原核生物系统学与物种形成

黄英

中国科学院微生物研究所

微生物资源前期开发国家重点实验室

系统学 (Systematics)

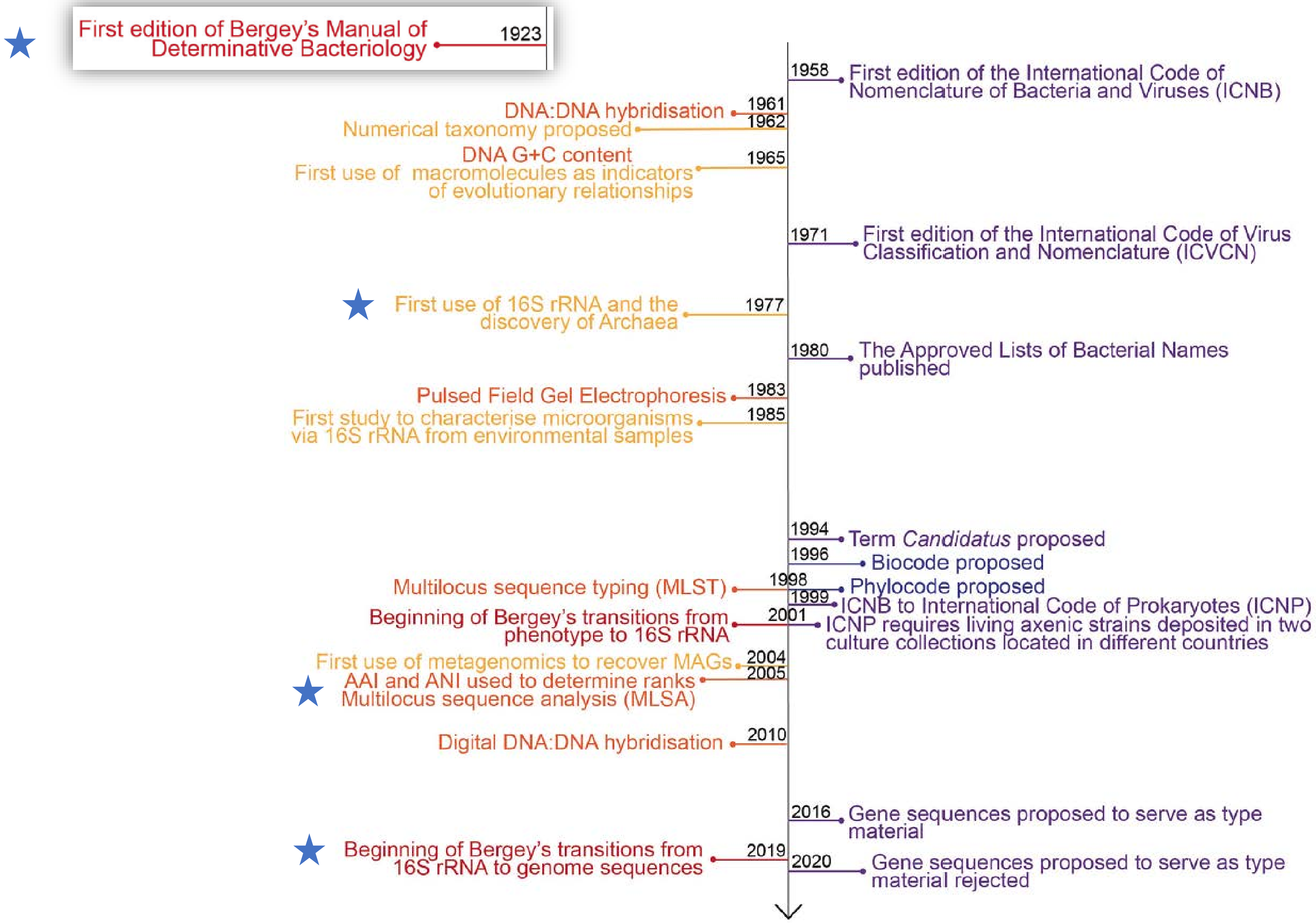
1. 分类 Classification
2. 描述 Description
3. 鉴定 Identification
4. 命名 Nomenclature

分类学 (Taxonomy)

5. 理解生物多样性 Comprehending biodiversity
   - Evolutionary history
   - Ecological adaptation

“The beginning of wisdom is calling things by their right names”
(名正言顺)

原核生物分类命名百年历程



★ 1923 First edition of Bergey's Manual of Determinative Bacteriology

1958 First edition of the International Code of Nomenclature of Bacteria and Viruses (ICNB)

1961 DNA:DNA hybridisation

1962 Numerical taxonomy proposed

1965 DNA G+C content
First use of macromolecules as indicators of evolutionary relationships

1971 First edition of the International Code of Virus Classification and Nomenclature (ICVCN)

★ 1977 First use of 16S rRNA and the discovery of Archaea

1980 The Approved Lists of Bacterial Names published

1983 Pulsed Field Gel Electrophoresis

1985 First study to characterise microorganisms via 16S rRNA from environmental samples

1994 Term *Candidatus* proposed

1996 Biocode proposed

1998 Multilocus sequence typing (MLST)
Phylocode proposed

1999 ICNB to International Code of Prokaryotes (ICNP)

2001 Beginning of Bergey's transitions from phenotype to 16S rRNA
ICNP requires living axenic strains deposited in two culture collections located in different countries

2004 First use of metagenomics to recover MAGs

★ 2005 AAI and ANI used to determine ranks
Multilocus sequence analysis (MLSA)

2010 Digital DNA:DNA hybridisation

2016 Gene sequences proposed to serve as type material

★ 2019 Beginning of Bergey's transitions from 16S rRNA to genome sequences

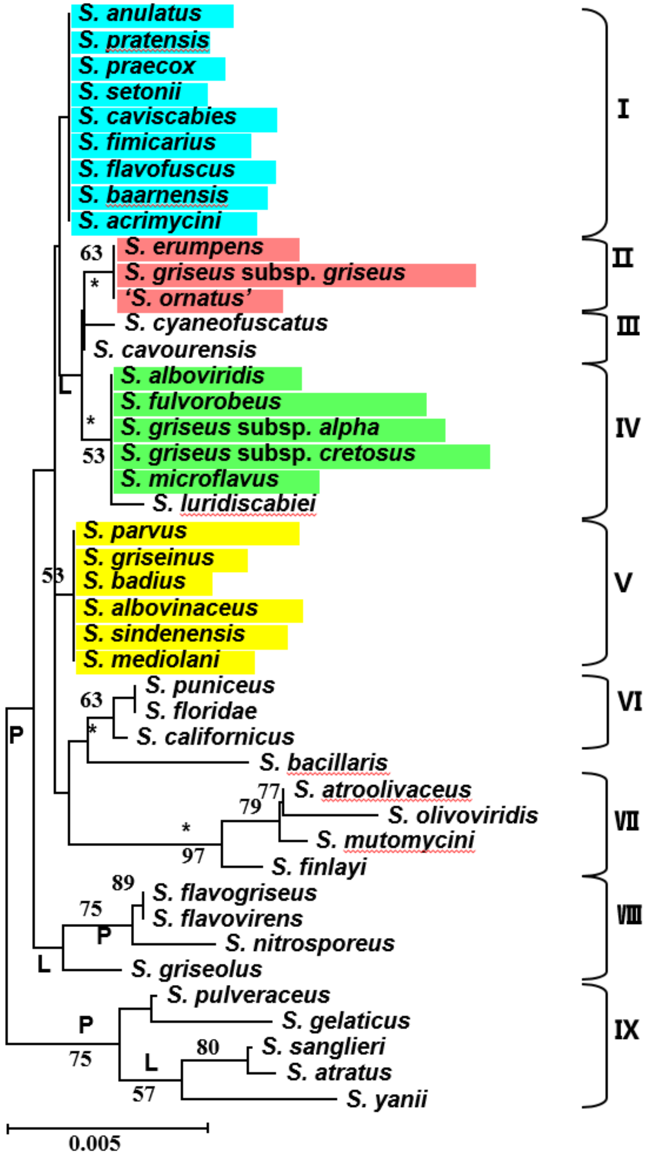
2020 Gene sequences proposed to serve as type material rejected

