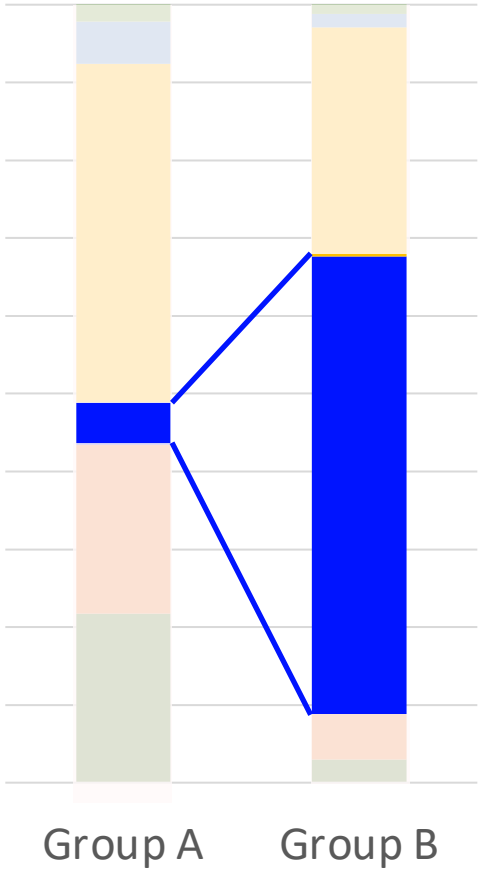
From reads to genomes

Silas Kieser

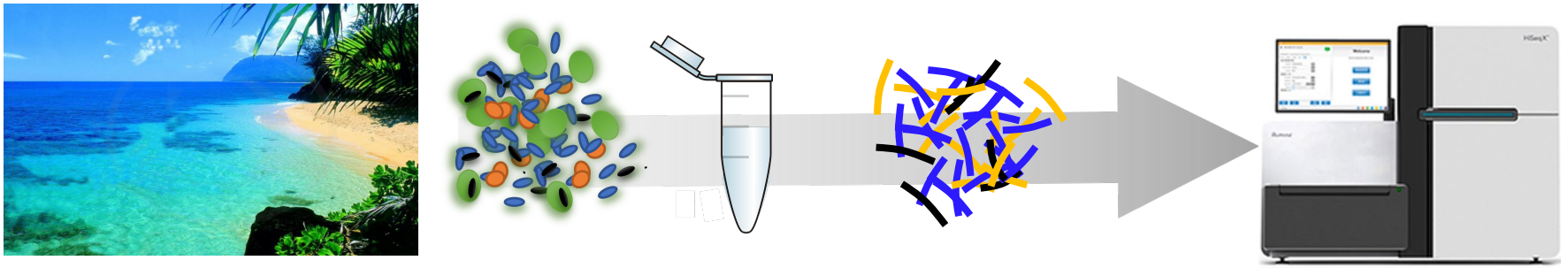

Group A
Group B

Which one is it?

First metagenome study

Stein et al. 1996

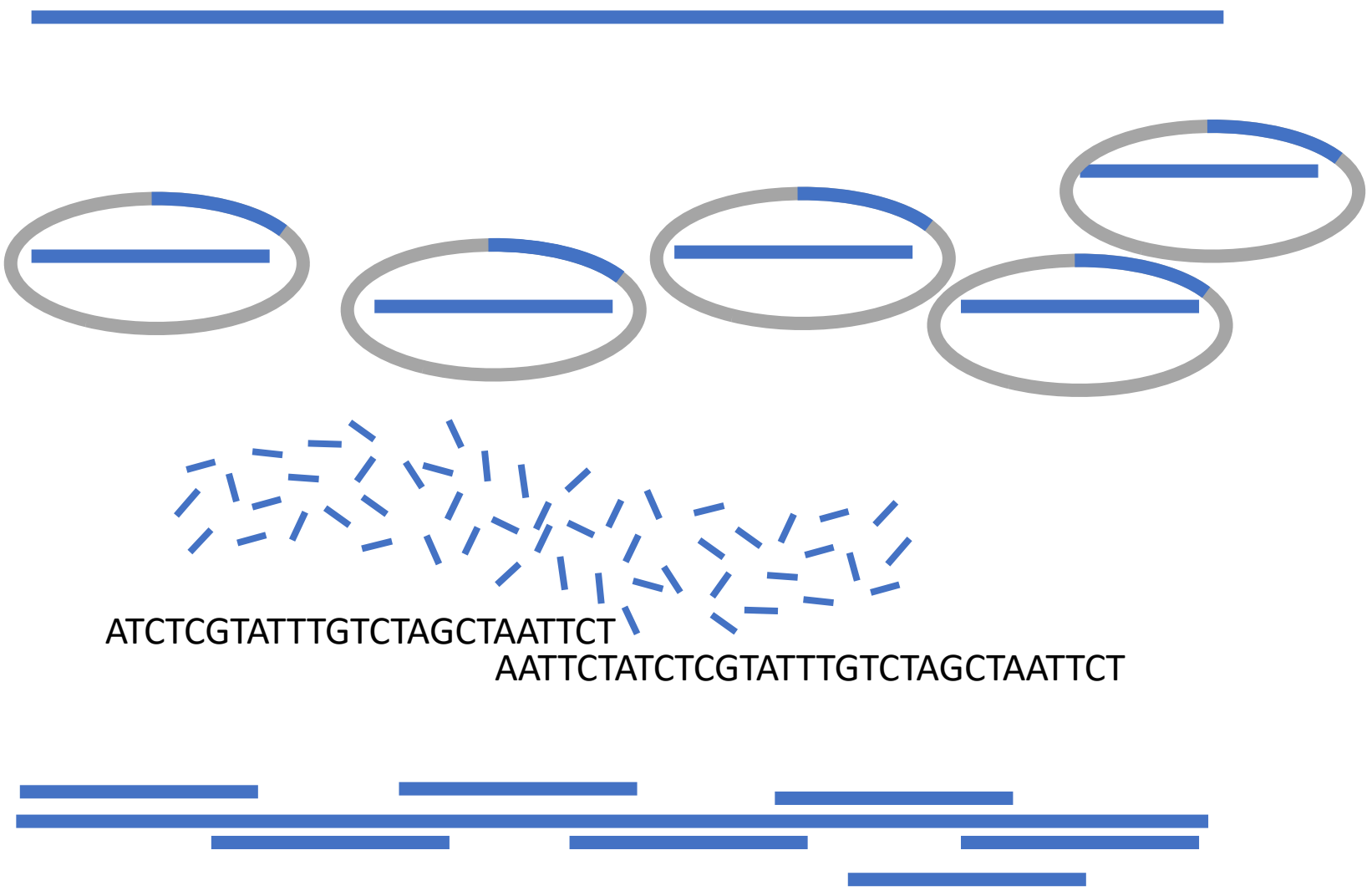
Sequence DNA directly form where they are

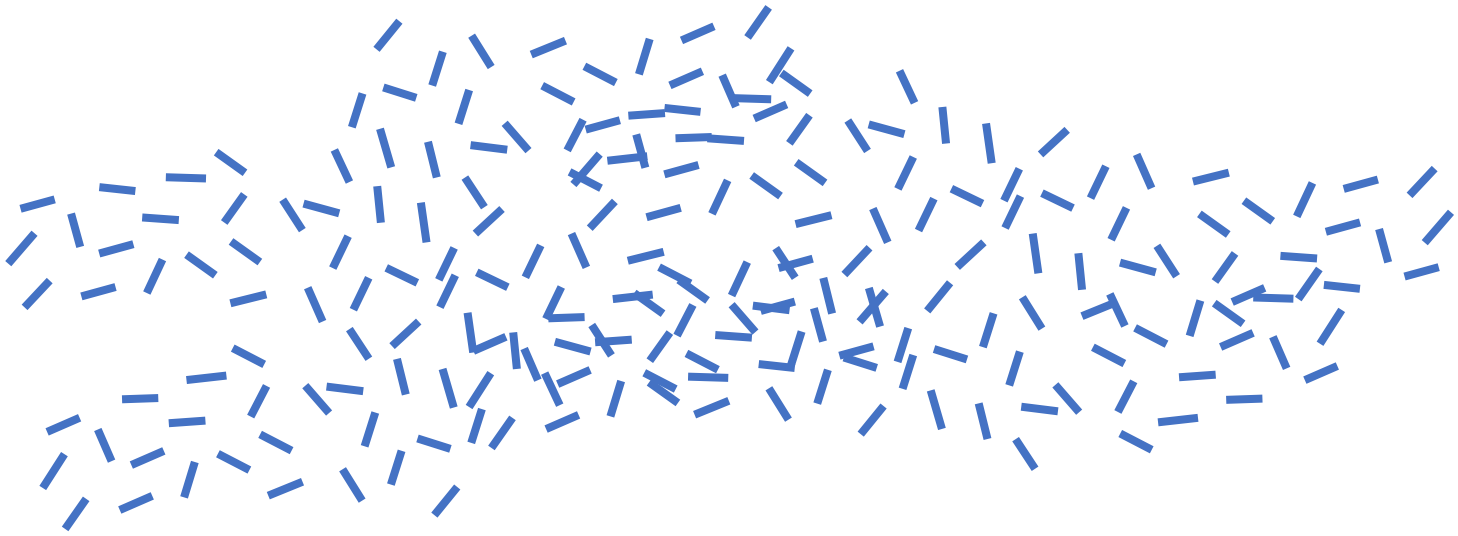


JOURNAL OF BACTERIOLOGY, Feb. 1996, p. 591–599 Vol. 178, No. 3
0021-9193/96/$04.00+0

