How to define a prokaryotic species

Joint Graduate Seminar

Supervisor: Prof. Margaret Ip
Student: LI, Jie
1st year PhD candidate

Date: 2021 December 14

Department: Microbiology

Outline

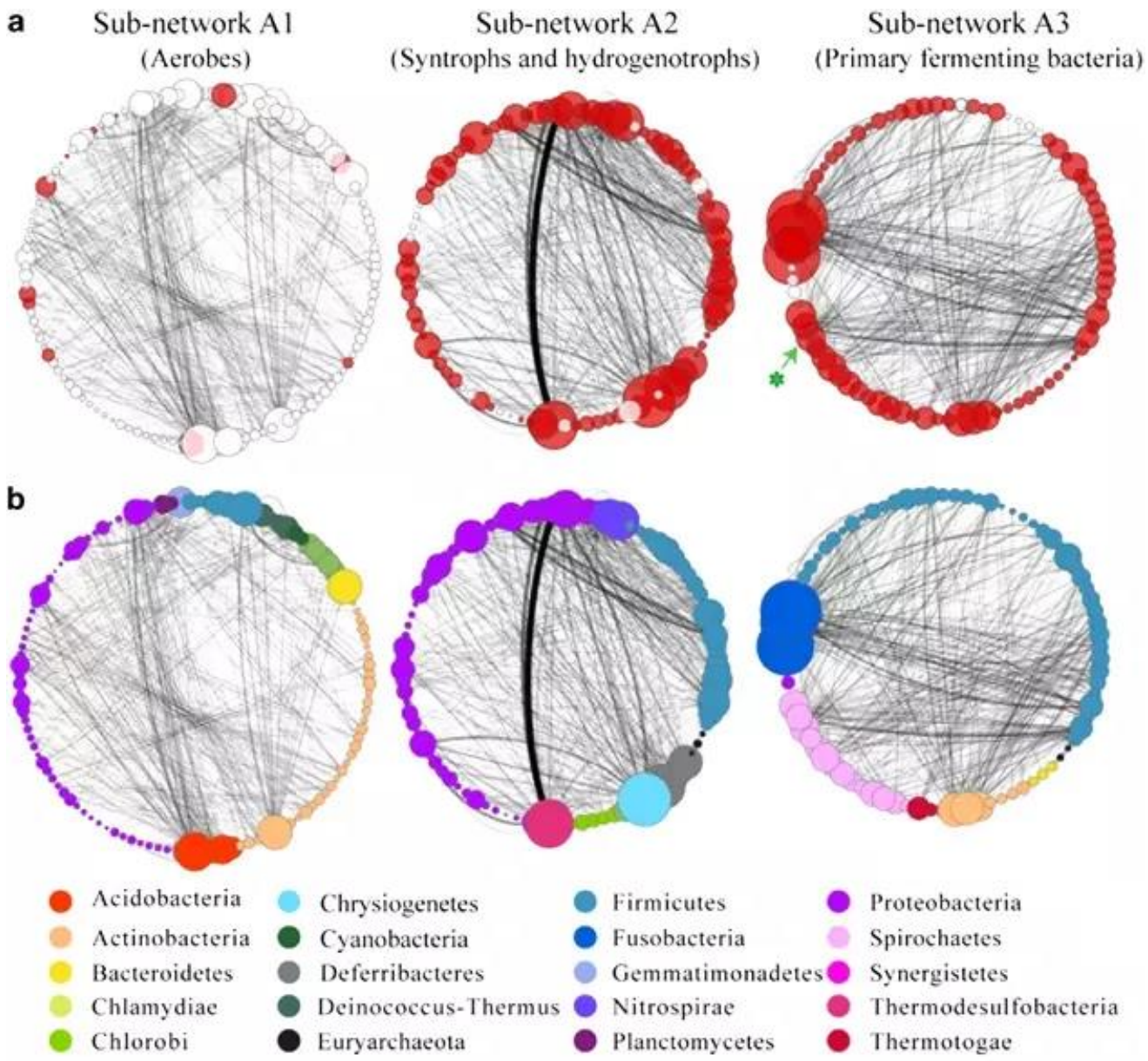
01 Challenges of applying the classical definition to prokaryotes

02 Practical importance and operational definitions

03 Concepts considering speciation mechanisms

04 Defining species and populations under the scope of microbial speciation

Challenges of applying the classical definition to prokaryotes



- Genetic exchange network among distantly related taxa
- Edges: the number of transferred genes
- Nodes: the genomes from different phyla

Caro-Quintero et al. ISME J. 2015.