Hugenholtz, et al.
ISME J, 2021

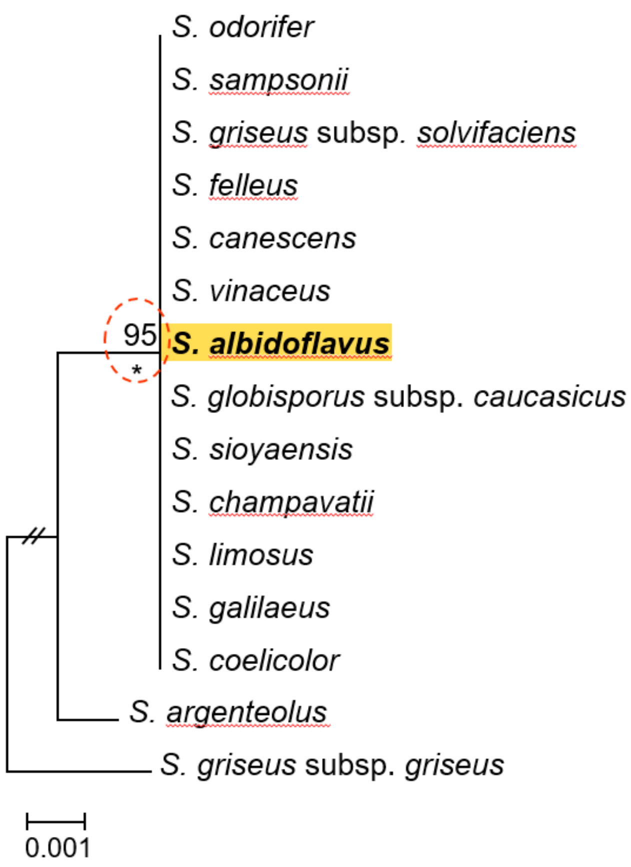
16S rRNA基因对一些原核物种的分辨率不足

- **16S rRNA基因对链霉菌种间和种内的分辨率均不足**

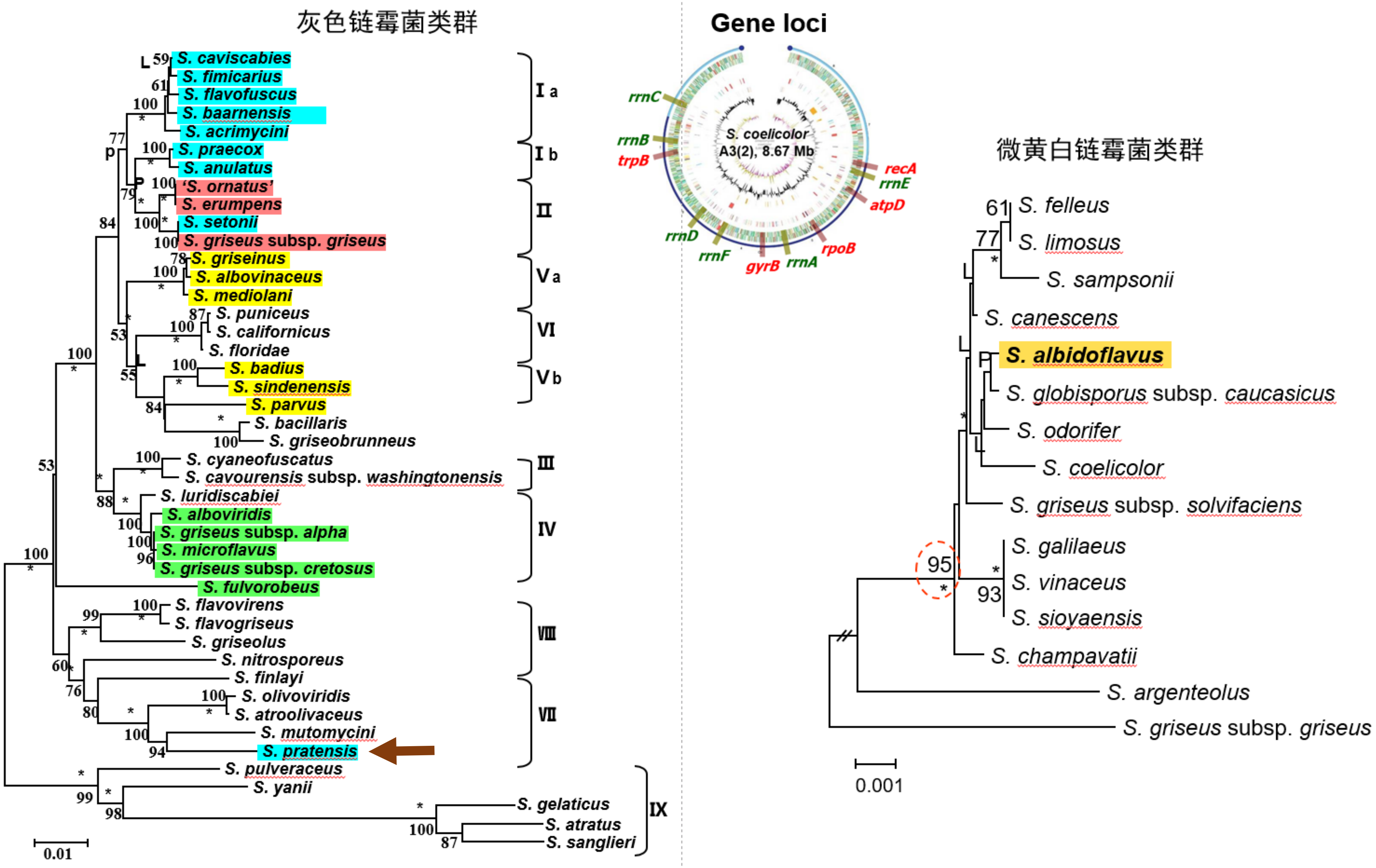
灰色链霉菌类群



微黄白链霉菌类群



多位点序列分析(MLSA)有效区分原核物种



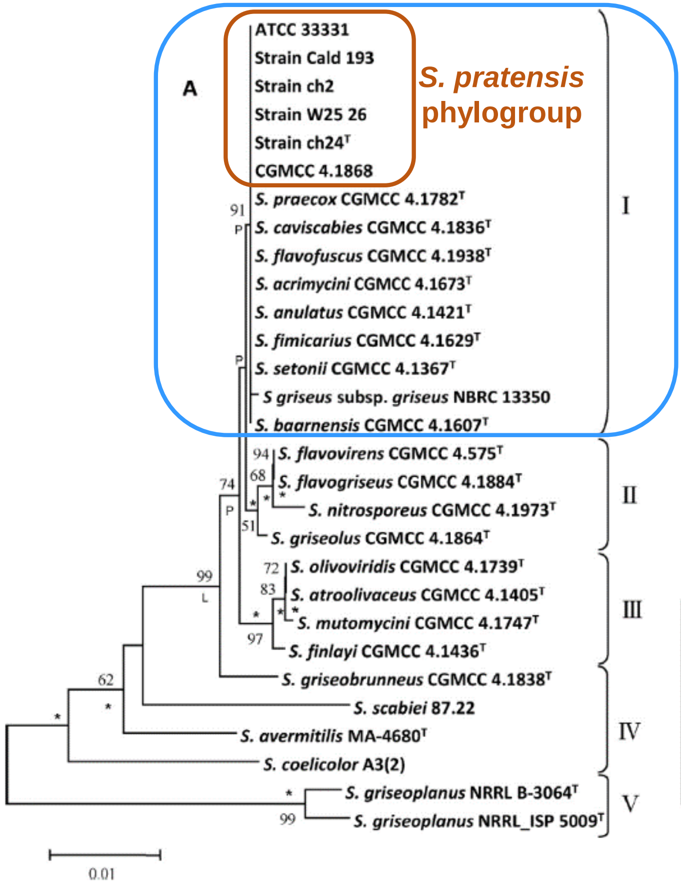
灰色链霉菌类群

Gene loci

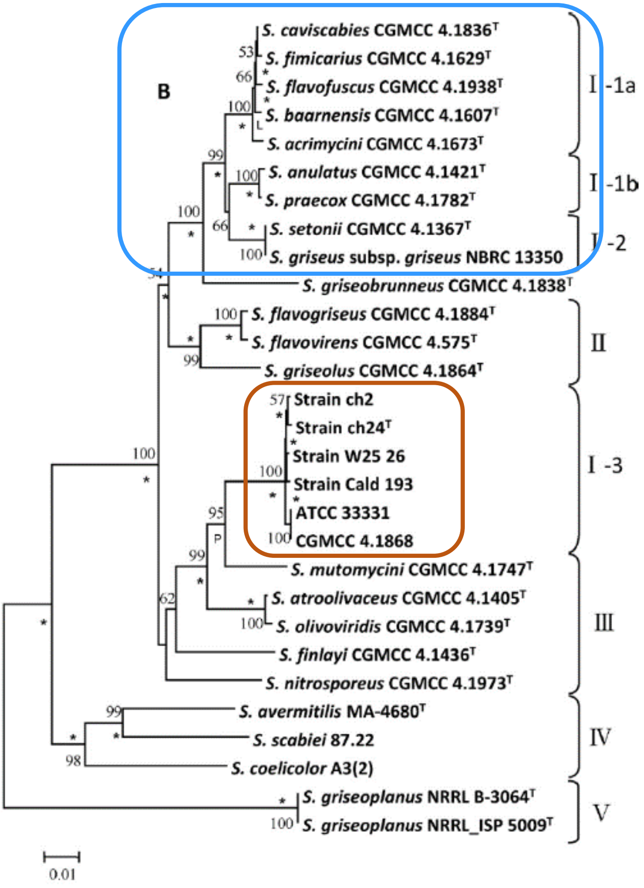
微黄白链霉菌类群

IJSEM, 2008, 2011
SAM, 2009, 2012

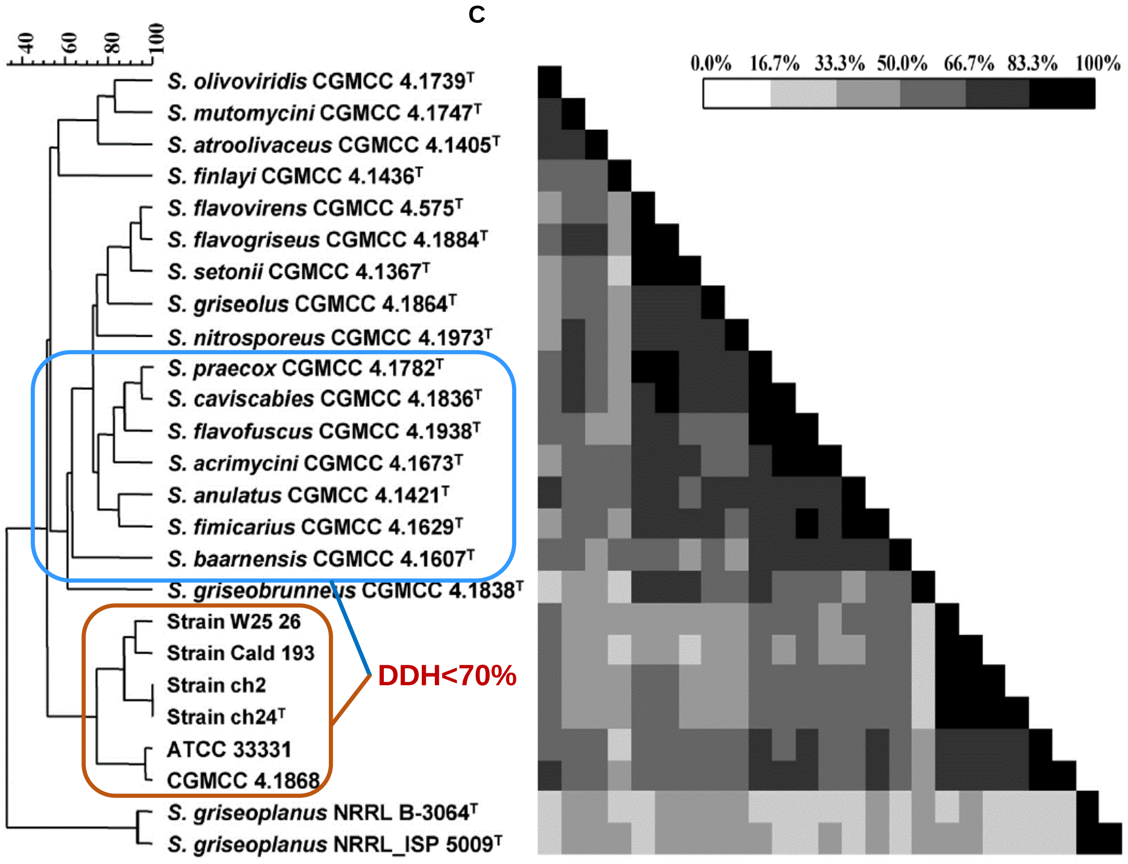
多位点序列分析(MLSA)有效区分原核物种



16S tree



5-gene MLSA tree

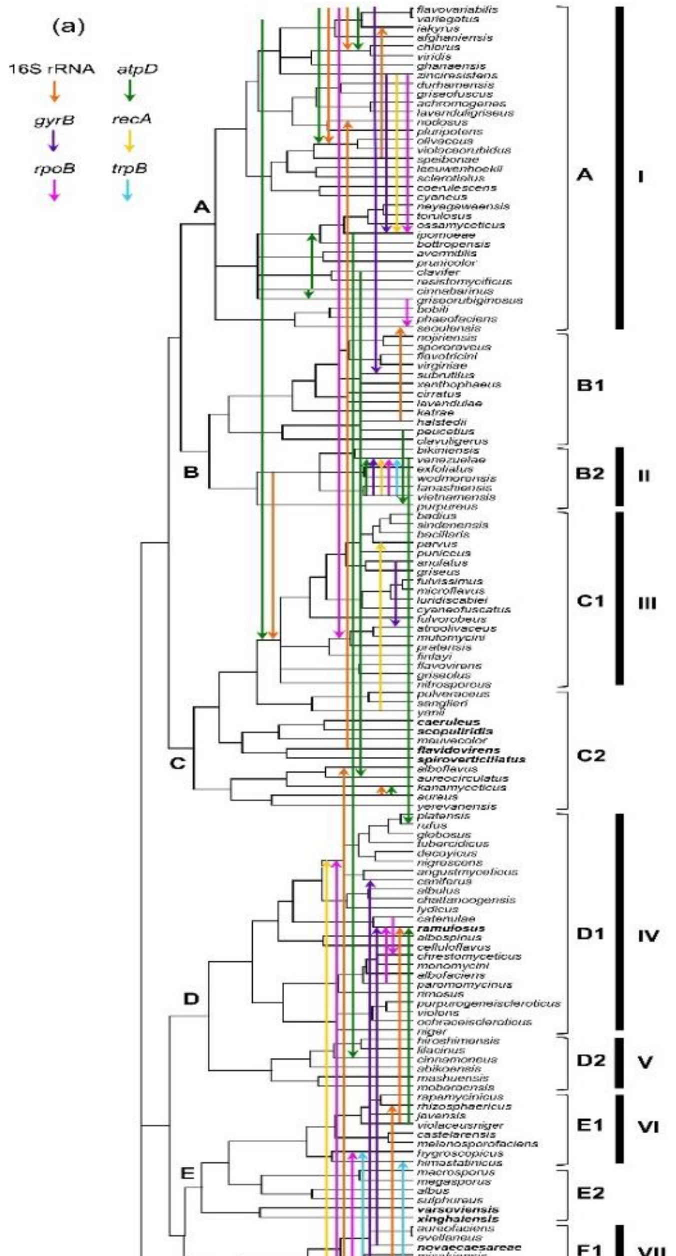


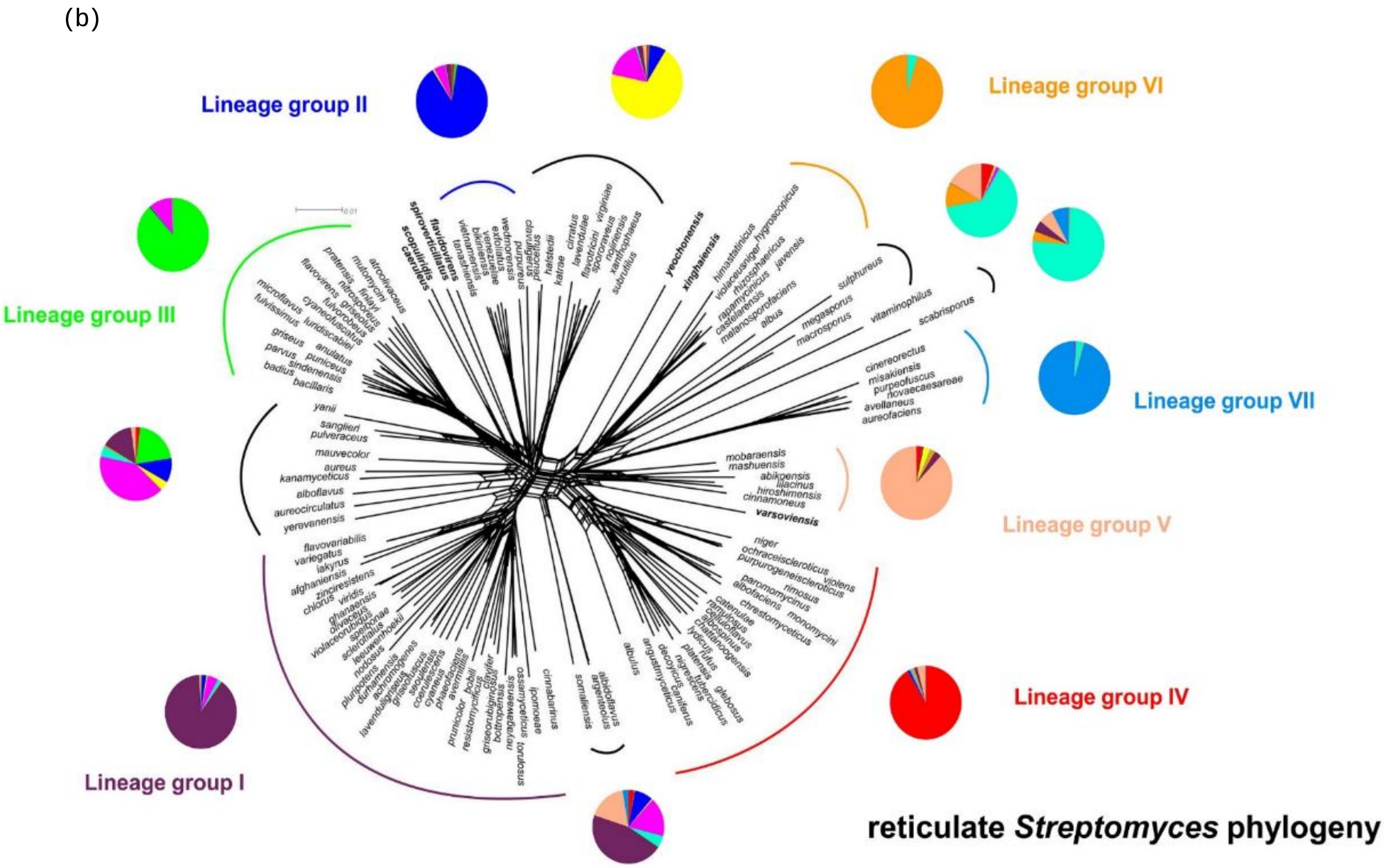
49-phenotype clustering

◆ **MLSA的结果得到表型数值分类和DDH的支持**