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

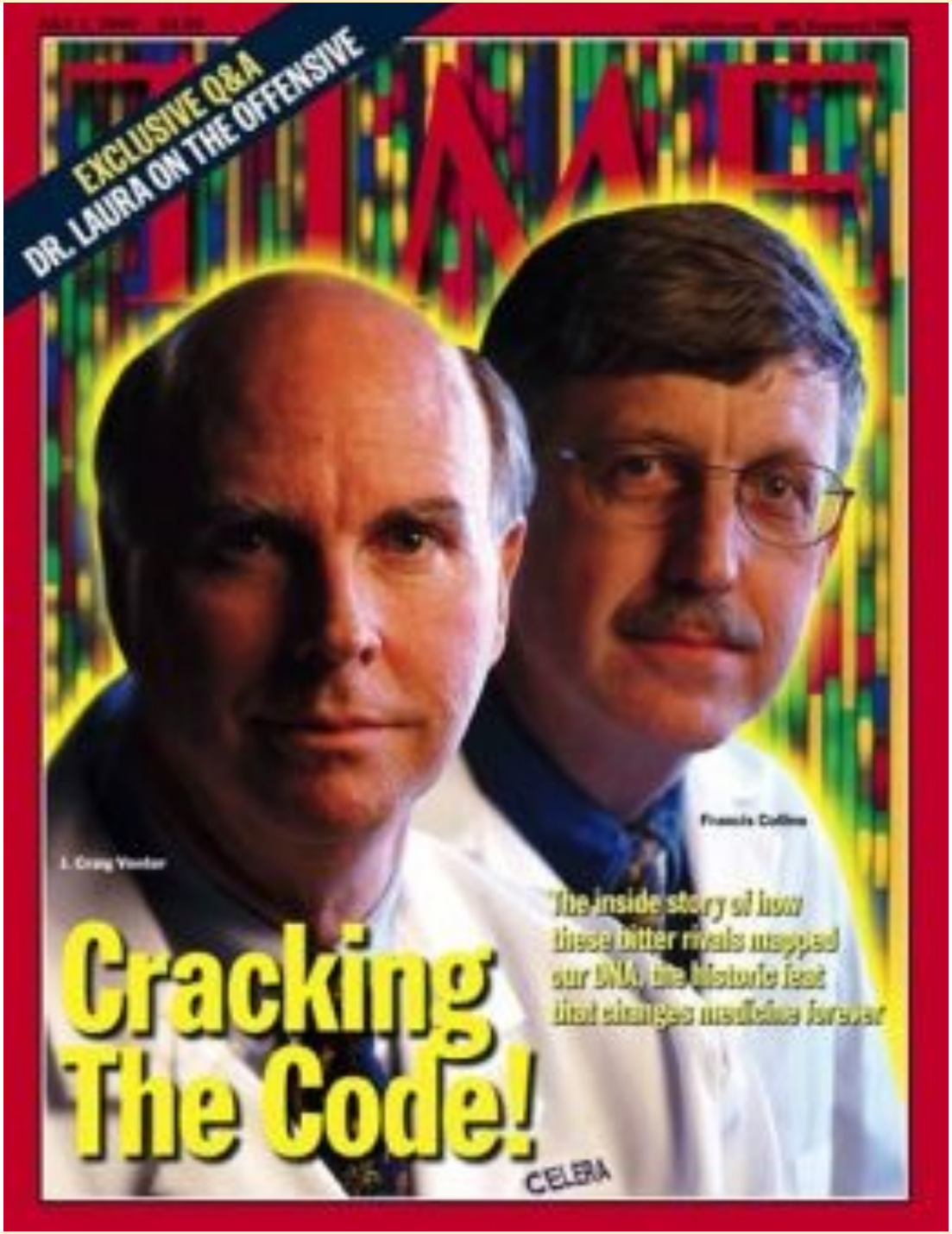
Human genome project

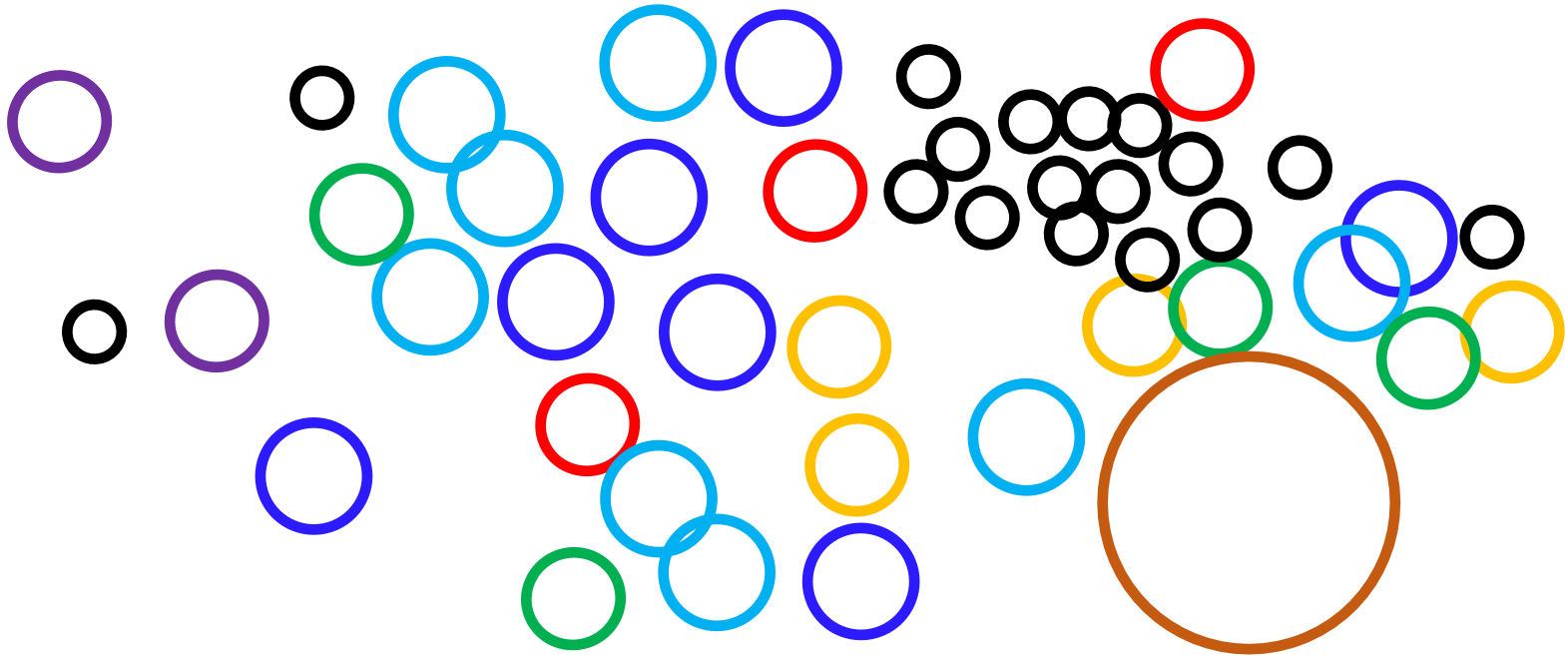

ATCTCGTATTTGTCTAGCTAATTCT
AATTCTATCTCGTATTTGTCTAGCTAATTCT



ATCTCGTATTTGTCTAGCTAATTCT

AATTCTATCTCGTATTTGTCTAGCTAATTCT


EXCLUSIVE Q&A
DR. LAURA ON THE OFFENSIVE
J. Craig Venter
Francis Collins
The inside story of how these bitter rivals mapped our DNA, the historic feat that changes medicine forever
Cracking The Code!
CELERA



How do algorithms cope with this complexity?

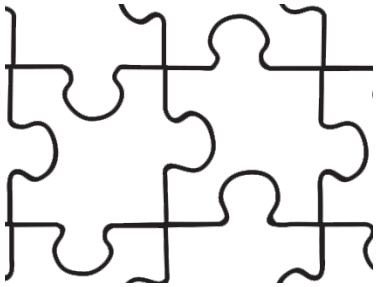
They don’t


Welcome
HiSeq X

Children like puzzles, and they usually assemble them by trying all possible pairs of pieces and putting together pieces that match.