Practical importance and operational definitions

Necessity:

- Clinical settings
- Agriculture
- Food industry

...

16S rRNA gene
Cutoff: 97%[1]

Average Nucleotide Identity (ANI)
Cutoff: 95% [1]

A set of **single copy conserved genes** [2]

Since 1960s, the gold standard:
DNA-DNA binding
Cutoff: 70% [1]

Polyphasic approach:
Combination of genotypic (DNA fingerprinting), phenotypic (biochemical tests, fatty-acid composition) properties and phylogenetic information (rRNA gene sequences)

[1] Goris et al. Int J Syst Evol Micrb. 2007
[2] Truong et al. Nat. Methods. 2015

Practical importance and operational definitions

nature biotechnology

RESOURCE

https://doi.org/10.1038/s41587-020-0501-8

Check for updates

A complete domain-to-species taxonomy for Bacteria and Archaea

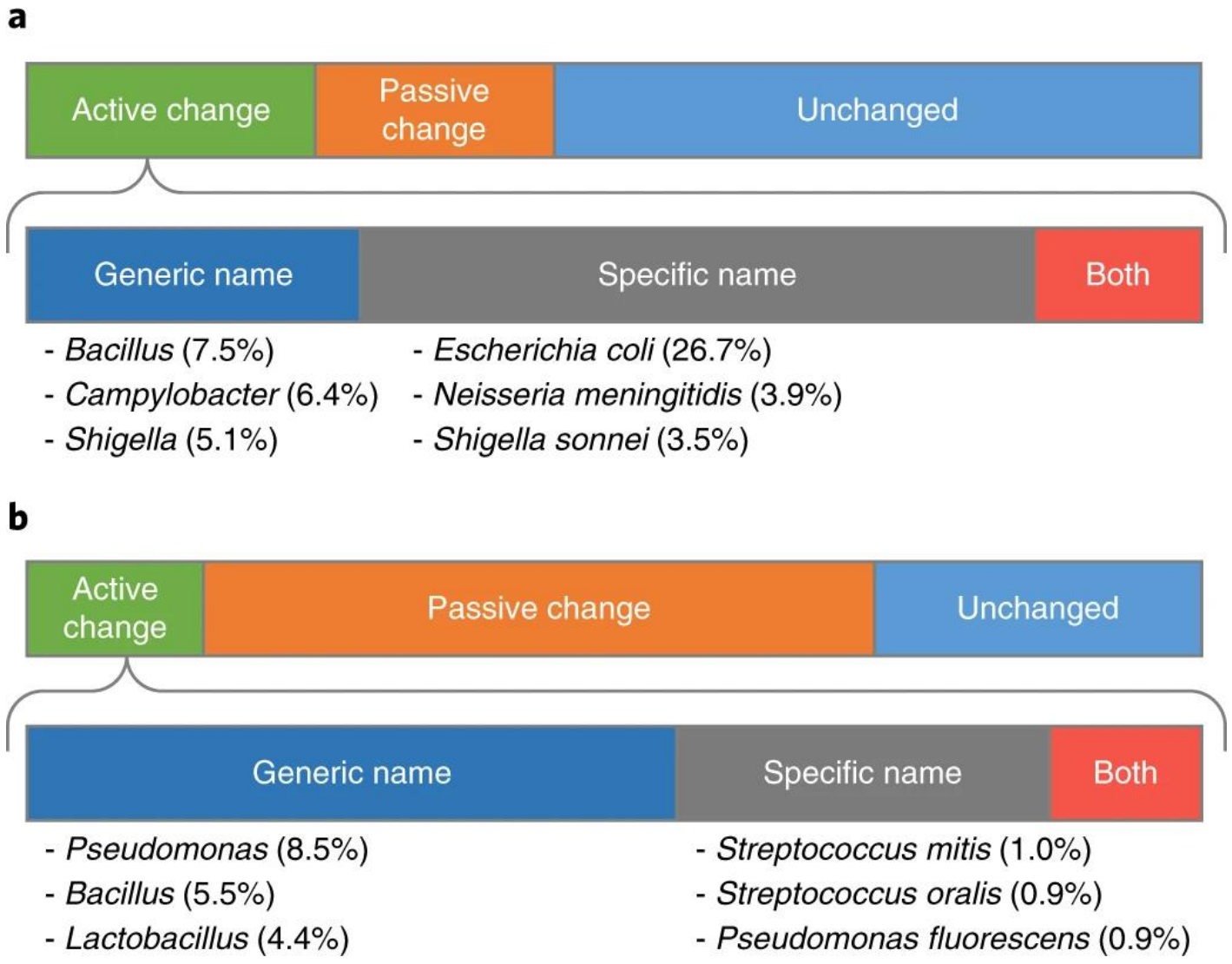
Donovan H. Parks ✉, Maria Chuvochina, Pierre-Alain Chaumeil, Christian Rinke, Aaron J. Mussig and Philip Hugenholtz