Syst Appl Microbiol, 2013
Method Microbiol, 2014

链霉菌的同源重组和网状进化





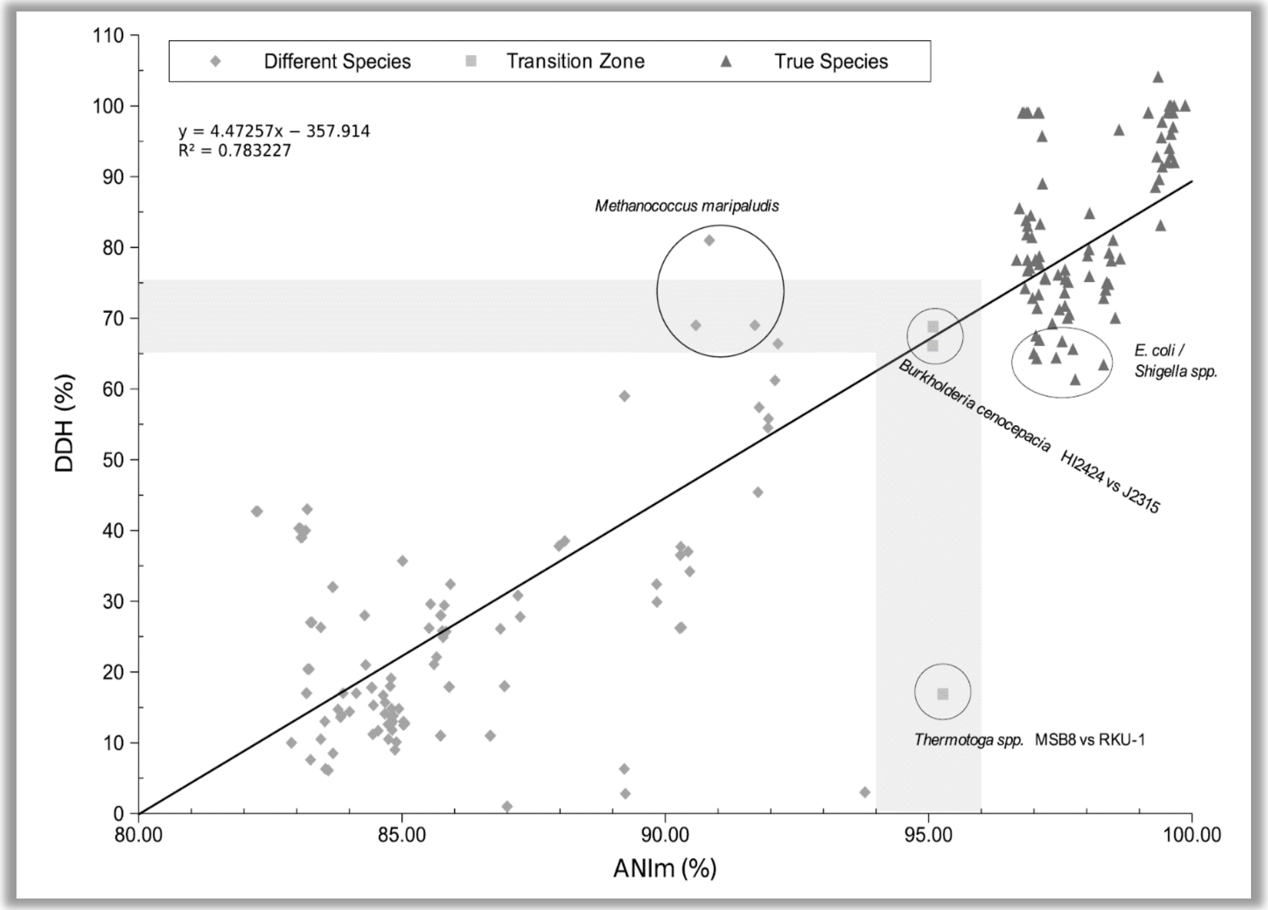
reticulate *Streptomyces* phylogeny

- 基因横向转移
- ~40%谱系受影响

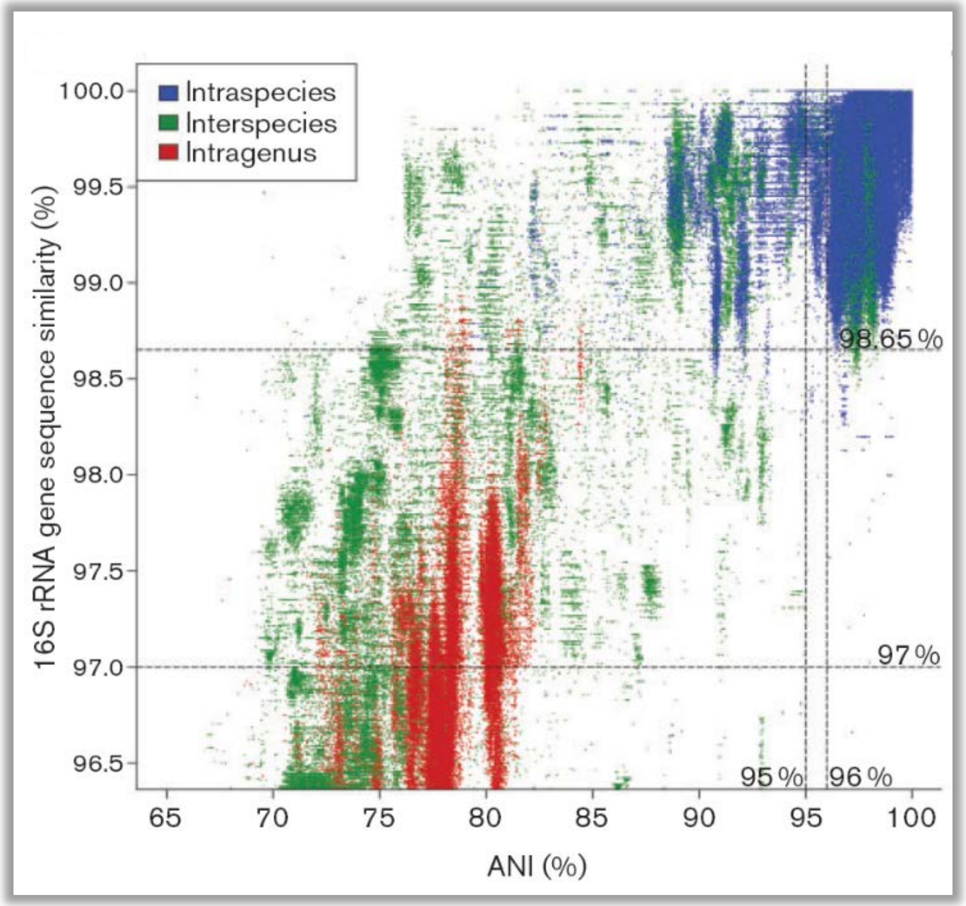
Cheng et al. Mol Phylogenet Evol, 2016

全基因组序列应用于原核生物物种界定

- **Genome-based average nucleotide identity (ANI) 与DDH呈线性相关，ANI=95-96%可区分绝大部分已知物种；绝大部分ANI>96%的菌株间16S相似性>98.65%。**



Shifting the genomic gold standard for the prokaryotic species definition. PNAS, 2009.

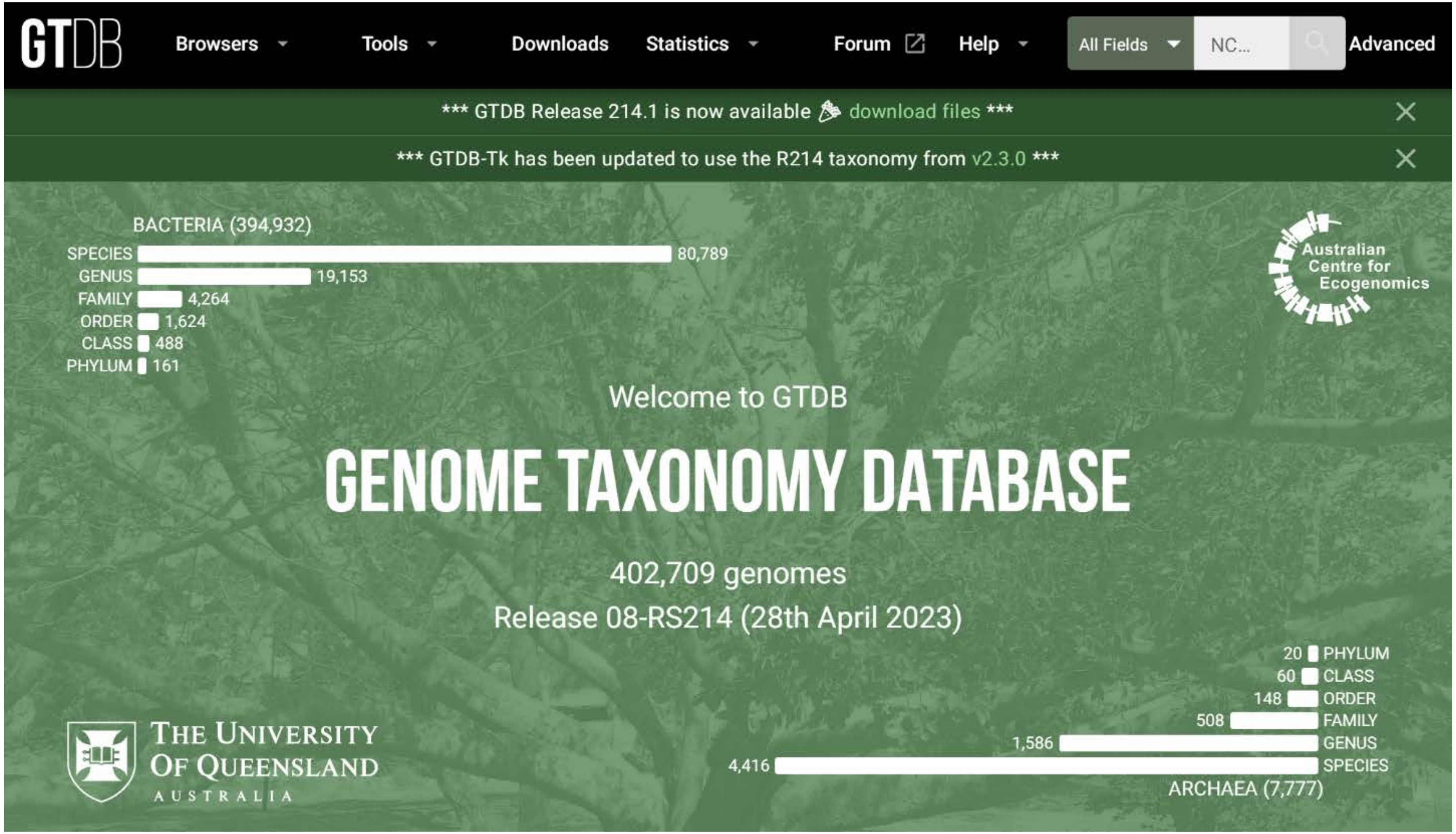


Towards a taxonomic coherence between ANI and 16S … IJSEM, 2014.