— Pevzner et al., 2001

Assembly before



ATCGTCACGTTT

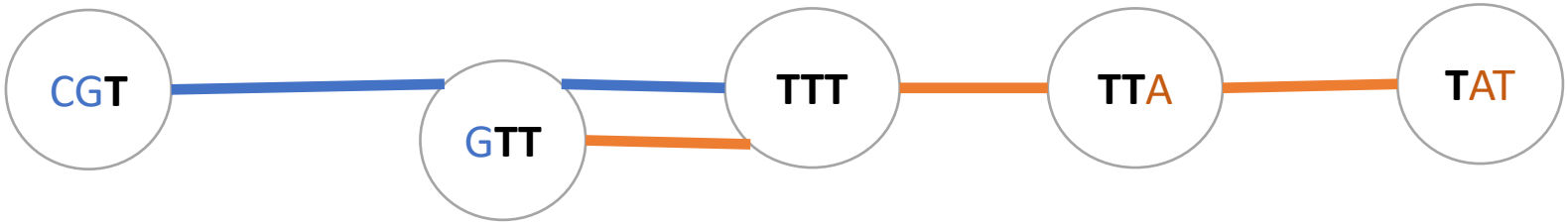
GTTTATCGTCTG

Assembly today

K=3

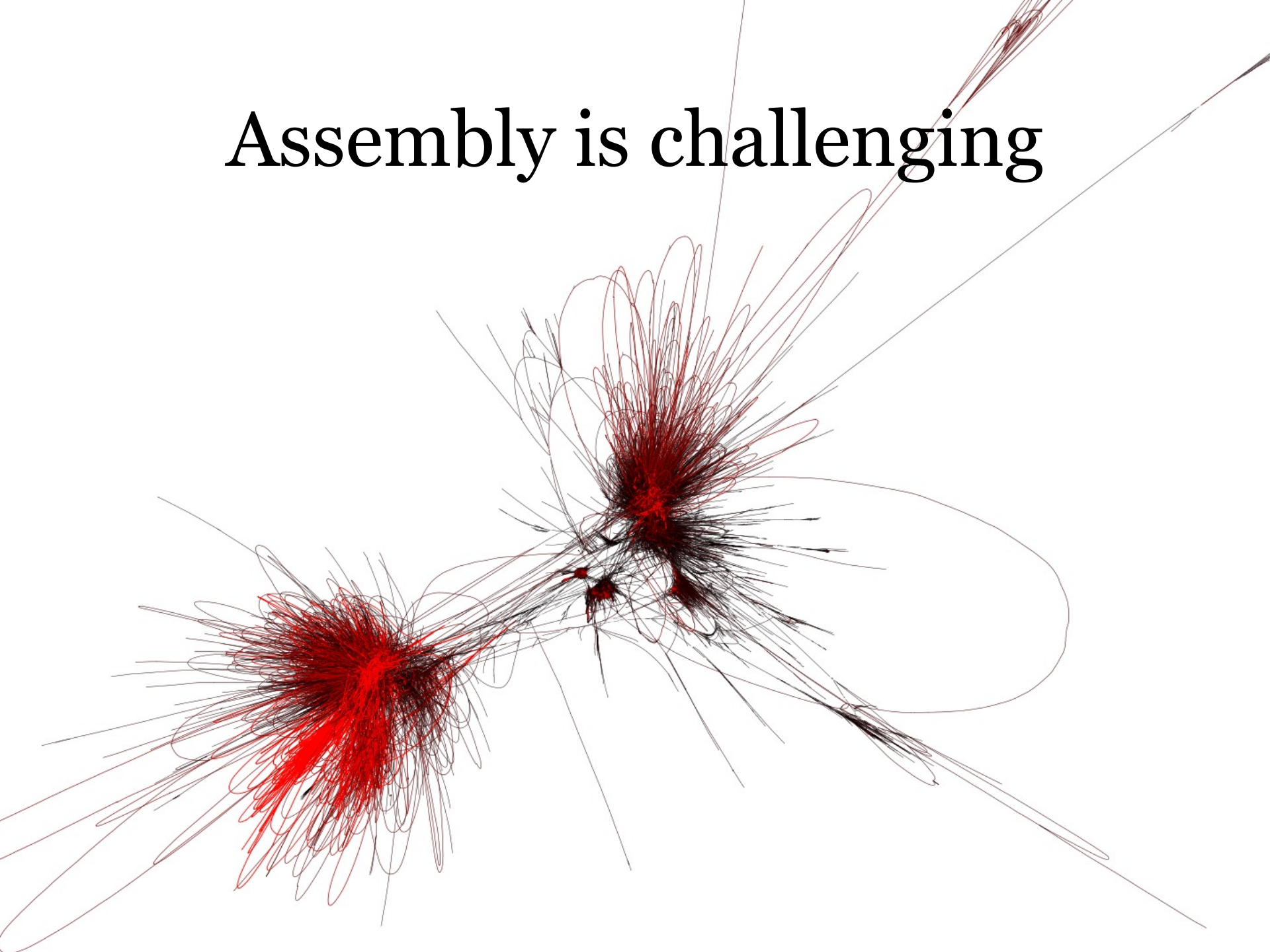
ATCGTCACGTTT

GTTTATCGTCTG

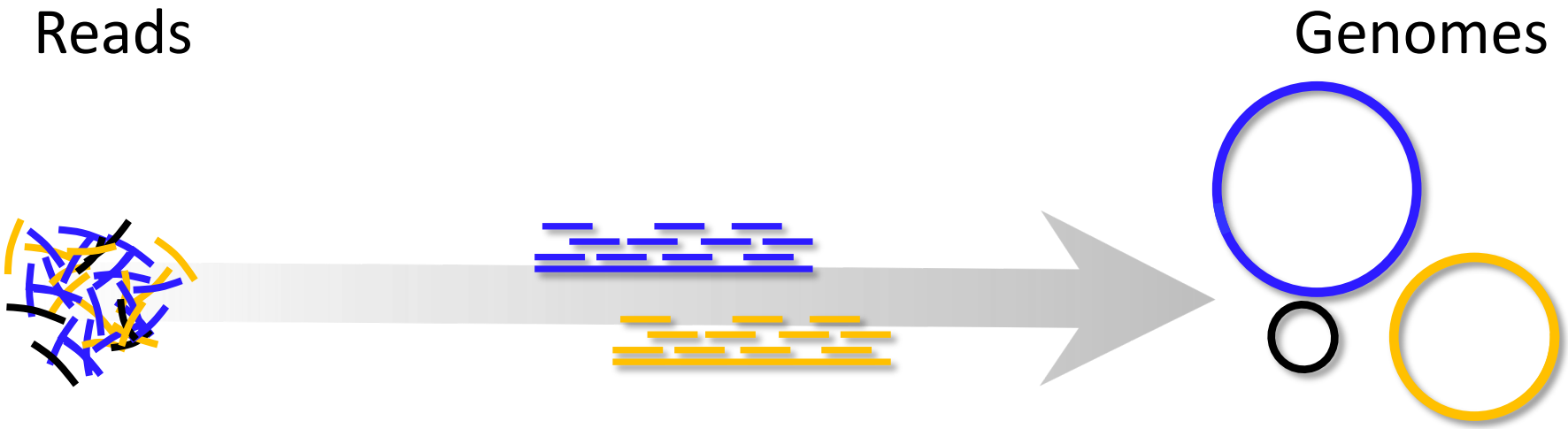


K-mer

Assembly is challenging



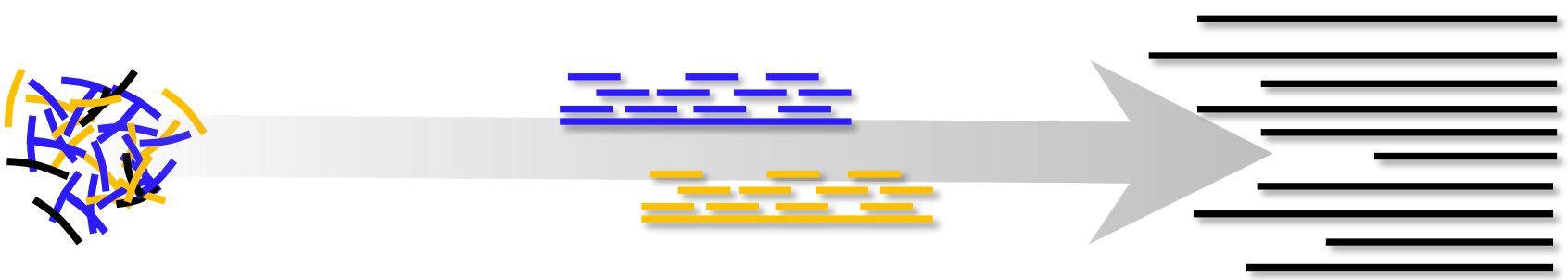
Goal: Assemble genomes from metagenomes



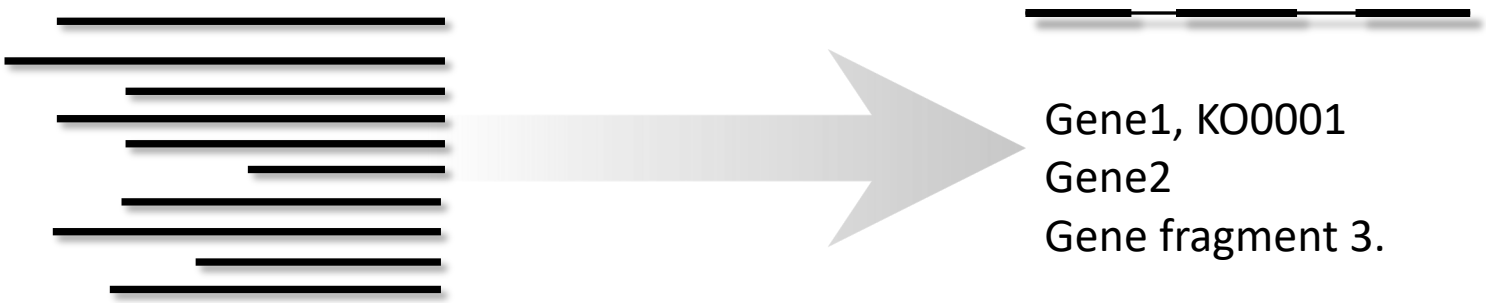
Goal: Assemble genomes from metagenomes

Reads

Contigs



Annotate genes on contigs

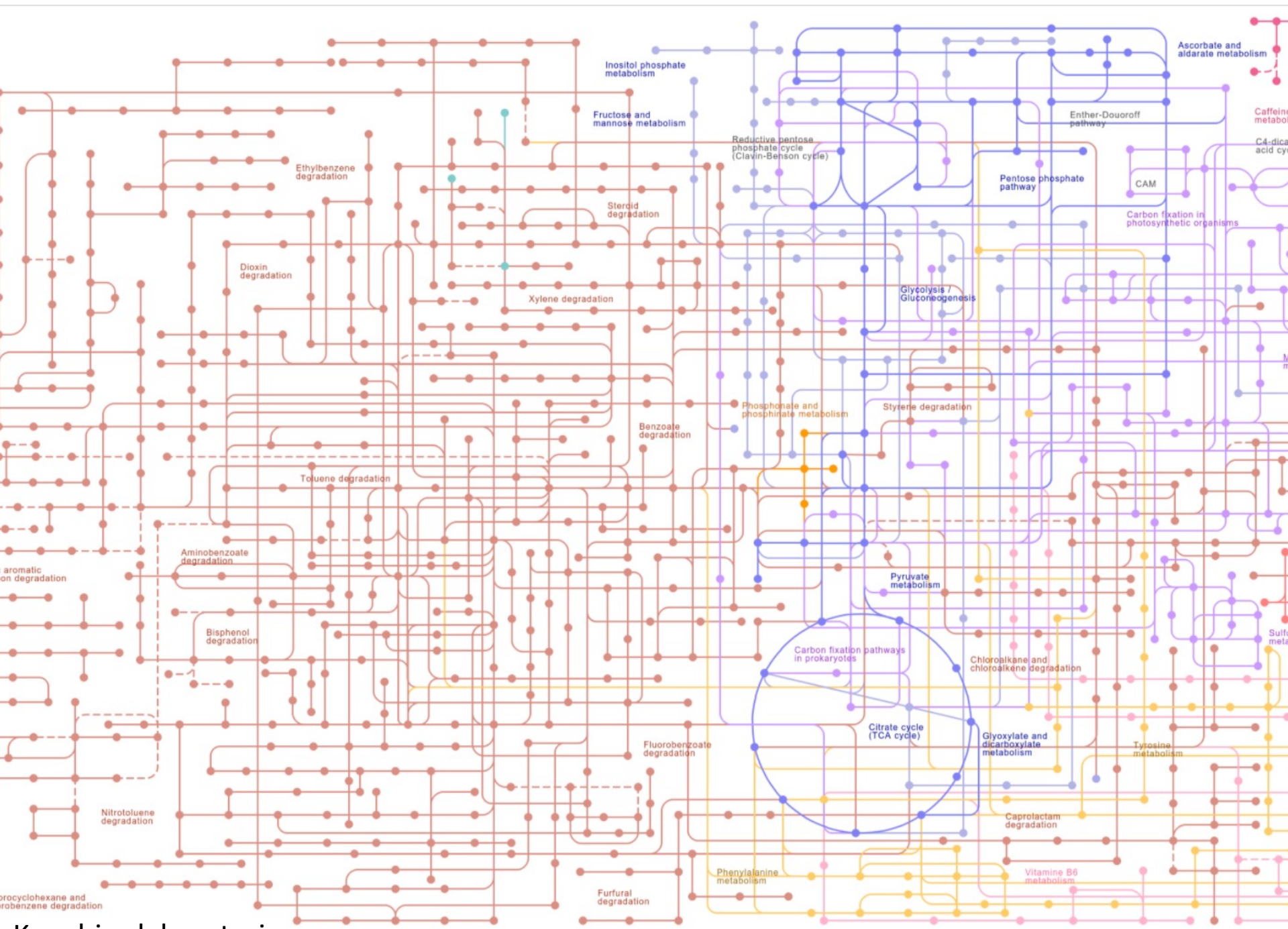


Era of gene catalogs

Functional potential

Bacteroidetes sp

Taxonomy



Kanehisa laboratories

Analyze a metagenome

Who is there?

What are they doing?