The Genome Taxonomy Database is a phylogenetically consistent, genome-based taxonomy that provides rank-normalized classifications for ~150,000 bacterial and archaeal genomes from domain to genus. However, almost 40% of the genomes in the Genome Taxonomy Database lack a species name. We address this limitation by using commonly accepted average nucleotide identity criteria to set bounds on species and propose species clusters that encompass all publicly available bacterial and archaeal genomes. Unlike previous average nucleotide identity studies, we chose a single representative genome to serve as the effective nomenclatural 'type' defining each species. Of the 24,706 proposed species clusters, 8,792 are based on published names. We assigned placeholder names to the remaining 15,914 species clusters to provide names to the growing number of genomes from uncultivated species. This resource provides a complete domain-to-species taxonomic framework for bacterial and archaeal genomes, which will facilitate research on uncultivated species and improve communication of scientific results.

- Pairwise ANI comparison against representative genomes
- Greedy clustering for de novo species clusters

Parks et al. Nat. Biotechnol. 2020

Results of the nomenclature for the unassigned genomes



Unchanged: the genome's binomial species name was identical in the NCBI taxonomy

Passively change: the NCBI taxonomy did not have a species assignment

Active change:

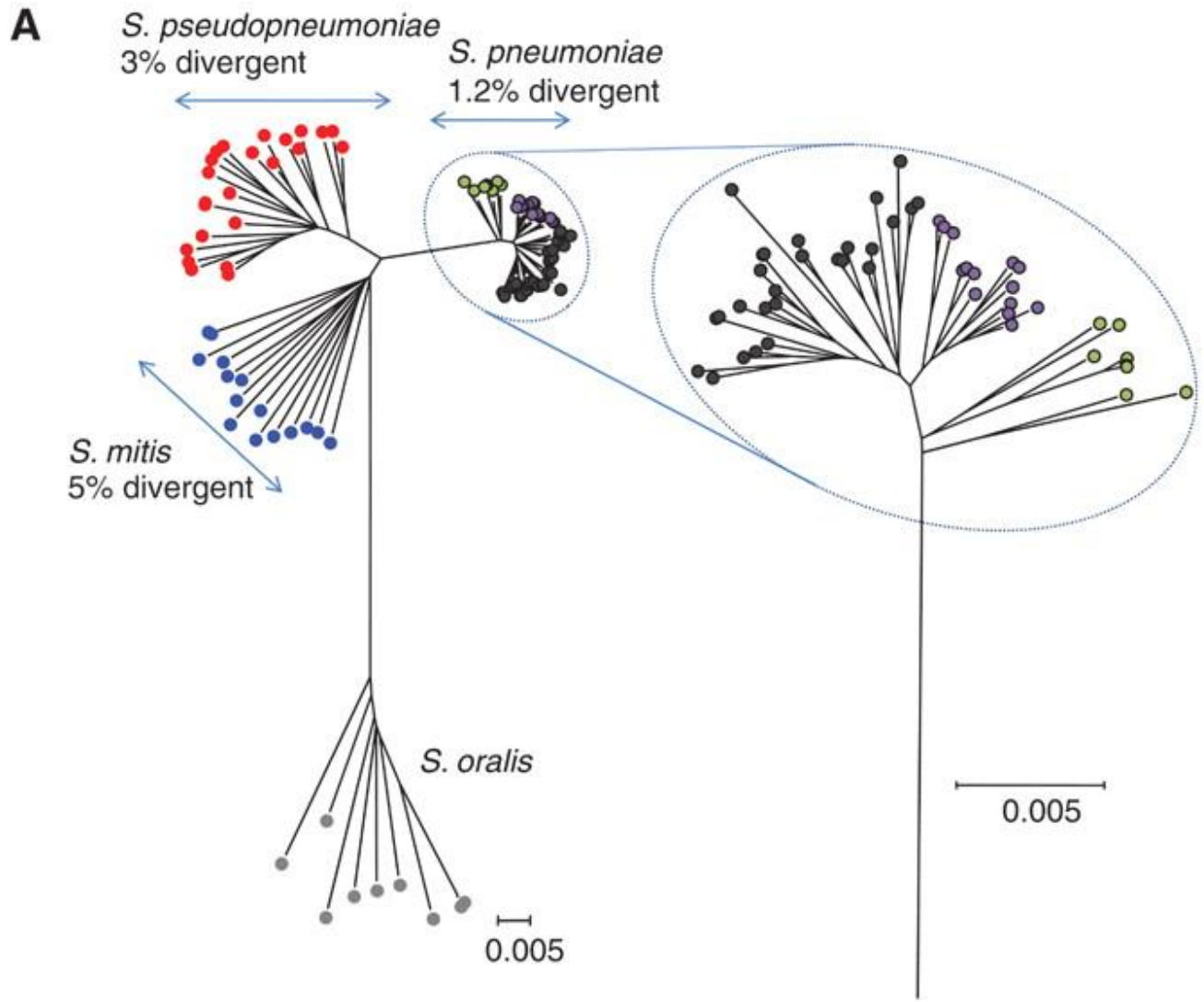
(1) Generic name: change in only the generic name of the species
(2) Specific name: change in only the specific name of the species
(3) Both: changes in both the generic and specific names

Fig. Results are shown for the 143,566 genomes (a) and 24,080 species representatives (b) with an NCBI taxonomic assignment.

Problems of the operational definitions

Fig. The evolution tree of 97 isolates of four Streptococcus species. The tree was built using concatenates of six housekeeping loci.

S. pneumoniae: a major human pathogen
S. mitis: a commensal bacteria
S. pseudopneumoniae: a recently described organism of uncertain status



Fraser et al. Science. 2009

Concepts considering speciation mechanisms

- The Ecological Species Concept (ESC): speciation is driven by natural selection

- The Biological Species Concept (BSC): speciation is driven by barriers to genetic exchange, which means rates of gene flow (recombination) are high within than between species, but cross-species gene transfer can still occur.

Shapiro et al. Cold Spring Harb Perspect Biol. 2015.