基于全基因组序列的原核生物系统分类平台--GTDB

- The **GTDB** taxonomy is based on genome trees inferred using FastTree from an aligned concatenated set of **120 single copy marker proteins for Bacteria (bac120)**,
- and with IQ-TREE from a concatenated set of **53 or 122 marker proteins for Archaea**.

https://gtdb.ecogenomic.org/



*** GTDB Release 214.1 is now available download files ***

*** GTDB-Tk has been updated to use the R214 taxonomy from v2.3.0 ***

Welcome to GTDB

GENOME TAXONOMY DATABASE

402,709 genomes

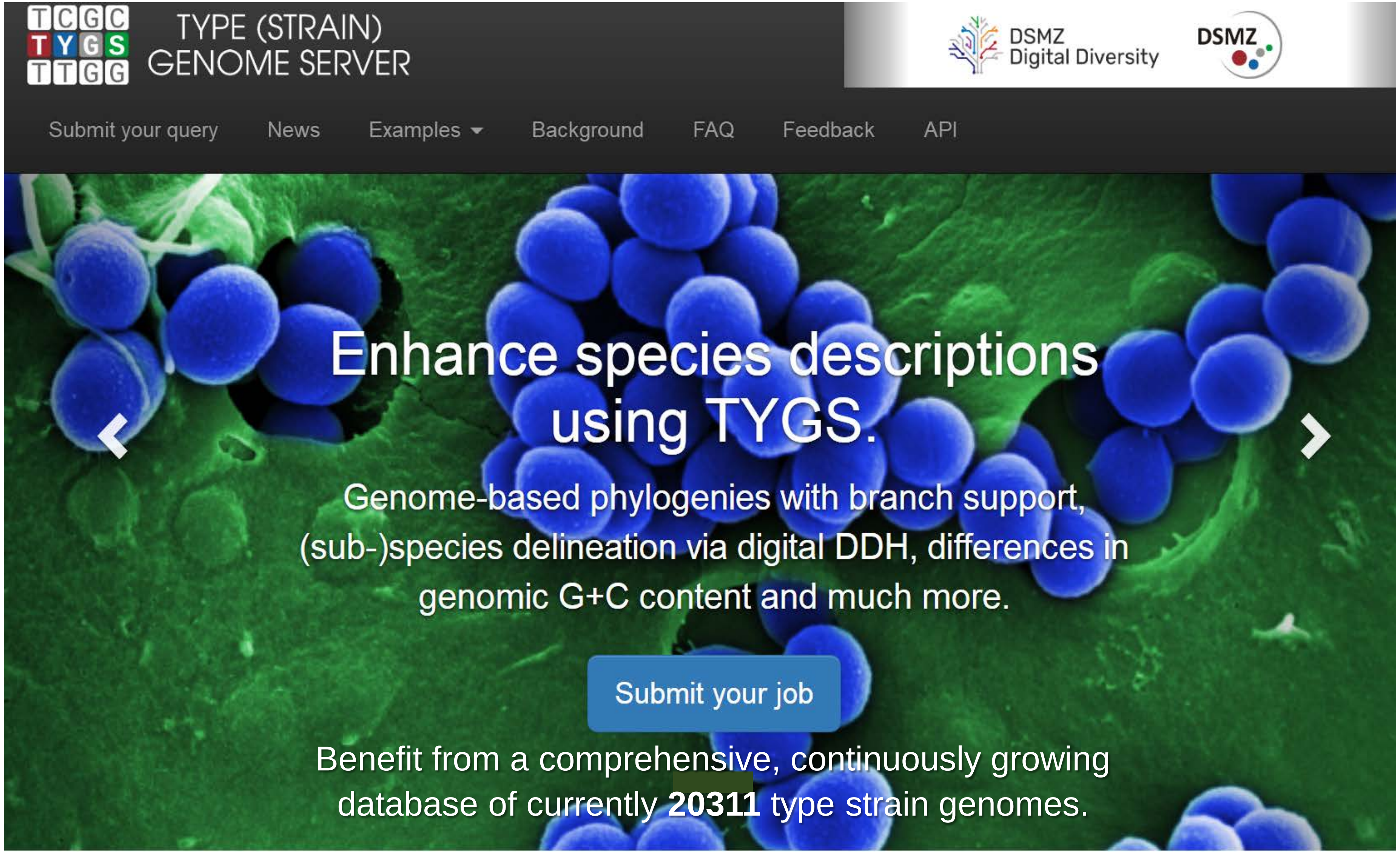
Release 08-RS214 (28th April 2023)

Nat Biotechnol, 2018

基于全基因组序列的原核生物系统分类平台--TYGS

- TYGS is the successor of the popular **Genome-to-Genome Distance Calculator**
- is a user-friendly high-throughput web server for **genome-based prokaryote taxonomy**
- is connected to a large, **continuously growing database** of genomic, taxonomic and nomenclatural information
- is highly interconnected with the **LPSN database** and depends on its expert-curated nomenclatural data
- provides a **multitude of tailored results.**

https://tygs.dsmz.de/



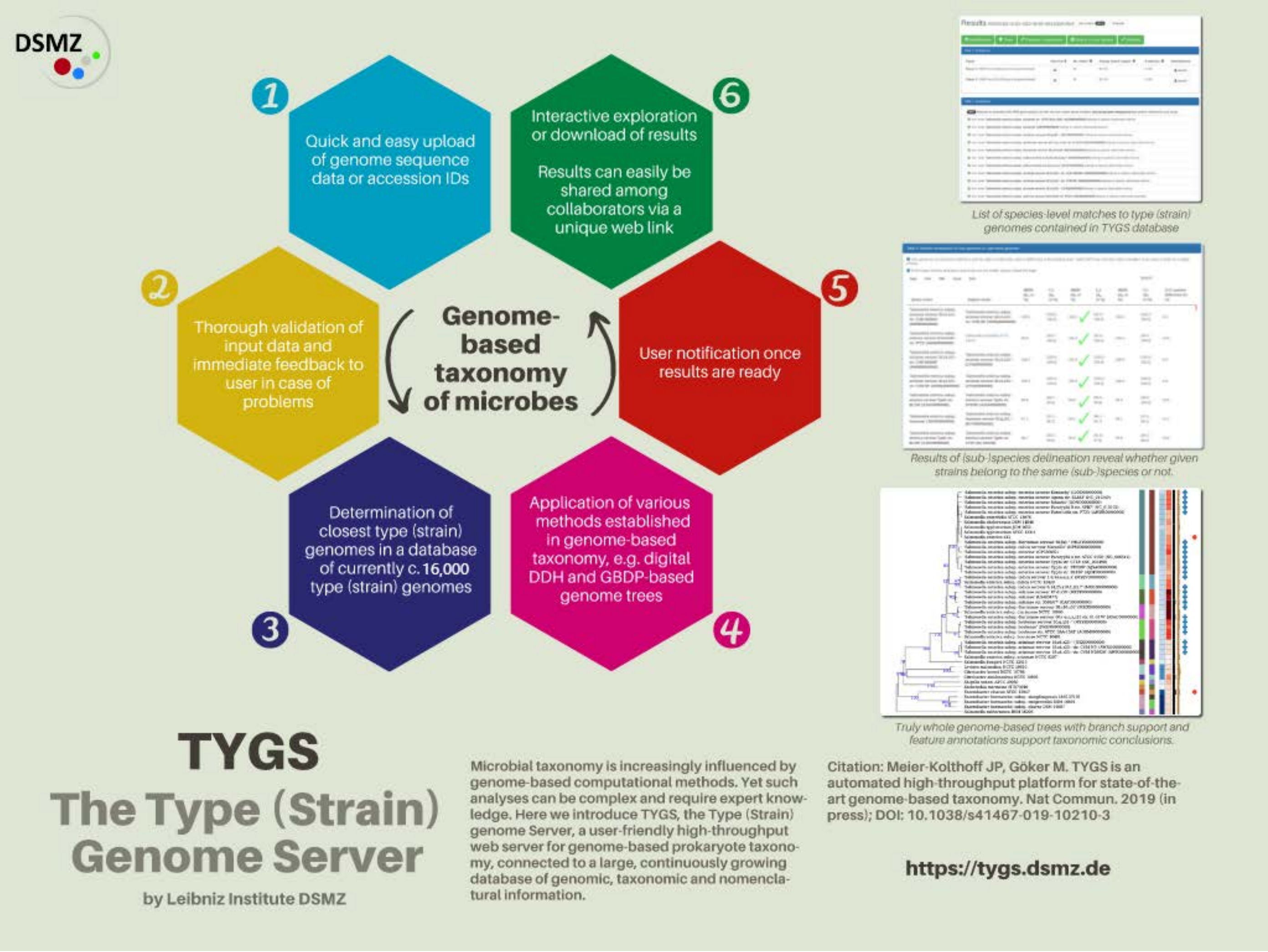
Enhance species descriptions using TYGS.

Genome-based phylogenies with branch support, (sub-)species delineation via digital DDH, differences in genomic G+C content and much more.

Benefit from a comprehensive, continuously growing database of currently **20311** type strain genomes.

Nat Commun, 2019; Nucleic Acids Res, 2022

基于全基因组序列的原核生物系统分类平台--TYGS



DSMZ

1 Quick and easy upload of genome sequence data or accession IDs

2 Thorough validation of input data and immediate feedback to user in case of problems

3 Determination of closest type (strain) genomes in a database of currently c. 16,000 type (strain) genomes

4 Application of various methods established in genome-based taxonomy, e.g. digital DDH and GBDP-based genome trees

5 User notification once results are ready

6 Interactive exploration or download of results

Results can easily be shared among collaborators via a unique web link

Genome-based taxonomy of microbes

List of species-level matches to type (strain) genomes contained in TYGS database

Results of (sub-)species delineation reveal whether given strains belong to the same (sub-)species or not.

Truly whole genome-based trees with branch support and feature annotations support taxonomic conclusions.