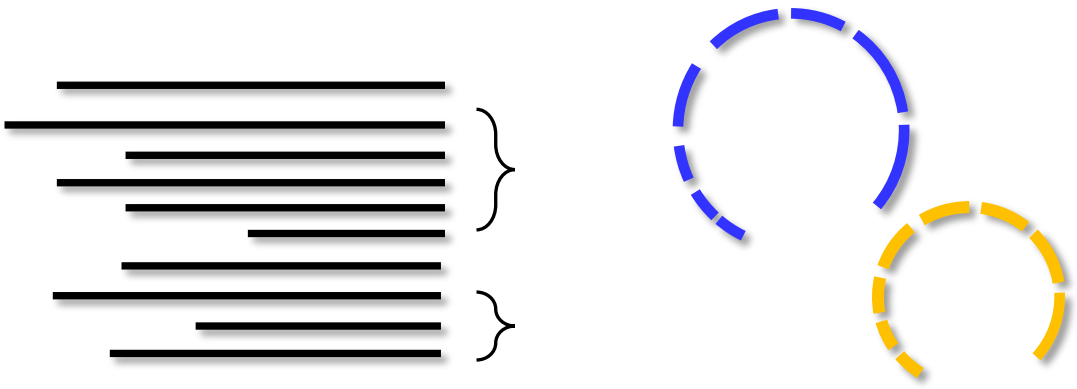
Analyze a metagenome

Who is doing
what?

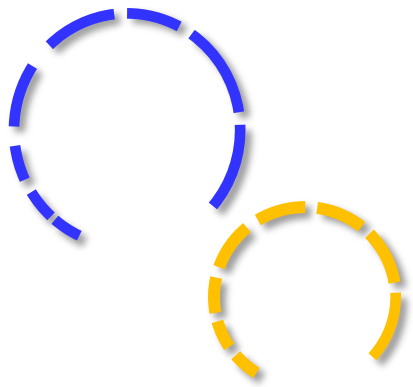
“Genes are expressed within cells, not in a homogenized cytoplasmic soup.”

Katherine McMahon

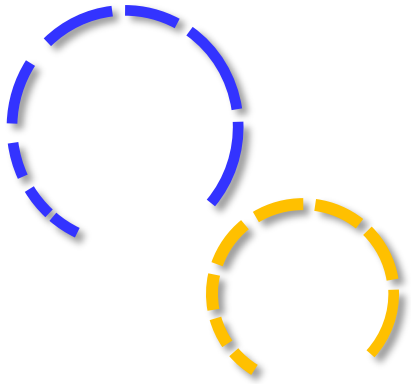
Binning



Binning



Metagenome-assembled genomes (MAGs)



The beginning of binning

Tyson et al. 2004

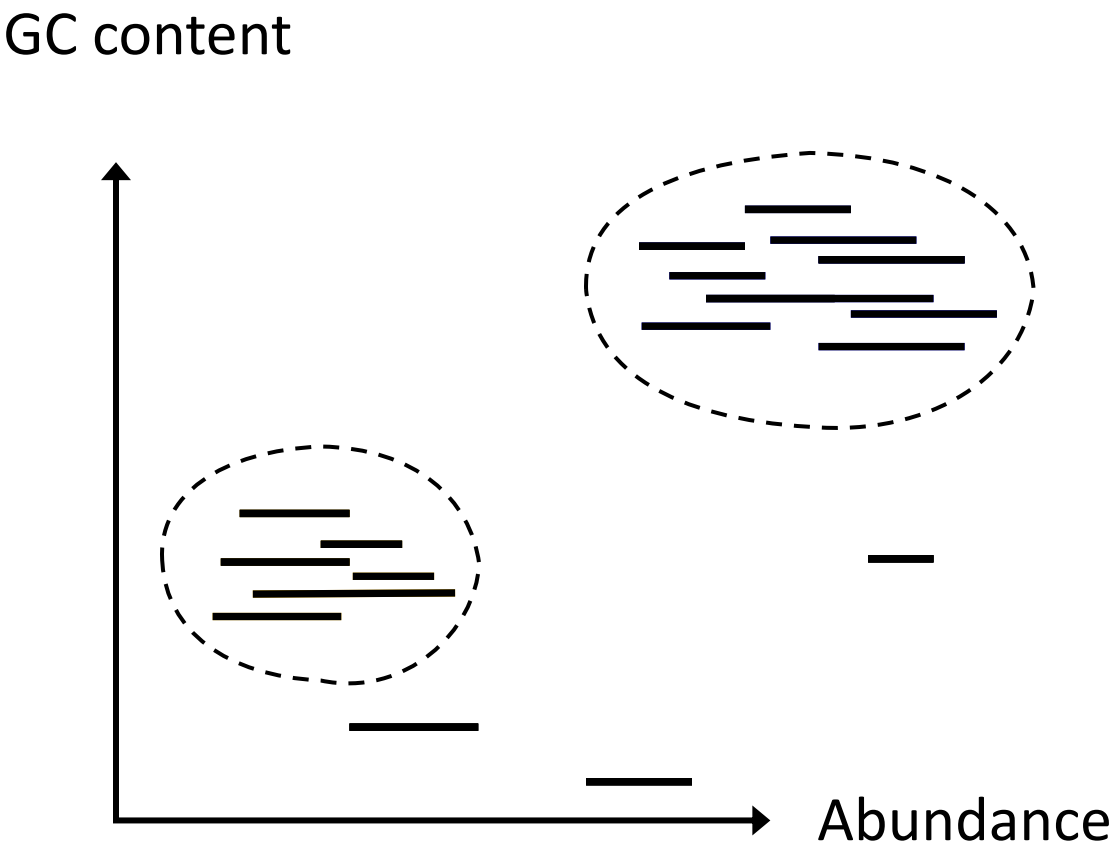
How do we bin contigs into genomes?