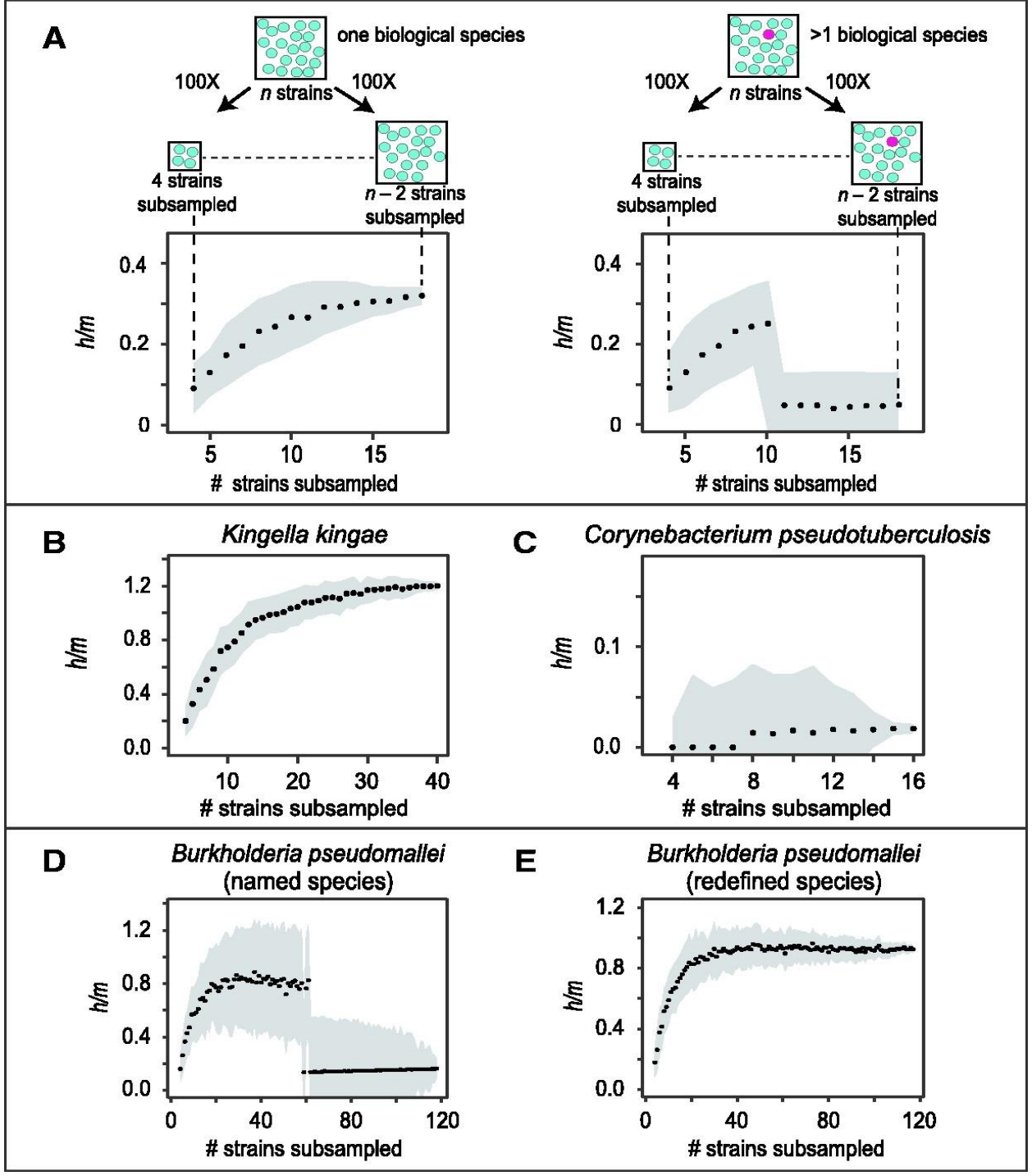
Redefining prokaryotic species by quantifying the gene flow (BSC-like)

h/m: the ratio of homoplasic to nonhomoplasic alleles

Rationale:

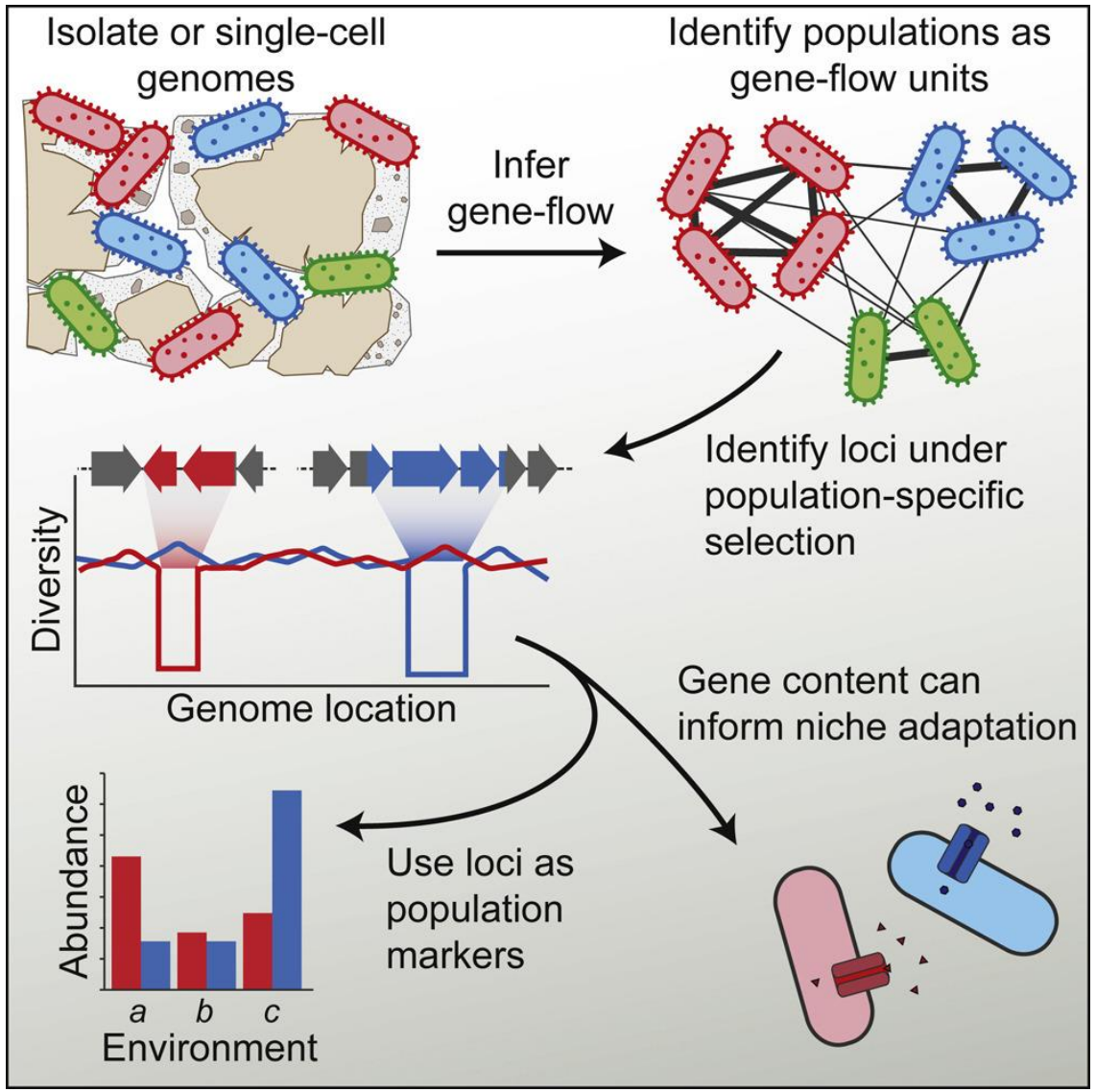
(1) If a group of strains are from one biological species, the gene flow will be coherent.

(2) If there are any strains from different species, a break of gene flow will occur.