TYGS

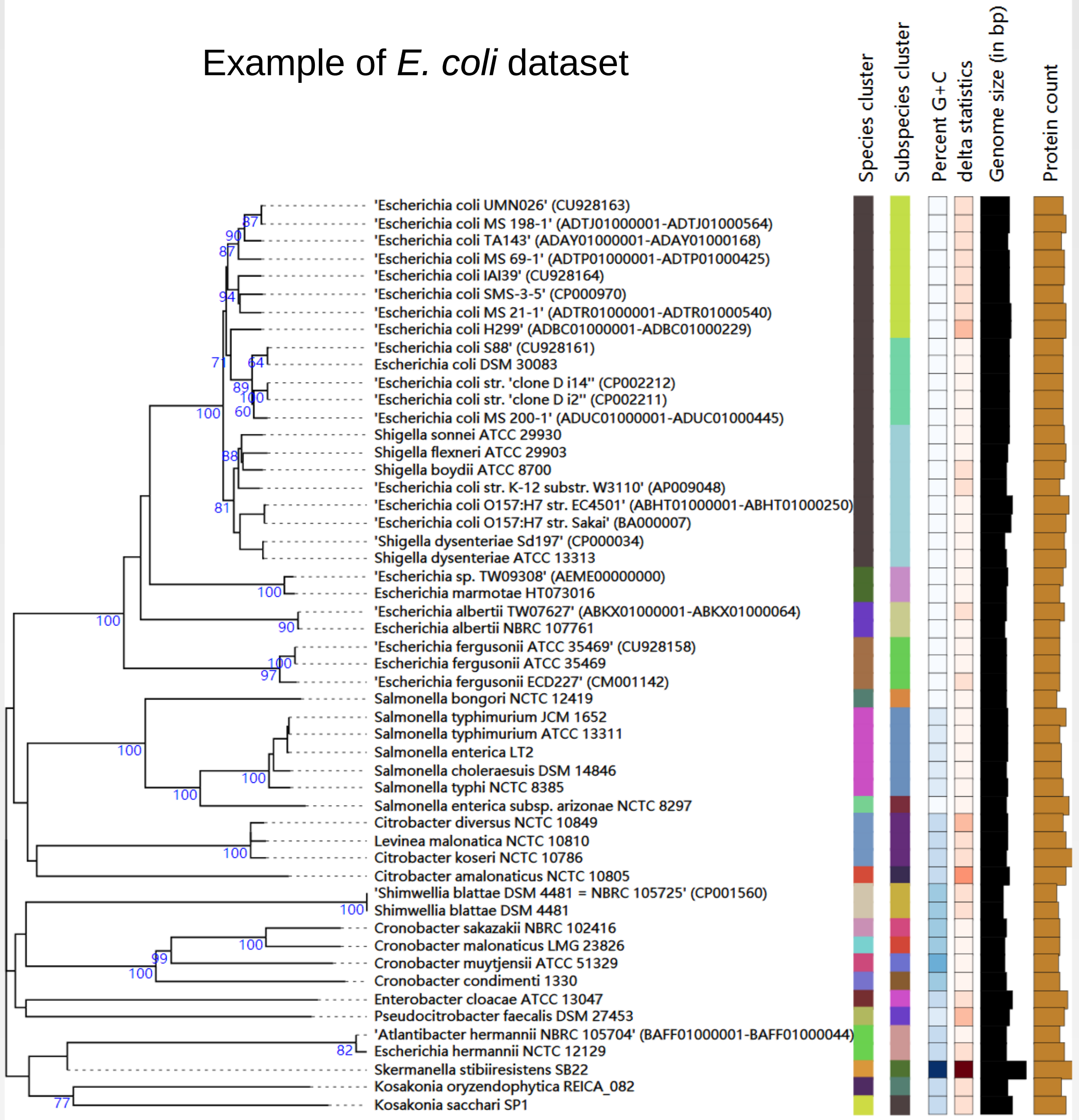
The Type (Strain) Genome Server

by Leibniz Institute DSMZ

Microbial taxonomy is increasingly influenced by genome-based computational methods. Yet such analyses can be complex and require expert knowledge. Here we introduce TYGS, the Type (Strain) Genome Server, a user-friendly high-throughput web server for genome-based prokaryote taxonomy, connected to a large, continuously growing database of genomic, taxonomic and nomenclatural information.

Citation: Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun. 2019 (in press); DOI: 10.1038/s41467-019-10210-3

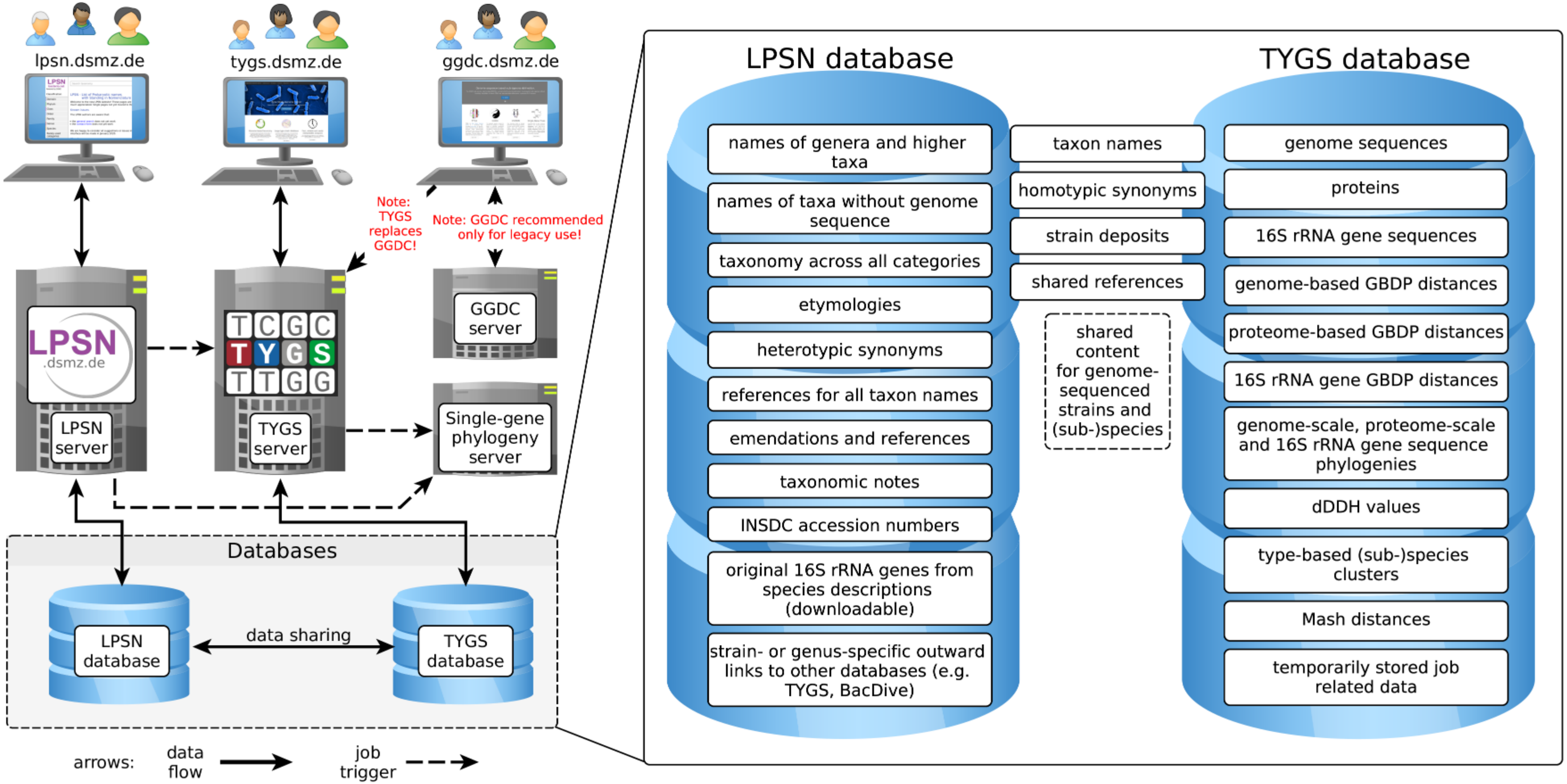
https://tygs.dsmz.de



Example of *E. coli* dataset

基于全基因组序列的原核生物系统分类平台--TYGS

Dependencies between the TYGS and LPSN databases



lpsn.dsmz.de

tygs.dsmz.de

ggdc.dsmz.de

Note: TYGS replaces GGDC!

Note: GGDC recommended only for legacy use!

LPSN .dsmz.de

LPSN server

T C G C
T Y G S
T T G G

TYGS server

GGDC server

Single-gene phylogeny server

Databases

LPSN database

data sharing

TYGS database

arrows: data flow → job trigger - - →

LPSN database

- names of genera and higher taxa
- names of taxa without genome sequence
- taxonomy across all categories
- etymologies
- heterotypic synonyms
- references for all taxon names
- emendations and references
- taxonomic notes
- INSDC accession numbers
- original 16S rRNA genes from species descriptions (downloadable)
- strain- or genus-specific outward links to other databases (e.g. TYGS, BacDive)

Shared

- taxon names
- homotypic synonyms
- strain deposits
- shared references

shared content for genome-sequenced strains and (sub-)species

TYGS database

- genome sequences
- proteins
- 16S rRNA gene sequences
- genome-based GBDP distances
- proteome-based GBDP distances
- 16S rRNA gene GBDP distances
- genome-scale, proteome-scale and 16S rRNA gene sequence phylogenies
- dDDH values
- type-based (sub-)species clusters
- Mash distances
- temporarily stored job related data

具有命名法地位的原核生物名录(LPSN)

- The **List of Prokaryotic names with Standing in Nomenclature (LPSN)** provides comprehensive information on the nomenclature of prokaryotes and much more.
- LPSN is a free to use service founded by Jean P. Euzéby in 1997 and later on maintained by Aidan C. Parte.

https://lpsn.dsmz.de/

Species *Bacillus amylolyticus*

ⓘ **Name:** *Bacillus amylolyticus* (*ex* Choukévitch 1911) Nakamura 1984

ⓘ **Category:** Species

ⓘ **Proposed as:** sp. nov., nom. rev.

ⓘ **Etymology:** a.my.lo.ly´ti.cus. Gr. neut. n. *amylon*, starch; Gr. masc. adj. *lytikos*, dissolving; N.L. masc. adj. *amylolyticus*, dissolving starch

ⓘ **Gender:** masculine

ⓘ **Type strain:** ATCC 9995; DSM 11730; DSM 11747; DSM 15211; DSM 3034; NRRL B-377; NRRL NRS-290

ⓘ See detailed strain information at **BacDive**

ⓘ Conduct genome-based taxonomy at TYGS

ⓘ **16S rRNA gene:** D85396 Analyse FASTA ENA NCBI

ⓘ **Valid publication:** Nakamura LK. *Bacillus amylolyticus* sp. nov., nom. rev., *Bacillus lautus* sp. nov., nom. rev., *Bacillus pabuli* sp. nov., nom. rev., and *Bacillus validus* sp. nov., nom. rev. *Int. J. Syst. Bacteriol.* 1984; **34**:224-226. doi

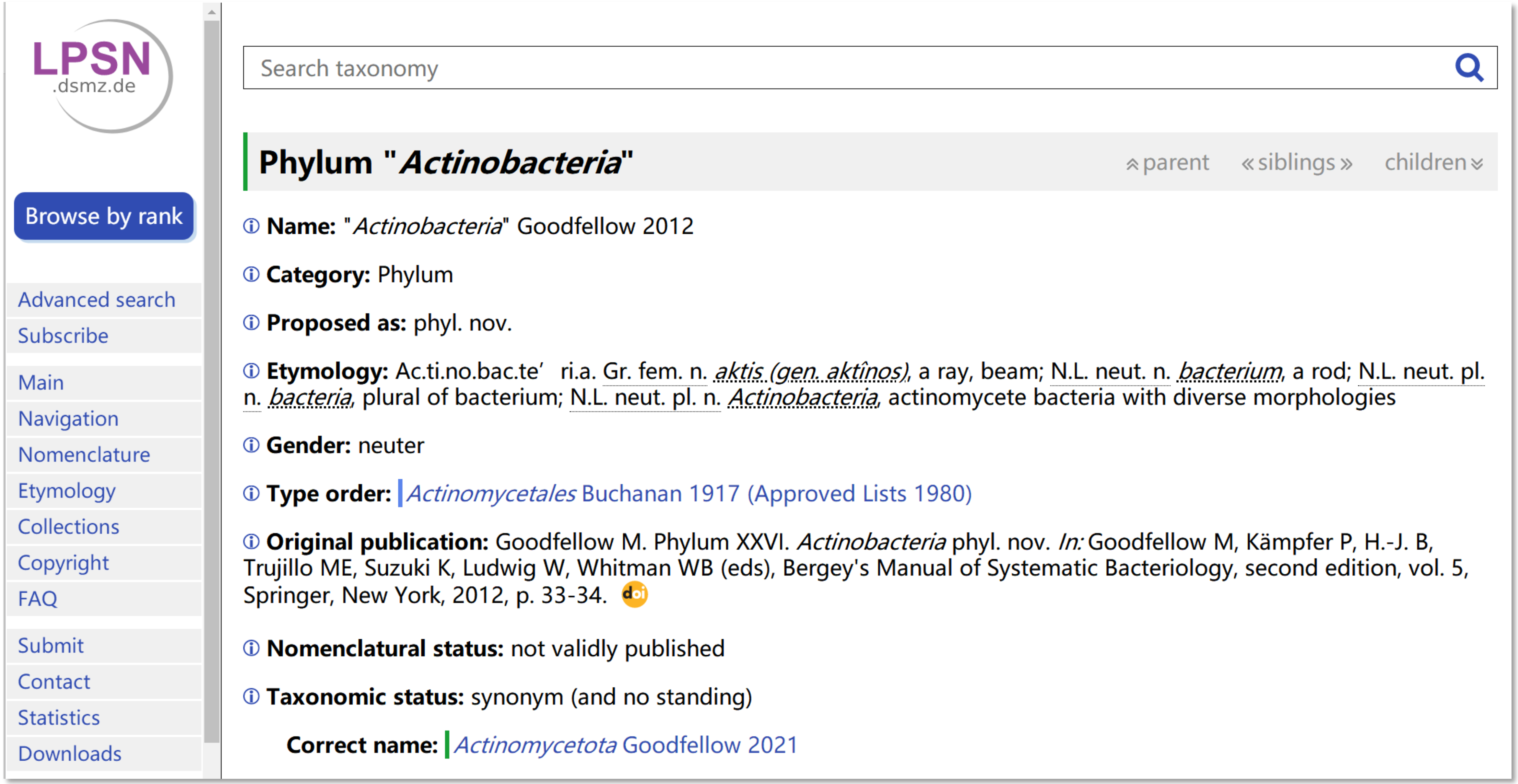
ⓘ **Nomenclatural status:** validly published under the ICNP

ⓘ **Taxonomic status:** synonym

Correct name: *Paenibacillus amylolyticus* (Nakamura 1984 *ex* Choukévitch 1911) Ash *et al.* 1994

ⓘ **Risk group:** 1

具有命名法地位的原核生物名录(LPSN)



Phylum "*Actinobacteria*"

ⓘ **Name:** "*Actinobacteria*" Goodfellow 2012

ⓘ **Category:** Phylum

ⓘ **Proposed as:** phyl. nov.