GC content

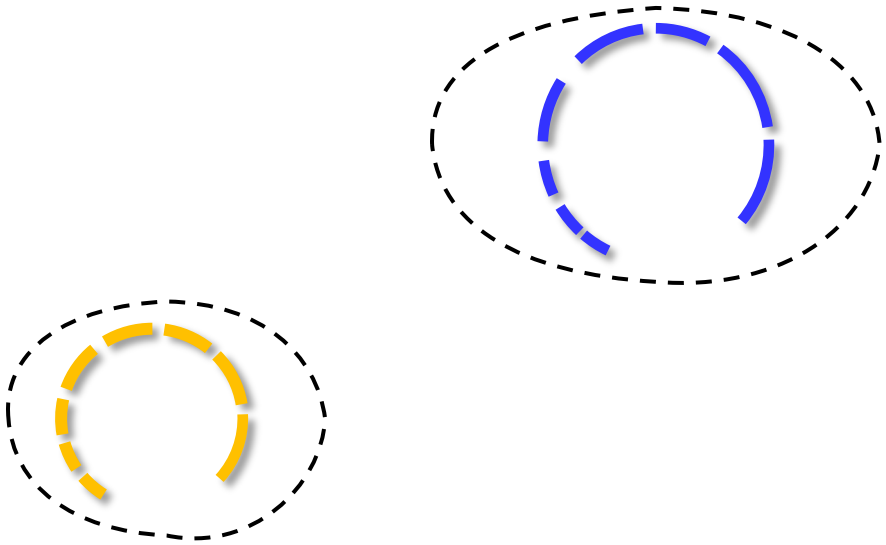
Abundance

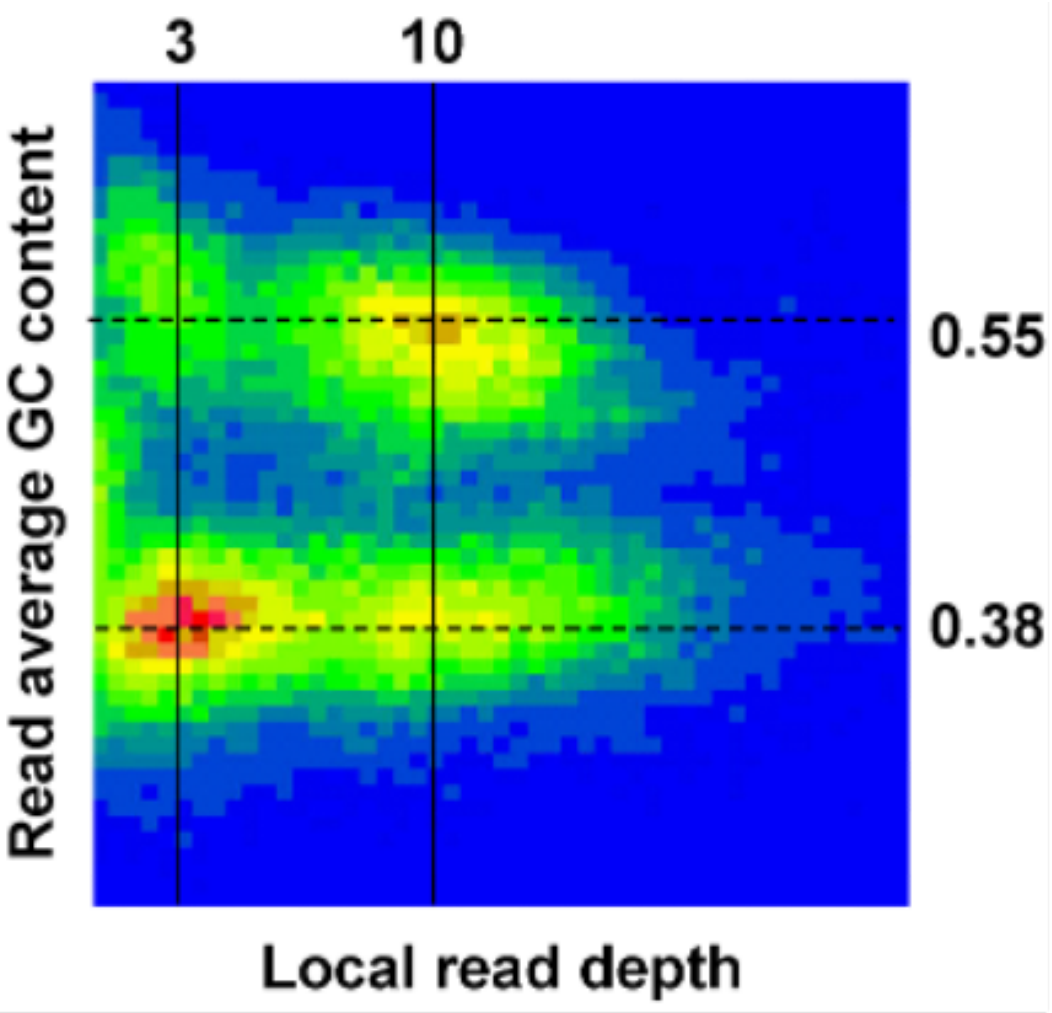
ATCGTCACG

Binning

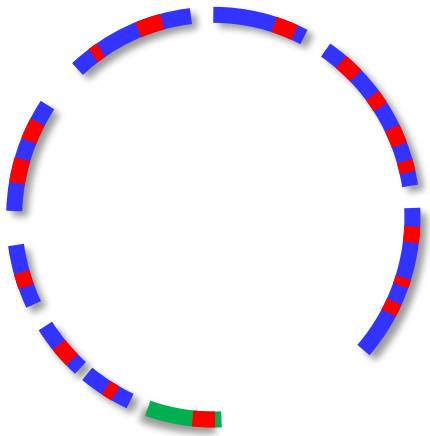


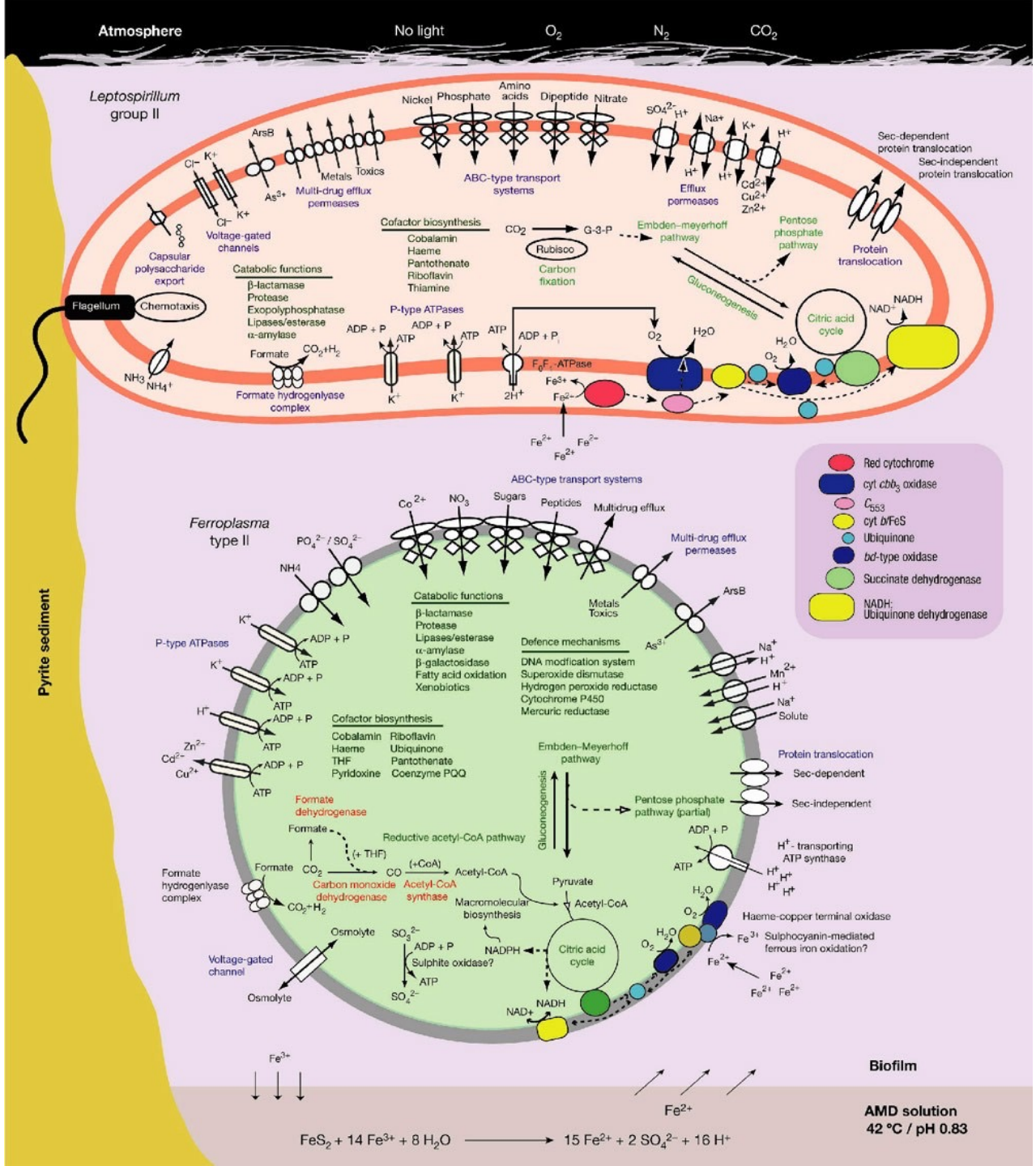
Metagenome-assembled genomes (MAGs)