Bobay et al. Genome Biol. Evol. 2017

Defining bacterial populations considering the speciation mechanisms



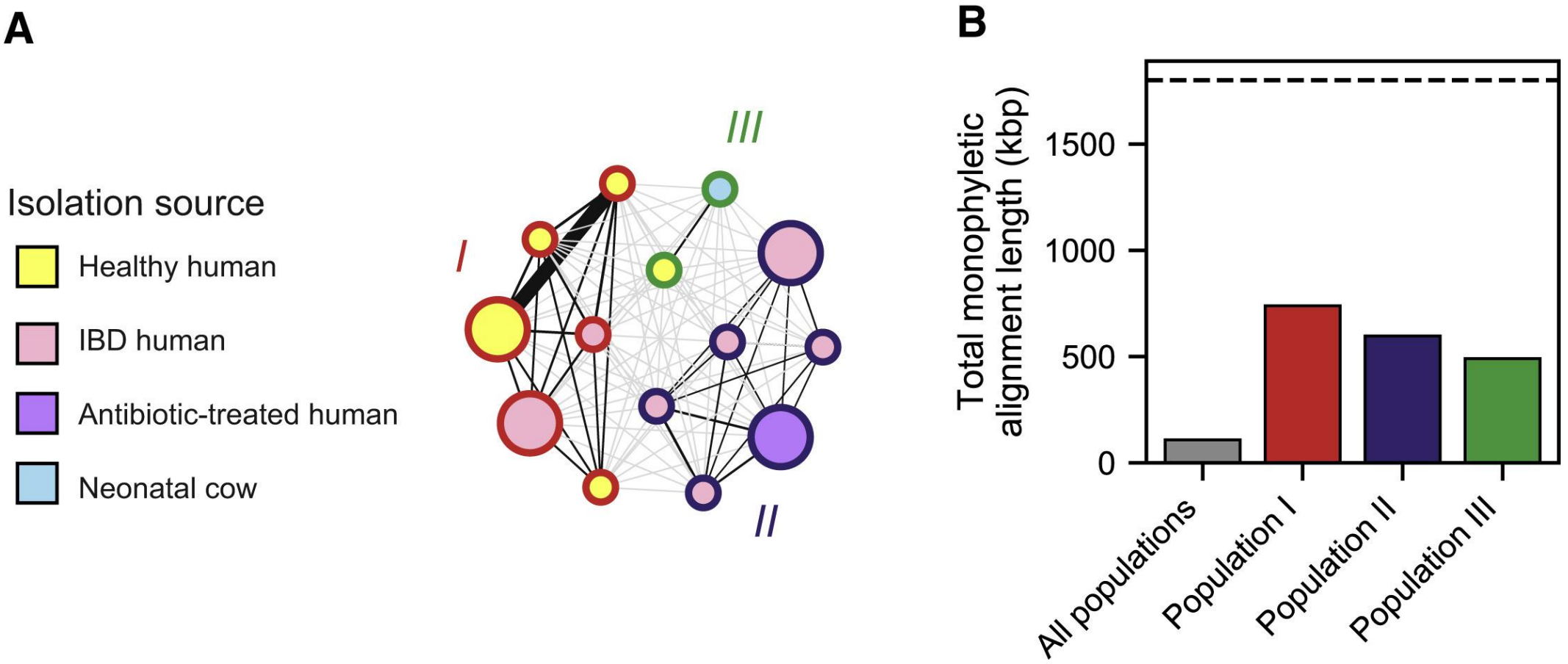
Species: gene flow units

Populations: locally co-existing members of a species that are reproductively isolated