ⓘ **Etymology:** Ac.ti.no.bac.te’ ri.a. Gr. fem. n. *aktis (gen. aktînos)*, a ray, beam; N.L. neut. n. *bacterium*, a rod; N.L. neut. pl. n. *bacteria*, plural of bacterium; N.L. neut. pl. n. *Actinobacteria*, actinomycete bacteria with diverse morphologies

ⓘ **Gender:** neuter

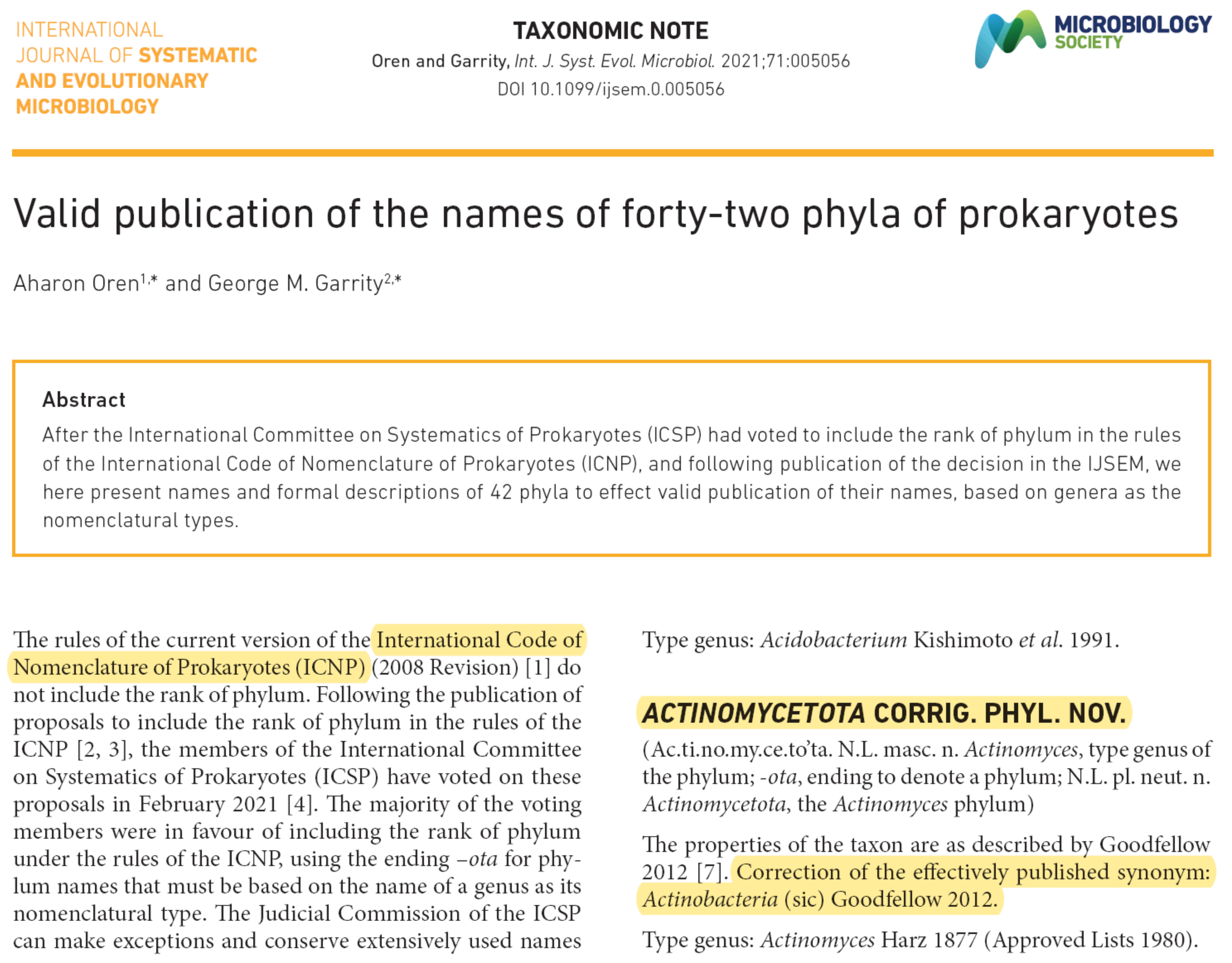
ⓘ **Type order:** *Actinomycetales* Buchanan 1917 (Approved Lists 1980)

ⓘ **Original publication:** Goodfellow M. Phylum XXVI. *Actinobacteria* phyl. nov. *In:* Goodfellow M, Kämpfer P, H.-J. B, Trujillo ME, Suzuki K, Ludwig W, Whitman WB (eds), Bergey's Manual of Systematic Bacteriology, second edition, vol. 5, Springer, New York, 2012, p. 33-34. doi

ⓘ **Nomenclatural status:** not validly published

ⓘ **Taxonomic status:** synonym (and no standing)

Correct name: *Actinomycetota* Goodfellow 2021



INTERNATIONAL
JOURNAL OF **SYSTEMATIC
AND EVOLUTIONARY
MICROBIOLOGY**

TAXONOMIC NOTE
Oren and Garrity, *Int. J. Syst. Evol. Microbiol.* 2021;71:005056
DOI 10.1099/ijsem.0.005056

MICROBIOLOGY SOCIETY

Valid publication of the names of forty-two phyla of prokaryotes

Aharon Oren[1,]* and George M. Garrity[2,]*

Abstract

After the International Committee on Systematics of Prokaryotes (ICSP) had voted to include the rank of phylum in the rules of the International Code of Nomenclature of Prokaryotes (ICNP), and following publication of the decision in the IJSEM, we here present names and formal descriptions of 42 phyla to effect valid publication of their names, based on genera as the nomenclatural types.

The rules of the current version of the International Code of Nomenclature of Prokaryotes (ICNP) (2008 Revision) [1] do not include the rank of phylum. Following the publication of proposals to include the rank of phylum in the rules of the ICNP [2, 3], the members of the International Committee on Systematics of Prokaryotes (ICSP) have voted on these proposals in February 2021 [4]. The majority of the voting members were in favour of including the rank of phylum under the rules of the ICNP, using the ending –*ota* for phylum names that must be based on the name of a genus as its nomenclatural type. The Judicial Commission of the ICSP can make exceptions and conserve extensively used names

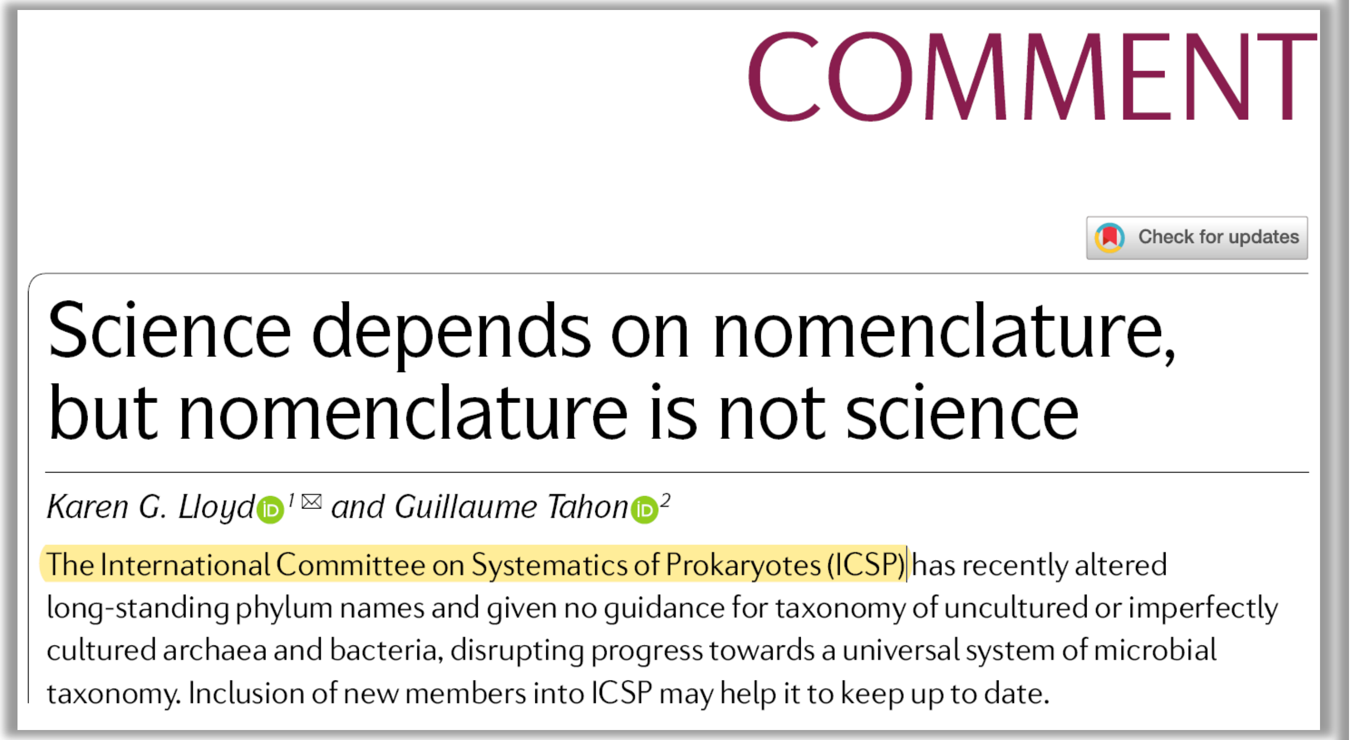
Type genus: *Acidobacterium* Kishimoto *et al*. 1991.

ACTINOMYCETOTA CORRIG. PHYL. NOV.

(Ac.ti.no.my.ce.to'ta. N.L. masc. n. *Actinomyces*, type genus of the phylum; -*ota*, ending to denote a phylum; N.L. pl. neut. n. *Actinomycetota*, the *Actinomyces* phylum)

The properties of the taxon are as described by Goodfellow 2012 [7]. Correction of the effectively published synonym: *Actinobacteria* (sic) Goodfellow 2012.

Type genus: *Actinomyces* Harz 1877 (Approved Lists 1980).



COMMENT

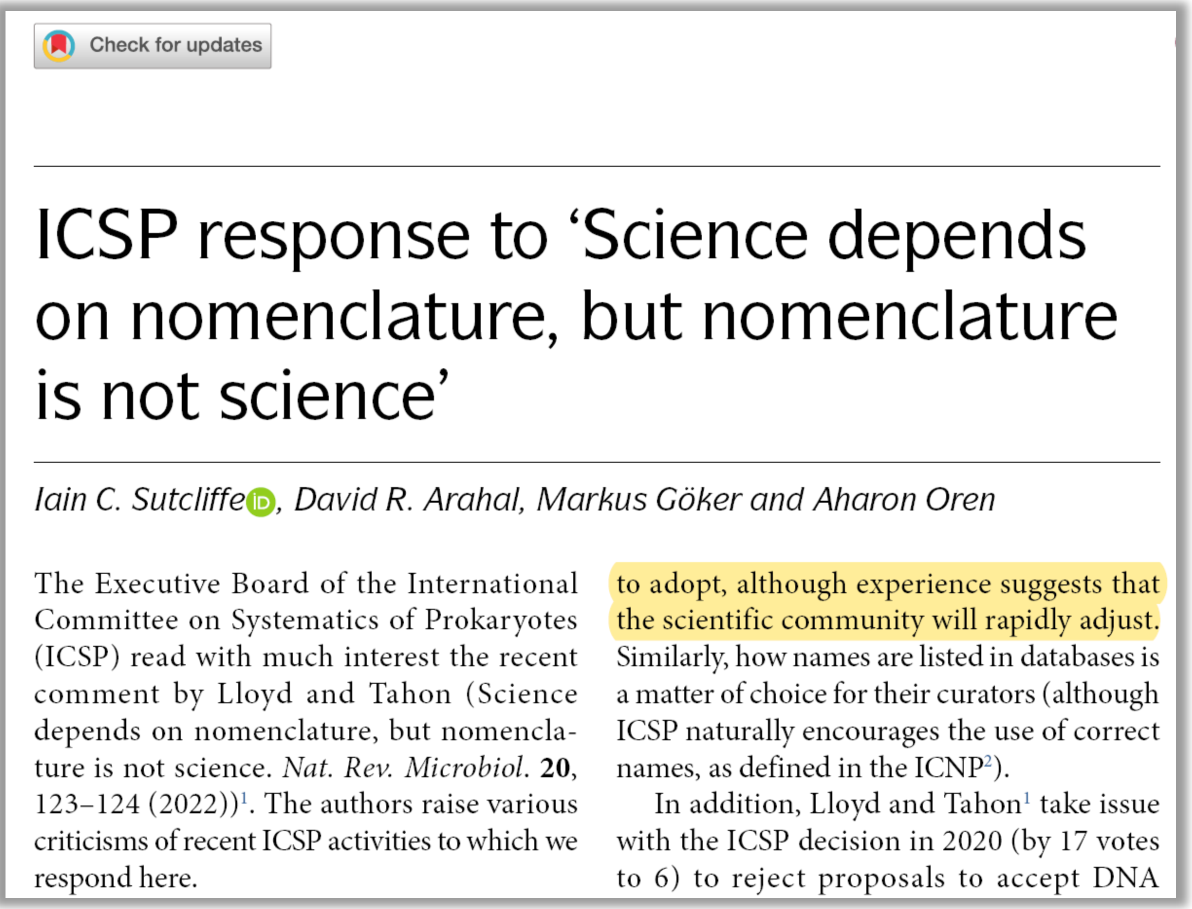
Check for updates

Science depends on nomenclature, but nomenclature is not science

Karen G. Lloyd[1] ✉ *and Guillaume Tahon*[2]

The International Committee on Systematics of Prokaryotes (ICSP) has recently altered long-standing phylum names and given no guidance for taxonomy of uncultured or imperfectly cultured archaea and bacteria, disrupting progress towards a universal system of microbial taxonomy. Inclusion of new members into ICSP may help it to keep up to date.