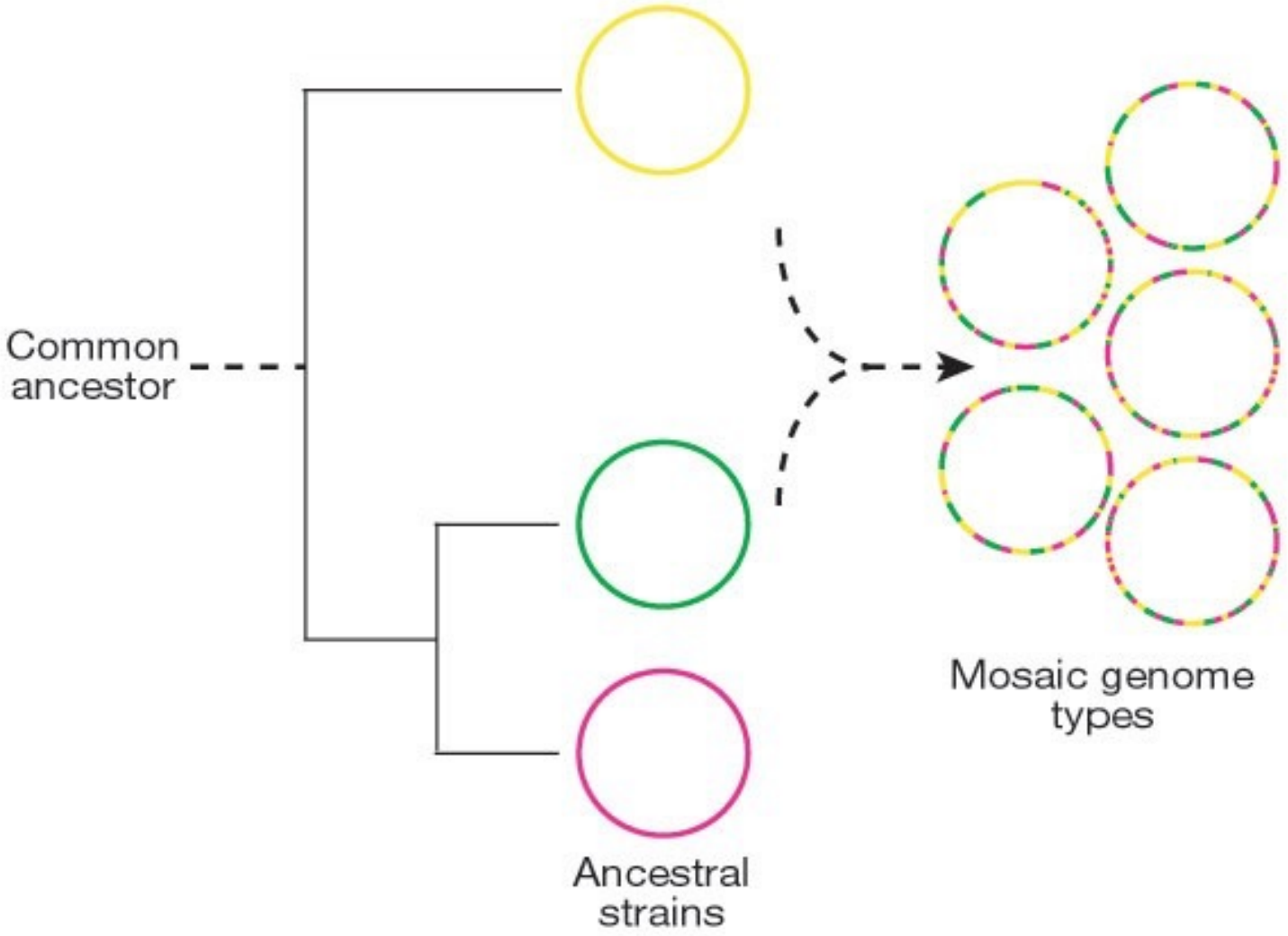


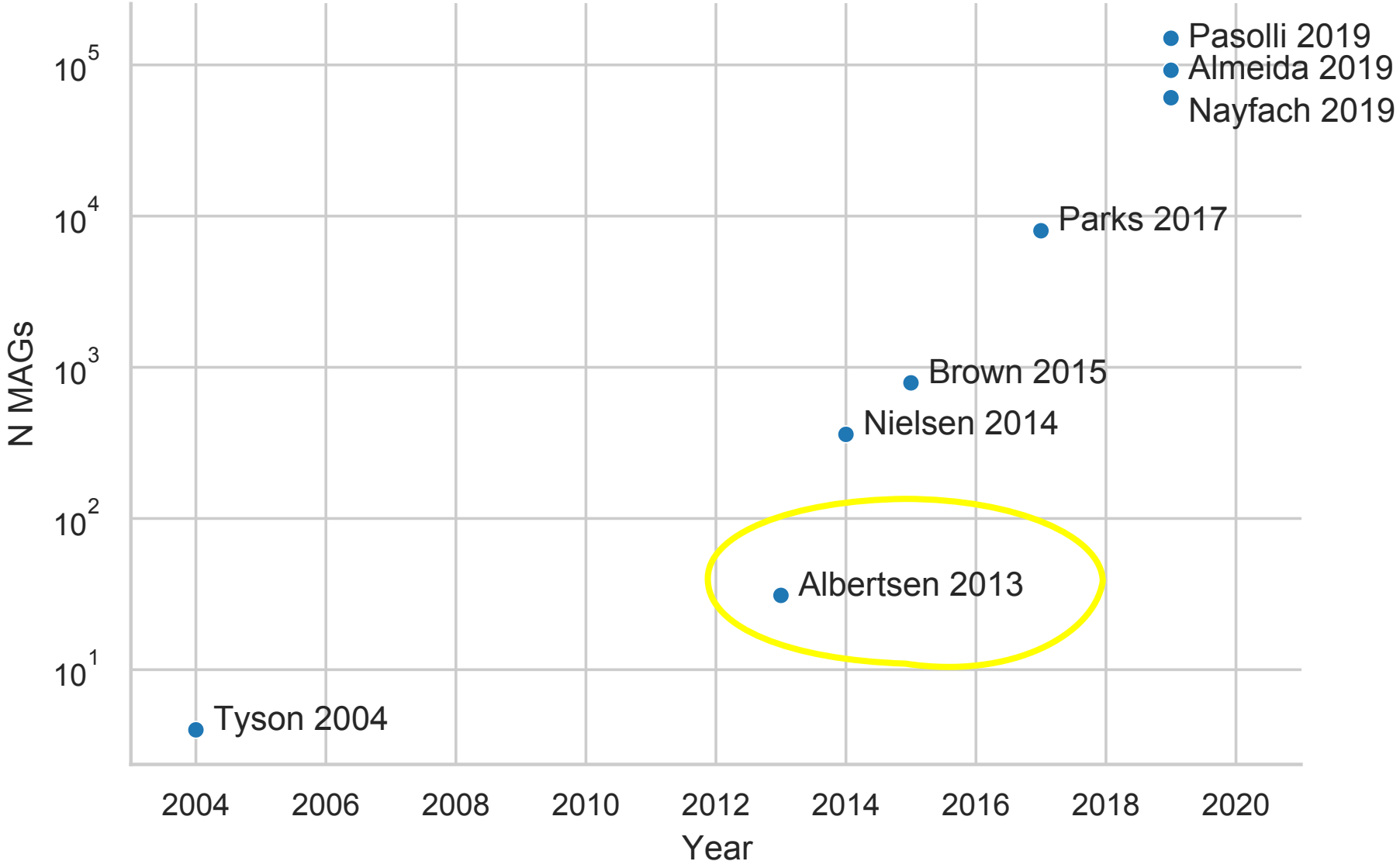

3
10
Read average GC content
0.55
0.38
Local read depth

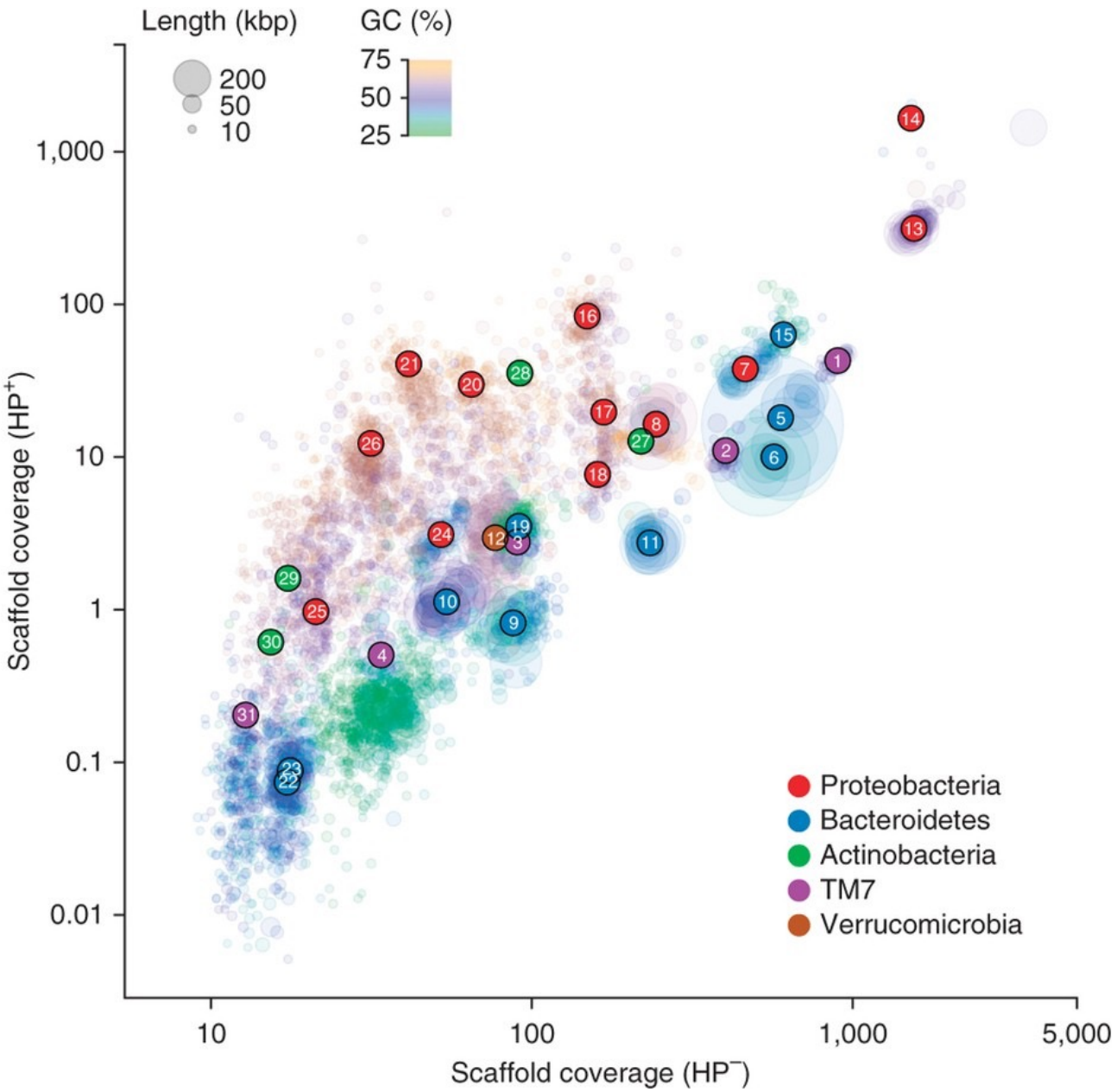
Quality assessment of MAGs using essential genes