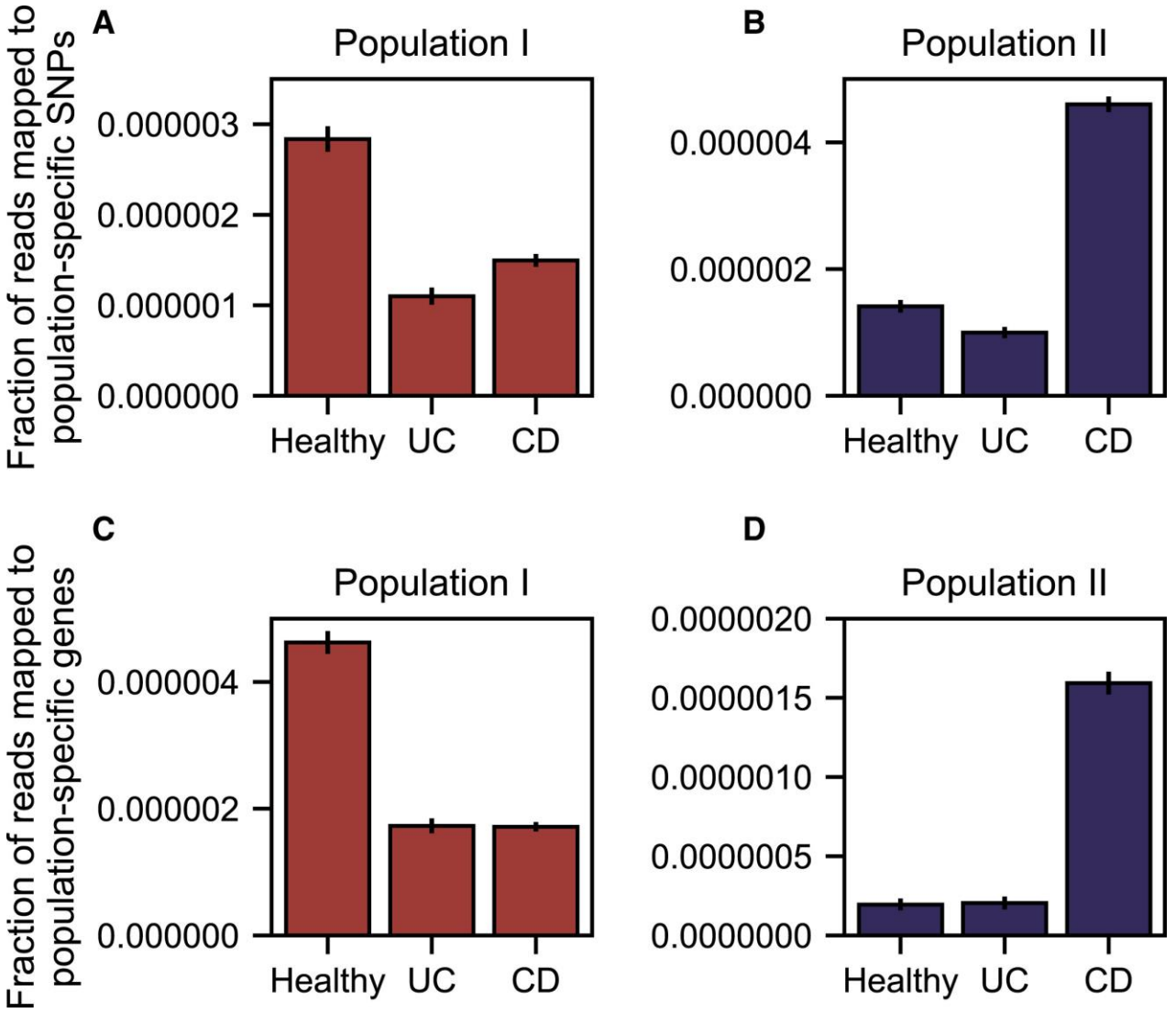
Recent research suggested that distinct populations can be identified based on cohesive ecological dynamics (natural selection) and preferential gene flow (recombination).

Arevalo et al. Cell. 2019.

Fine Scale Population Structure Is Evident in R. gnavus Populations



Verification of the R. gnavus Populations with metagenomic reads coverage



Concluding remarks

- Ordering microbial diversity into ecologically and genetically cohesive units is both theoretically and practically necessary

- Without considering the speciation mechanisms, scientists demarcate the prokaryotes into species according to the genomic similarities.

- Defining prokaryotic species and populations considering speciation mechanisms