Nat Rev Microbiol, 2022, **20**,123–124



Check for updates

ICSP response to 'Science depends on nomenclature, but nomenclature is not science'

Iain C. Sutcliffe, David R. Arahal, Markus Göker and Aharon Oren

The Executive Board of the International Committee on Systematics of Prokaryotes (ICSP) read with much interest the recent comment by Lloyd and Tahon (Science depends on nomenclature, but nomenclature is not science. *Nat. Rev. Microbiol.* **20**, 123–124 (2022))[1]. The authors raise various criticisms of recent ICSP activities to which we respond here.

to adopt, although experience suggests that the scientific community will rapidly adjust. Similarly, how names are listed in databases is a matter of choice for their curators (although ICSP naturally encourages the use of correct names, as defined in the ICNP[2]).

In addition, Lloyd and Tahon[1] take issue with the ICSP decision in 2020 (by 17 votes to 6) to reject proposals to accept DNA

Nat Rev Microbiol, 2022, **20**,249–250

原核生物分类命名百年历程

物种定义多基于核酸相似性及给定的阈值，特定阈值并不适用于所有原核生物；

物种的定义可以是实际可操作性质的，但**物种的概念需要基于物种形成的机制。**

Hugenholtz, et al.
ISME J, 2021

物种是什么?

◆ 一个所有生物学家都在用，但是至今尚无统一定论的概念

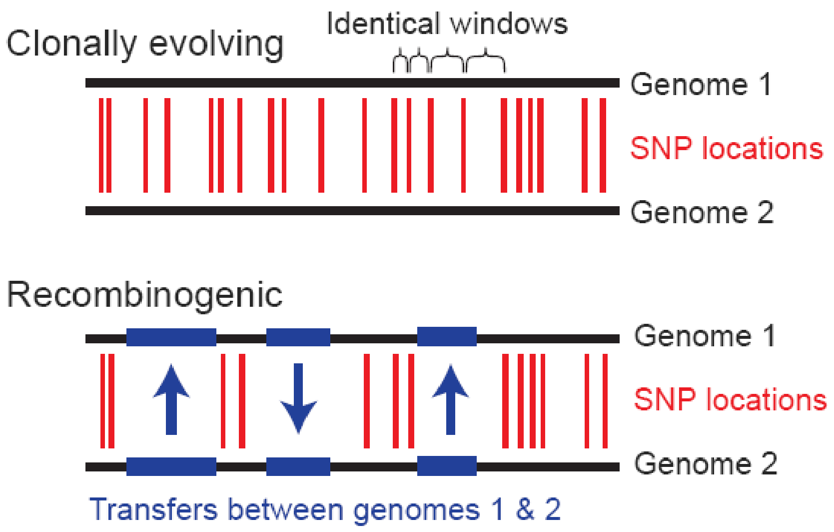
◆ **Biological Species Concept (BSC):** a species concept based on restricted gene flow.

- 真核生物：可交配群体；生殖隔离
- 原核生物：群体内部比群体间具有更多的基因交流

◆ **物种：独立进化的集合种群。**

Species are recognizable as distinct clusters, based on genetic similarity across the genome and differences from other species.

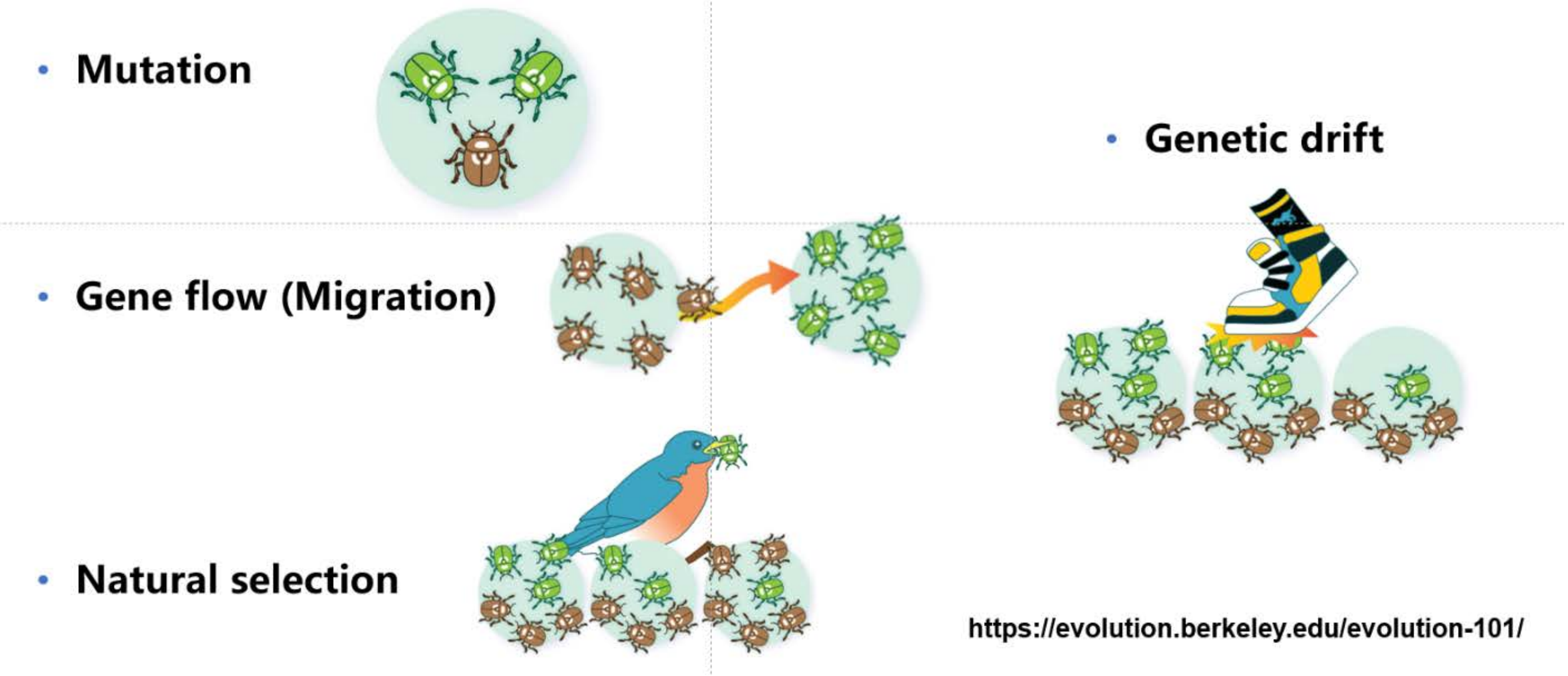
de Queiroz. Syst Biol, 2007
Hendry. Nature, 2009
Shapiro and Polz. Trends Microbiol, 2014
Shapiro et al. Plos Genet, 2016
Arevalo et al. Cell, 2019



微生物微进化

◆ **微进化（microevolution）**：物种内遗传组成的改变和多样性形成的过程

◆ **机制：** 突变、基因交流、遗传漂变、自然选择

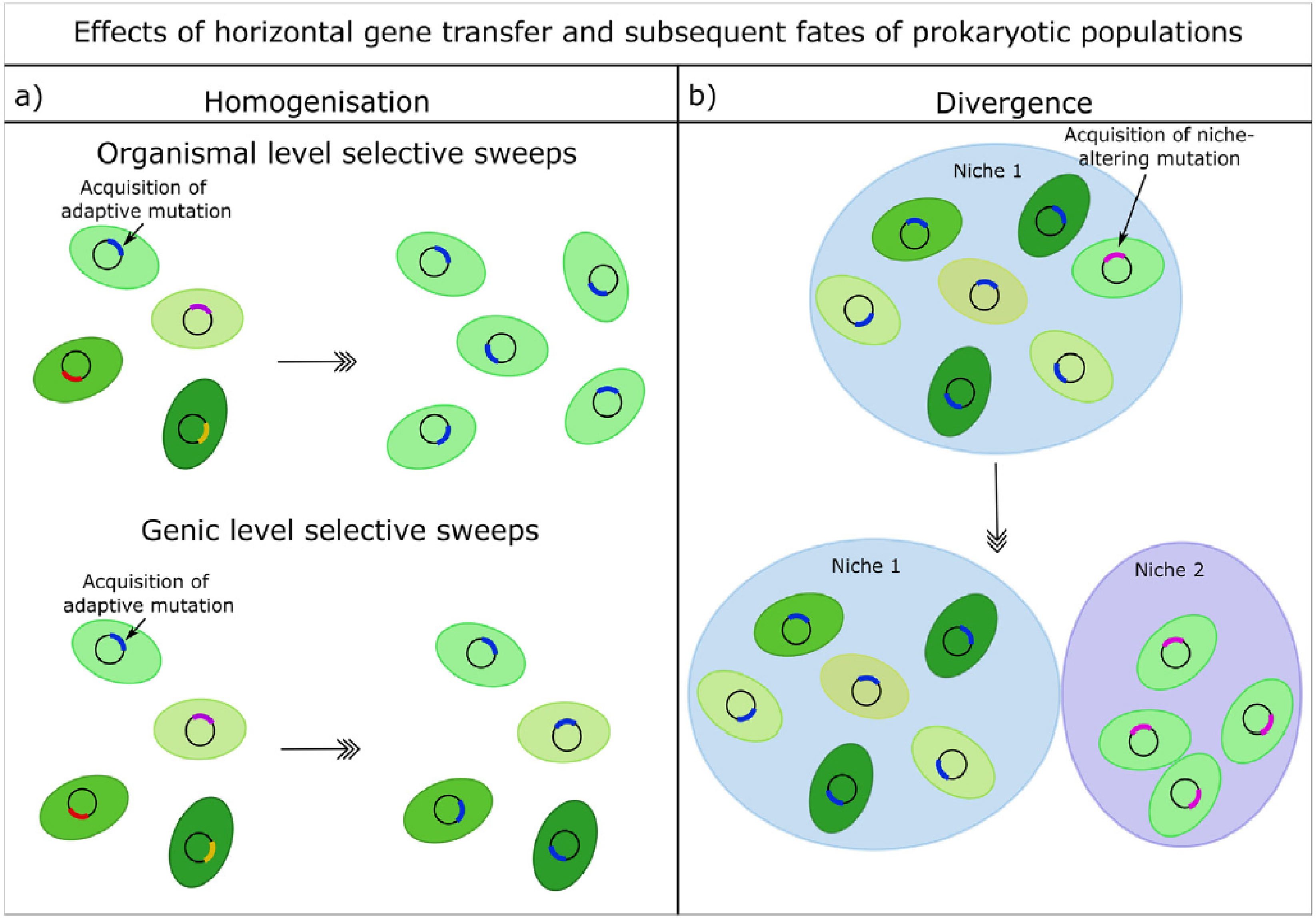


https://evolution.berkeley.edu/evolution-101/

◆ **突变和基因交流引入种内变异，遗传漂变和自然选择决定变异的命运**

基因交流的结果

- 原核生物的水平基因转移(HGT)是单向的，通过可移动遗传元件(MGE)等介质实现
- **种内HGT可能代表着种群内部强大的均质力**，这种HGT介导的过程涉及在生物水平和基因水平上的选择性清除，以保持个体间的凝聚力
- 与种内HGT相比，**不同物种成员之间的HGT可以帮助分化**



Palmer et al. 2019

什么决定了物种的多样性?