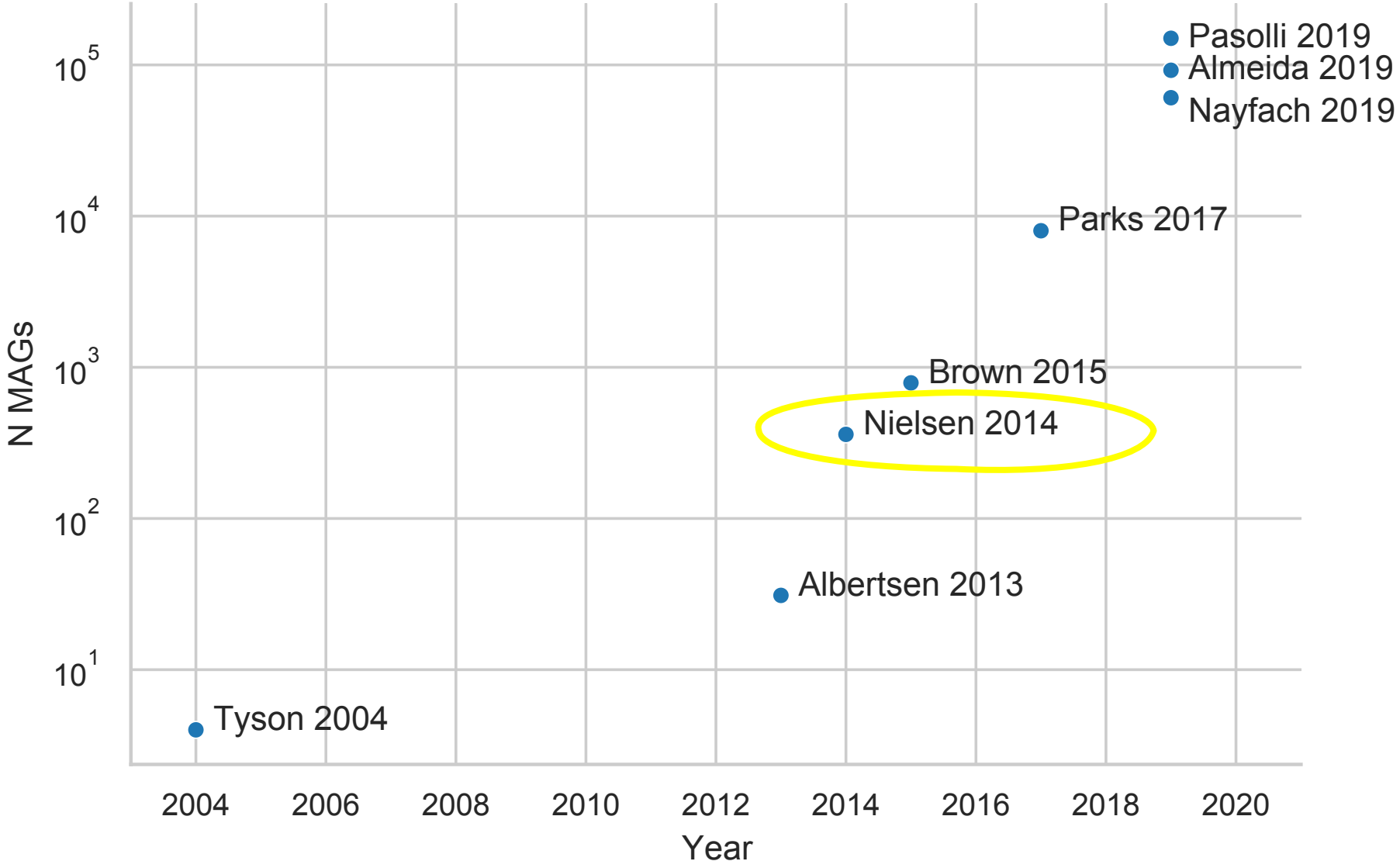


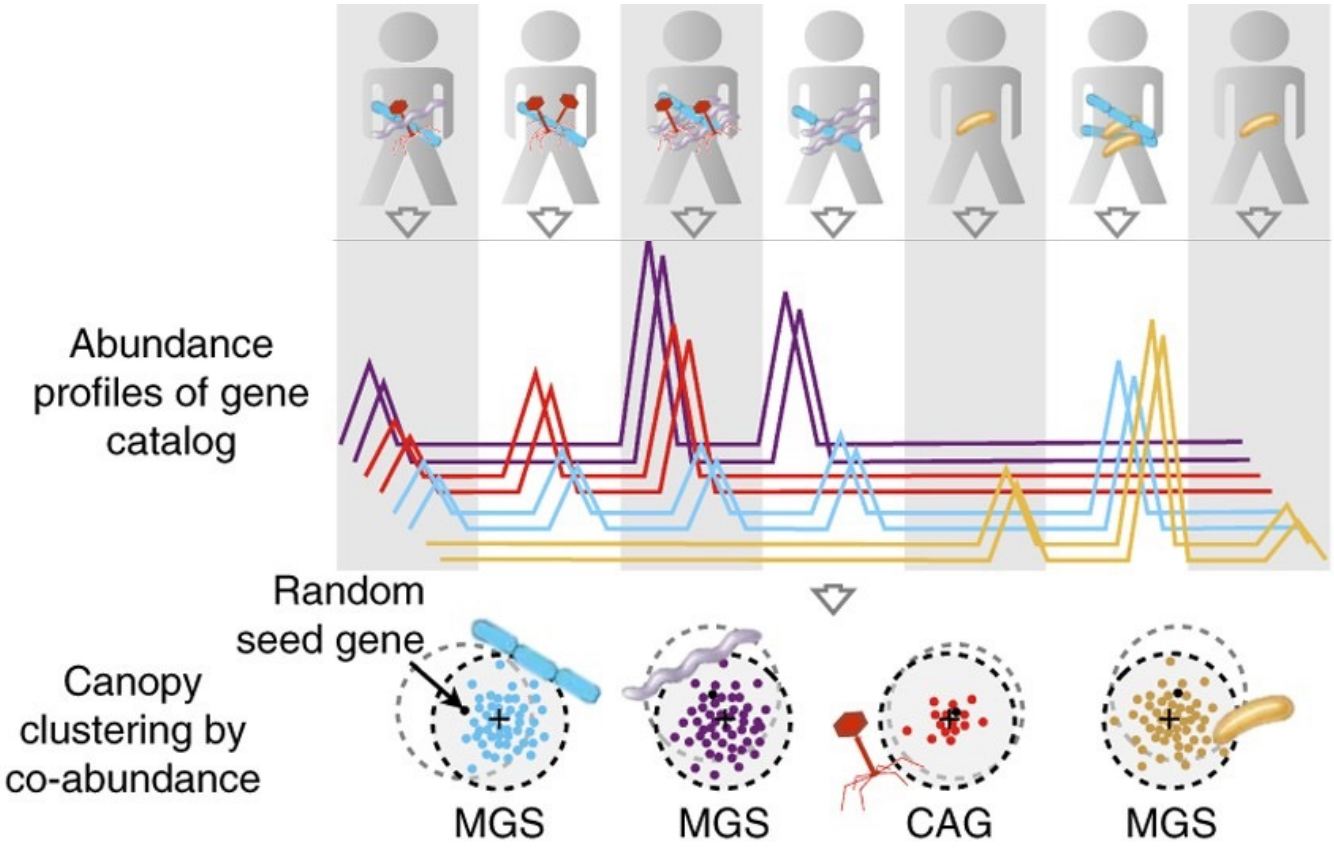

Atmosphere
No light
O_2
N_2
CO_2
Leptospirillum group II
Nickel
Phosphate
Amino acids
Dipeptide
Nitrate
ABC-type transport systems
Multi-drug efflux permeases
Efflux permeases
Sec-dependent protein translocation
Sec-independent protein translocation
Protein translocation
Capsular polysaccharide export
Voltage-gated channels
Cofactor biosynthesis
Cobalamin
Haeme
Pantothenate
Riboflavin
Thiamine
Rubisco
Carbon fixation
Embden–meyerhoff pathway
Pentose phosphate pathway
Gluconeogenesis
Catabolic functions
β-lactamase
Protease
Exopolyphosphatase
Lipases/esterase
α-amylase
Flagellum
Chemotaxis
P-type ATPases
Formate
Formate hydrogenlyase complex
F_0F_1-ATPase
Citric acid cycle
NADH
Pyrite sediment
Ferroplasma type II
ABC-type transport systems
Multidrug efflux
Multi-drug efflux permeases
Metals
Toxics
ArsB
Defence mechanisms
DNA modfication system
Superoxide dismutase
Hydrogen peroxide reductase
Cytochrome P450
Mercuric reductase
Solute
Embden–Meyerhoff pathway
Protein translocation
Sec-dependent
Sec-independent
Pentose phosphate pathway (partial)
Formate dehydrogenase
Reductive acetyl-CoA pathway
(+ THF)
(+CoA)
Carbon monoxide dehydrogenase
Acetyl-CoA synthase
Acetyl-CoA
Pyruvate
Macromolecular biosynthesis
H+ - transporting ATP synthase
Haeme-copper terminal oxidase
Sulphocyanin-mediated ferrous iron oxidation?
Osmolyte
Voltage-gated channel
Sulphite oxidase?
Red cytochrome
cyt cbb_3 oxidase
C_{553}
cyt b/FeS
Ubiquinone
bd-type oxidase
Succinate dehydrogenase
NADH: Ubiquinone dehydrogenase
Biofilm
AMD solution
42 °C / pH 0.83
$FeS_2 + 14\ Fe^{3+} + 8\ H_2O \longrightarrow 15\ Fe^{2+} + 2\ SO_4^{2-} + 16\ H^+$

