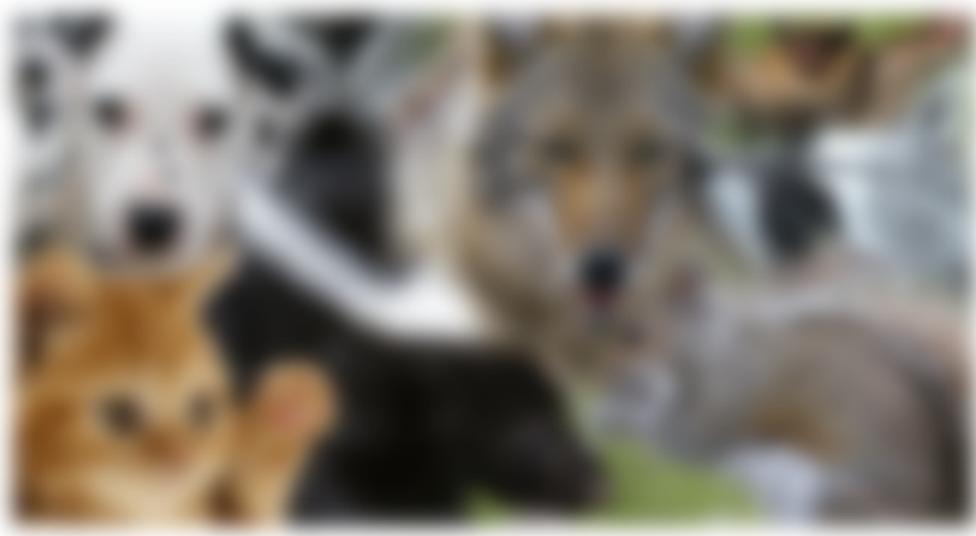
Read- vs. assembly-based



Bioinformatics

Profiling:

Read- vs. assembly-based



Physalia
Courses

Environmental metagenomics
April 2021

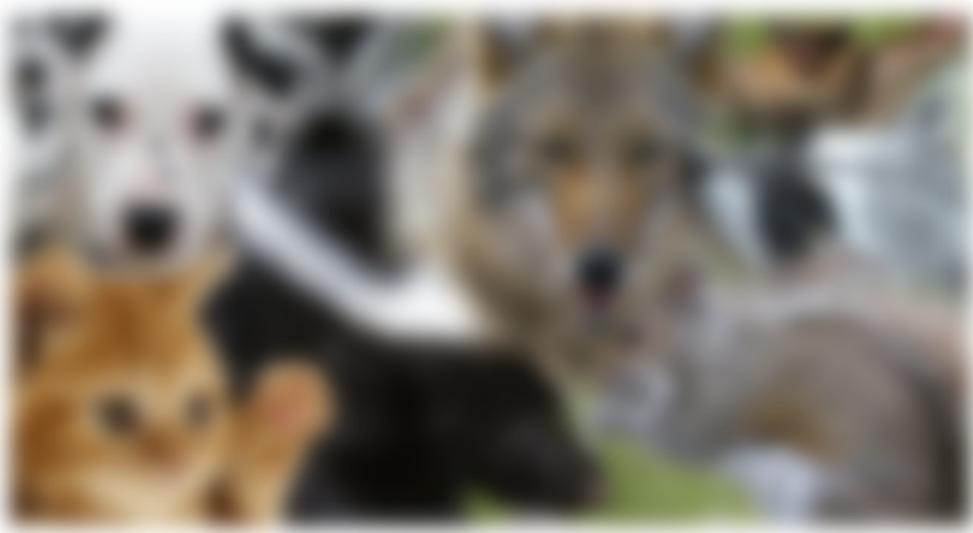
Igor S. Pessi & Antti Karkman, University of Helsinki

45

Bioinformatics

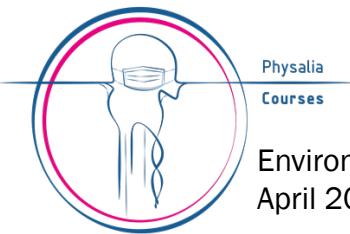
Profiling:

Read- vs. assembly-based



Resources for self-learning

- Happy Belly Bioinformatics: <https://astrobiomike.github.io/>
- Merenlab: <https://merenlab.org/>
- Huttenhower Lab: <https://huttenhower.sph.harvard.edu/tools/>
- Rob Edwards YouTube:
<https://www.youtube.com/user/RobEdwardsSDSU>



Physalia
Courses

Environmental metagenomics
April 2021

Igor S. Pessi & Antti Karkman, University of Helsinki