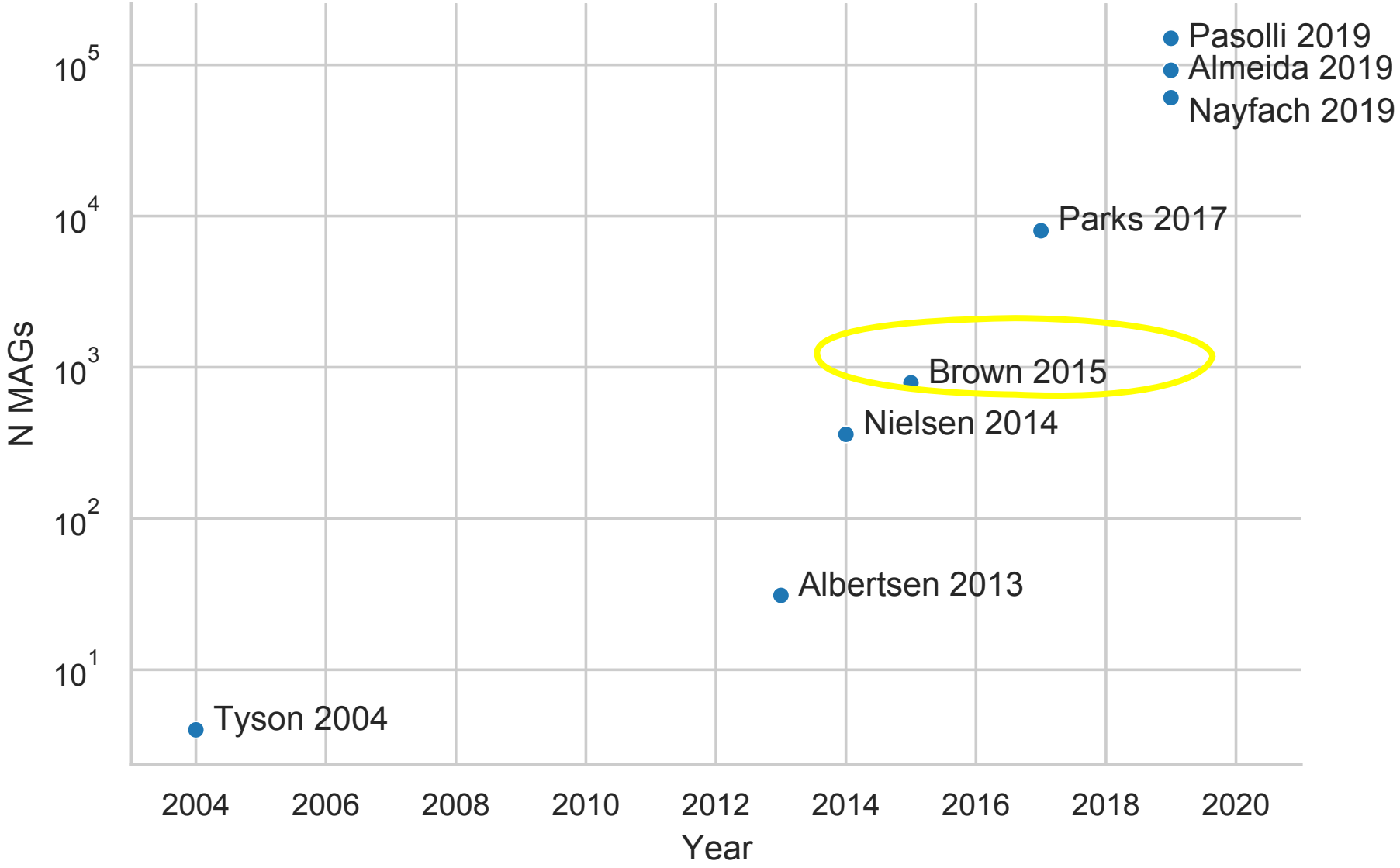

Common ancestor
Ancestral strains
Mosaic genome types

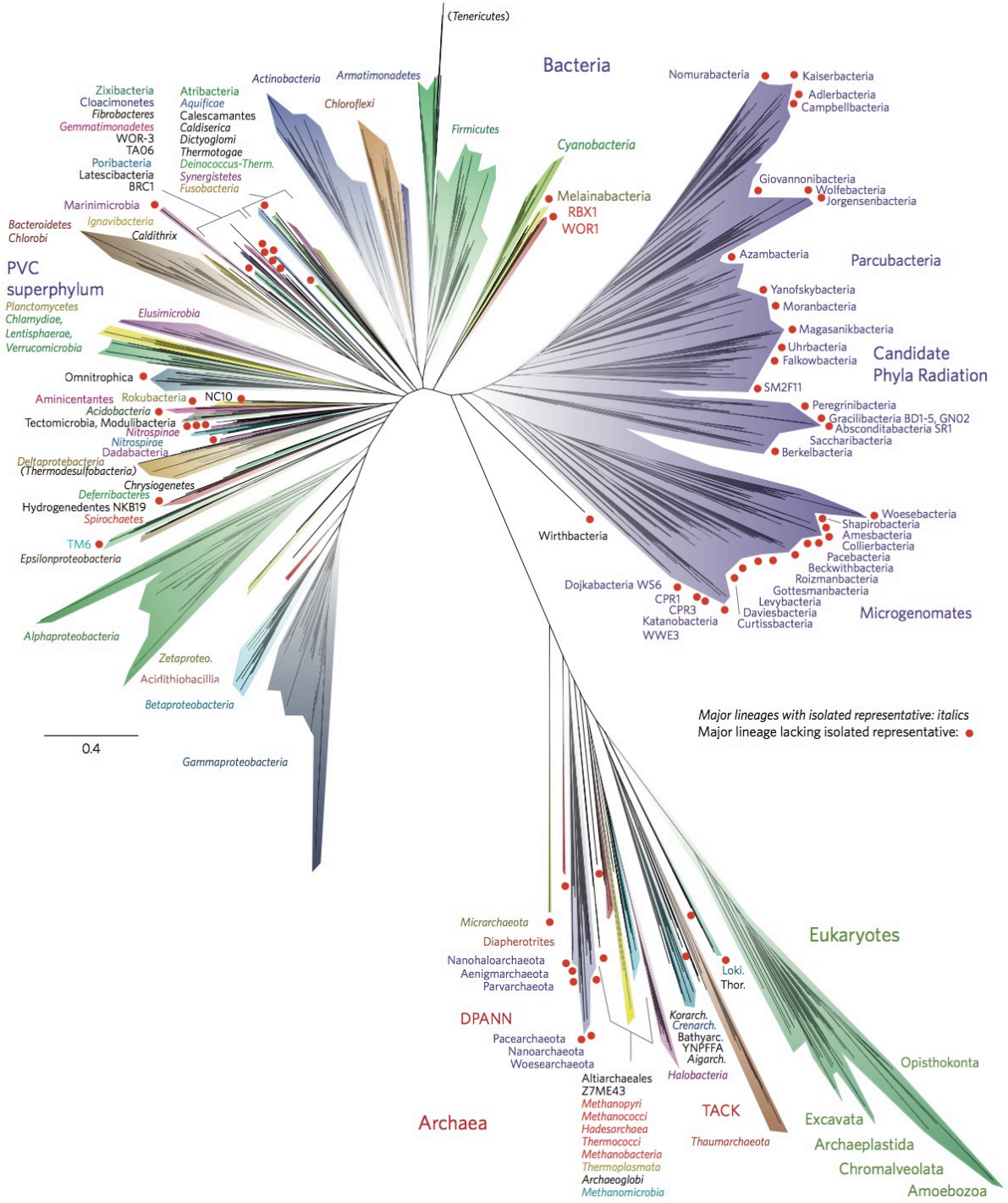

Pasolli 2019
Almeida 2019
Nayfach 2019
Parks 2017
Brown 2015
Nielsen 2014
Albertsen 2013
Tyson 2004
N MAGs
10^5
10^4
10^3
10^2
10^1
2004
2006
2008
2010
2012
2014
2016
2018
2020
Year


Length (kbp)
200
50
10
GC (%)
75
50
25
Scaffold coverage (HP+)
1,000
100
10
1
0.1
0.01
10
100
1,000
5,000
Scaffold coverage (HP−)
Proteobacteria
Bacteroidetes
Actinobacteria
TM7
Verrucomicrobia

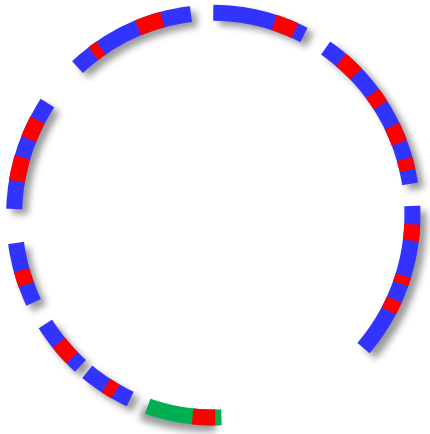

Pasolli 2019
Almeida 2019
Nayfach 2019
Parks 2017
Brown 2015
Nielsen 2014
Albertsen 2013
Tyson 2004
N MAGs
10^5
10^4
10^3
10^2
10^1
2004
2006
2008
2010
2012
2014
2016
2018
2020
Year

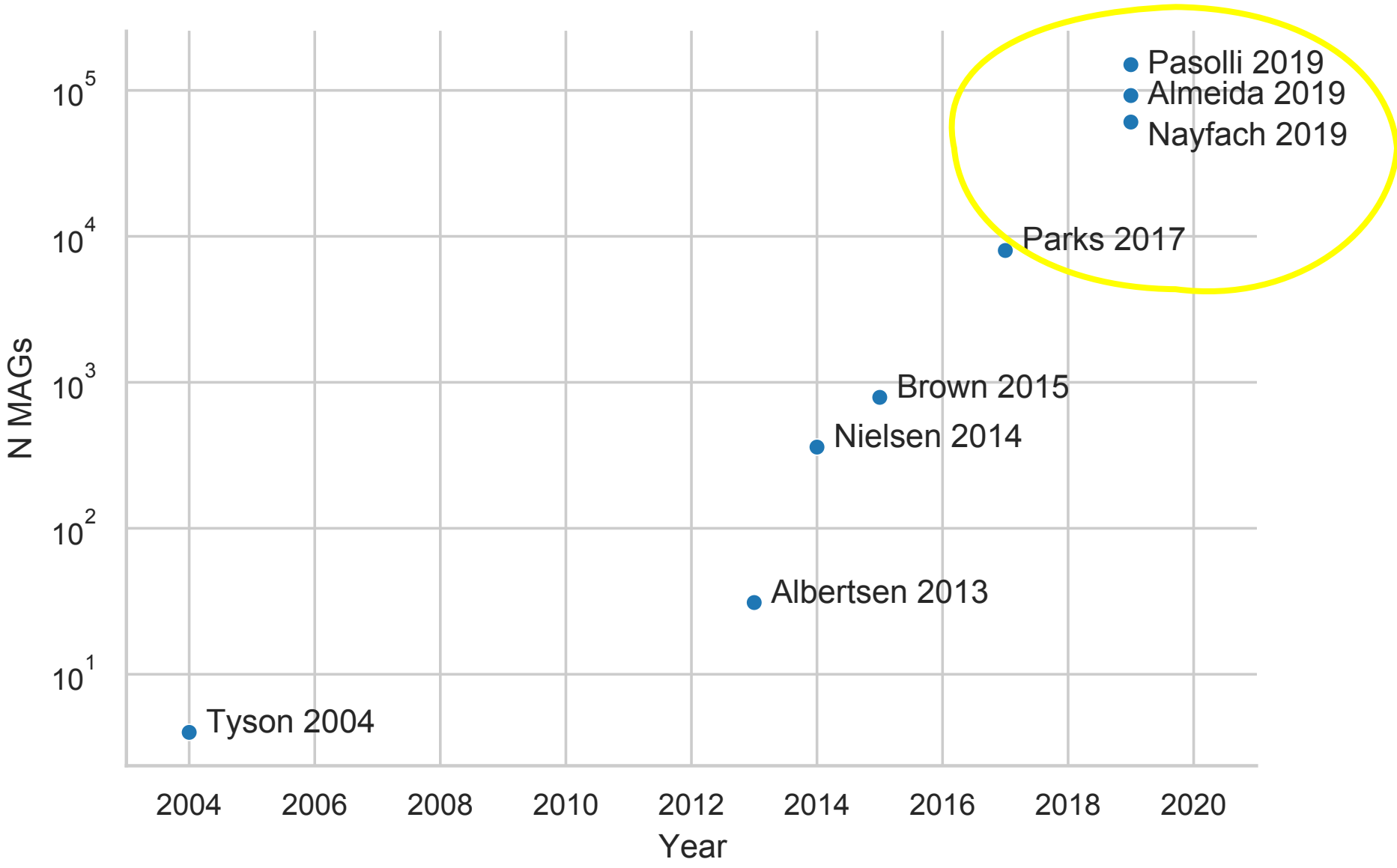

Abundance profiles of gene catalog
Random seed gene
Canopy clustering by co-abundance
MGS
MGS
CAG
MGS


Pasolli 2019
Almeida 2019
Nayfach 2019
Parks 2017
Brown 2015
Nielsen 2014
Albertsen 2013
Tyson 2004
N MAGs
10^5
10^4
10^3
10^2
10^1
2004
2006
2008
2010
2012
2014
2016
2018
2020
Year


(Tenericutes)
Bacteria
Actinobacteria
Armatimonadetes
Chloroflexi
Firmicutes
Cyanobacteria
Melainabacteria
RBX1
WOR1
Zixibacteria
Cloacimonetes
Fibrobacteres
Gemmatimonadetes
WOR-3
TA06
Poribacteria
Latescibacteria
BRC1
Atribacteria
Aquificae
Calescamantes
Caldiserica
Dictyoglomi
Thermotogae
Deinococcus-Therm.
Synergistetes
Fusobacteria
Marinimicrobia
Bacteroidetes
Chlorobi
Ignavibacteria
Caldithrix
PVC
superphylum
Planctomycetes
Chlamydiae,
Lentisphaerae,
Verrucomicrobia
Elusimicrobia
Omnitrophica
Aminicentantes
Rokubacteria
NC10
Acidobacteria
Tectomicrobia, Modulibacteria
Nitrospinae
Nitrospirae
Dadabacteria
Deltaprotebacteria
(Thermodesulfobacteria)
Chrysiogenetes
Deferribacteres
Hydrogenedentes NKB19
Spirochaetes
TM6
Epsilonproteobacteria
Alphaproteobacteria
Zetaproteo.
Acidithiobacillia
Betaproteobacteria
Gammaproteobacteria
0.4
Nomurabacteria
Kaiserbacteria
Adlerbacteria
Campbellbacteria
Giovannonibacteria
Wolfebacteria
Jorgensenbacteria
Azambacteria
Parcubacteria
Yanofskybacteria
Moranbacteria
Magasanikbacteria
Uhrbacteria
Falkowbacteria
Candidate
Phyla Radiation
SM2F11
Peregrinibacteria
Gracilibacteria BD1-5, GN02
Absconditabacteria SR1
Saccharibacteria
Berkelbacteria
Wirthbacteria
Woesebacteria
Shapirobacteria
Amesbacteria
Collierbacteria
Pacebacteria
Beckwithbacteria
Roizmanbacteria
Gottesmanbacteria
Levybacteria
Daviesbacteria
Curtissbacteria
Microgenomates
Dojkabacteria WS6
CPR1
CPR3
Katanobacteria
WWE3
Major lineages with isolated representative: italics
Major lineage lacking isolated representative:
Micrarchaeota
Diapherotrites
Nanohaloarchaeota
Aenigmarchaeota
Parvarchaeota
DPANN
Pacearchaeota
Nanoarchaeota
Woesearchaeota
Loki.
Thor.
Korarch.
Crenarch.
Bathyarc.
YNPFFA
Aigarch.
Halobacteria
Altiarchaeales
Z7ME43
Methanopyri
Methanococci
Hadesarchaea
Thermococci
Methanobacteria
Thermoplasmata
Archaeoglobi
Methanomicrobia
Archaea
TACK
Thaumarchaeota
Eukaryotes
Opisthokonta
Excavata
Archaeplastida
Chromalveolata
Amoebozoa

Patescibacteria don't have all universal marker genes!




Pasolli 2019
Almeida 2019
Nayfach 2019
Parks 2017
Brown 2015
Nielsen 2014
Albertsen 2013
Tyson 2004
N MAGs
10^5
10^4
10^3
10^2
10^1
2004
2006
2008
2010
2012
2014
2016
2018
2020
Year

Large-scale re-assembly era

nature microbiology

ARTICLES

DOI: 10.1038/s41564-017-0012-7

OPEN

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/s41587-020-0603-3

OPEN

A unified catalog of 204,938 reference genomes from the human gut microbiome

Post-assembly era

Analyze a metagenome