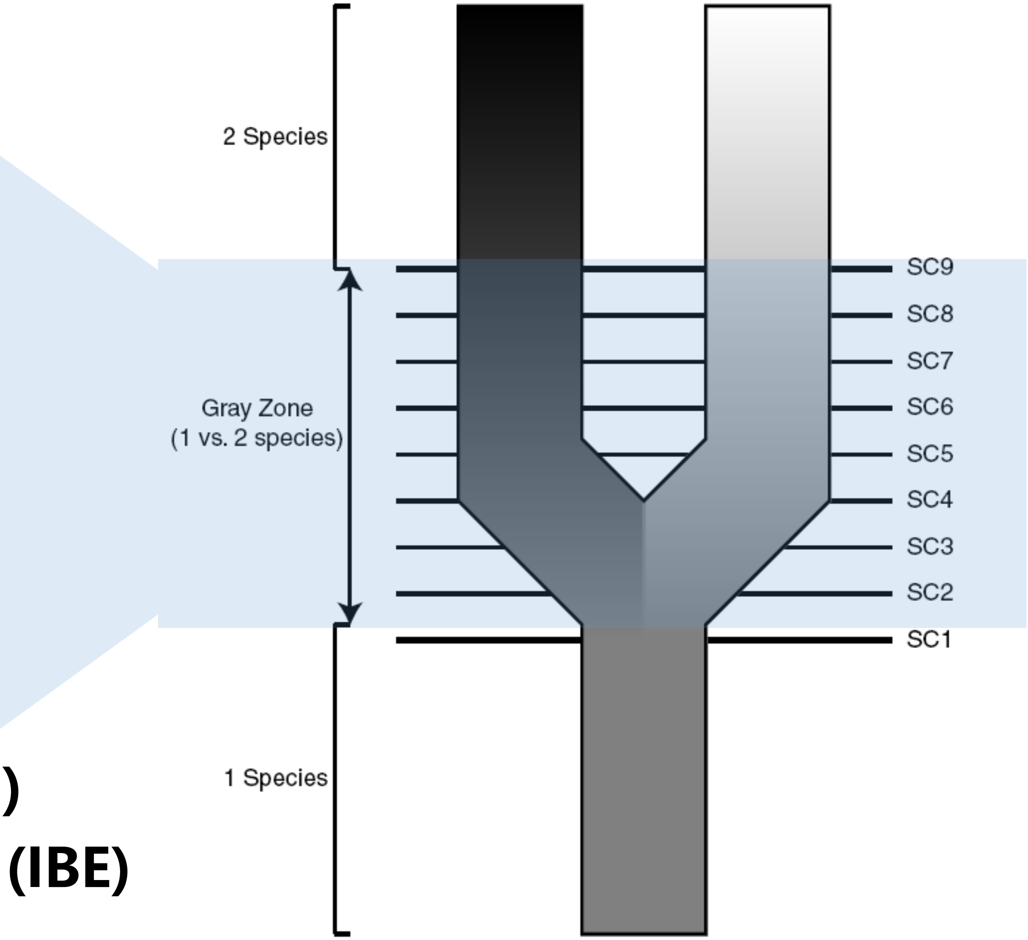
◆ 影响物种形成的主要因素：

- **重组/突变—— 主要内因**
- **选 择 —— 主要外因**
- **扩散/迁移 —— 重要因素，内外因结合**
- **遗传漂变 —— 重要因素，内外因结合**

◆ 外因：

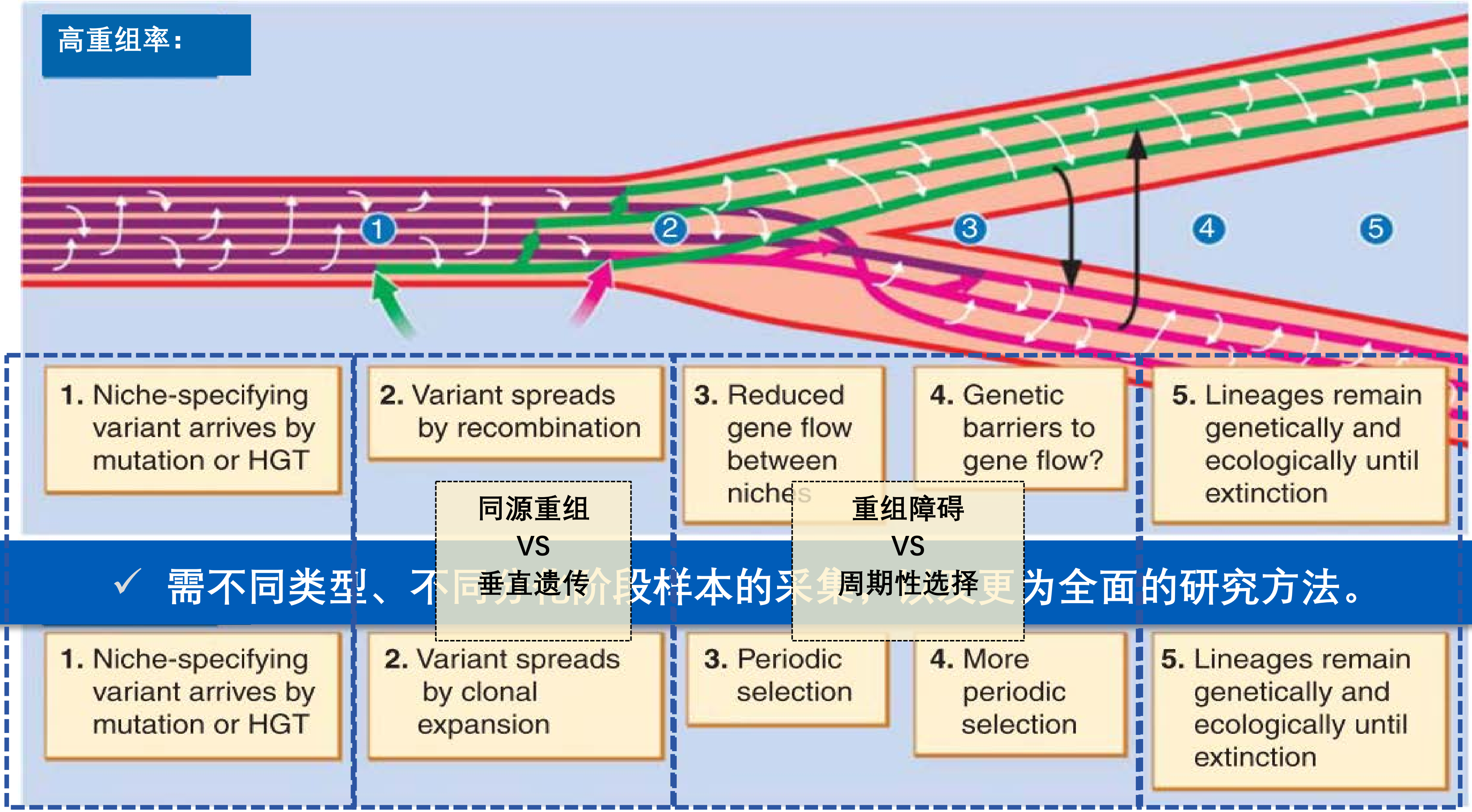
- **地理因素：isolation by distance (IBD)**
- **生态因素：isolation by environment (IBE)**

Hanson et al. Nat Rev Microbiol, 2012；Sexton et al. Evolution, 2013



de Queiroz. Syst Biol, 2007

原核生物物种形成的阶段



高重组率：

1. Niche-specifying variant arrives by mutation or HGT
2. Variant spreads by recombination
3. Reduced gene flow between niches
4. Genetic barriers to gene flow?
5. Lineages remain genetically and ecologically until extinction

同源重组
VS
垂直遗传

重组障碍
VS
周期性选择

✓ 需不同类型、不同分化阶段样本的采集，以及更为全面的研究方法。

1. Niche-specifying variant arrives by mutation or HGT
2. Variant spreads by clonal expansion
3. Periodic selection
4. More periodic selection
5. Lineages remain genetically and ecologically until extinction

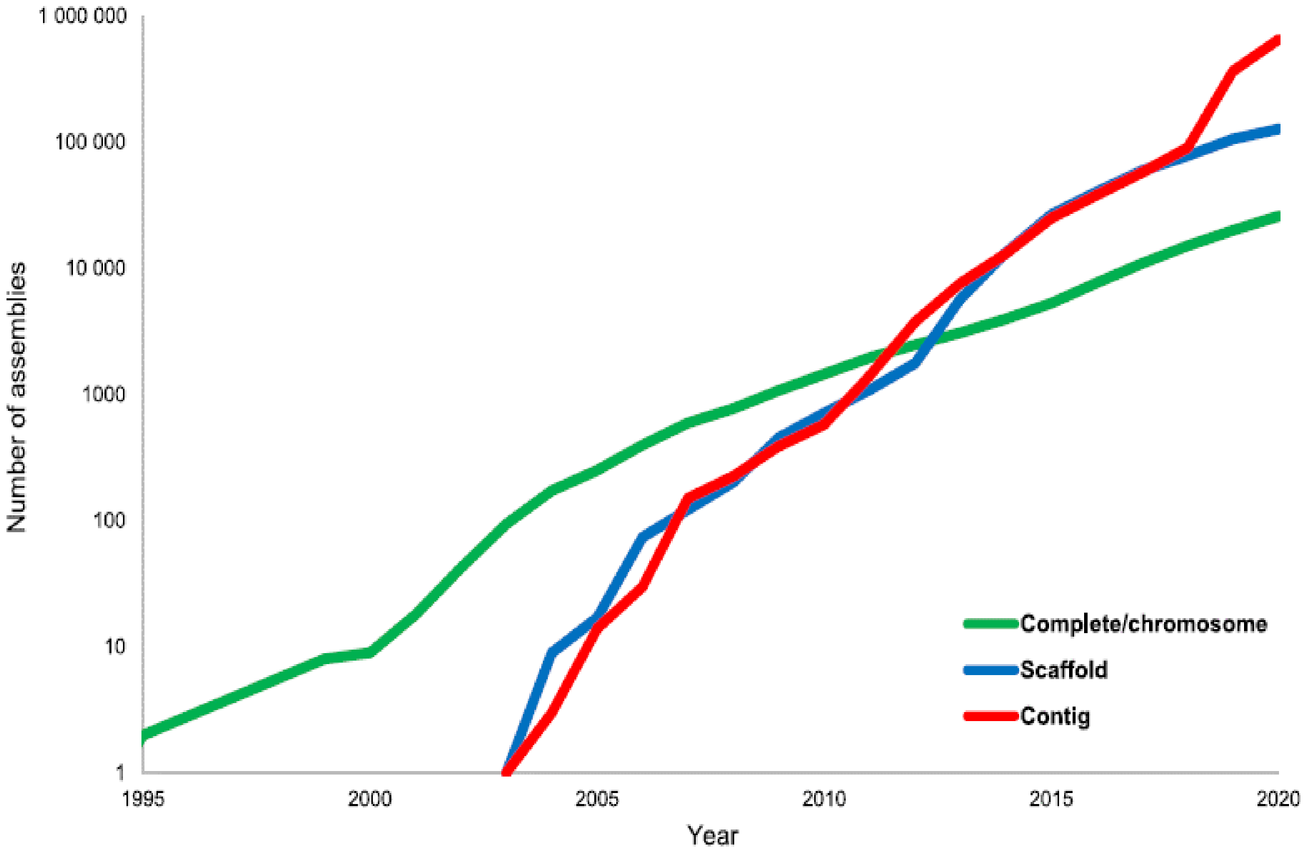
不同重组/选择率条件下的细菌物种形成模型

Shapiro and Polz, 2014, 2015

群体遗传学与群体基因组学

- **群体遗传学** 是遗传学的分支学科，通过对群体内等位基因分布和频率在空间和时间上的表征来推断塑造种群模式（如聚类、分化、分歧等）的进化过程。

- **群体基因组学**作为基因组学与群体遗传学的结合体，通过对全基因组范围的多态性位点进行研究，来更好地理解进化过程是如何塑造基因组和种群的变异。已经成为微生物领域界定物种边界和研究物种形成机制的有力工具。



Koonin, et al., 2021

研究群体基因组学的工具

◆ **PopCOGenT**

Populations as **c**lusters **o**f **gen**e **t**ransfer

https://github.com/philarevalo/PopCOGenT

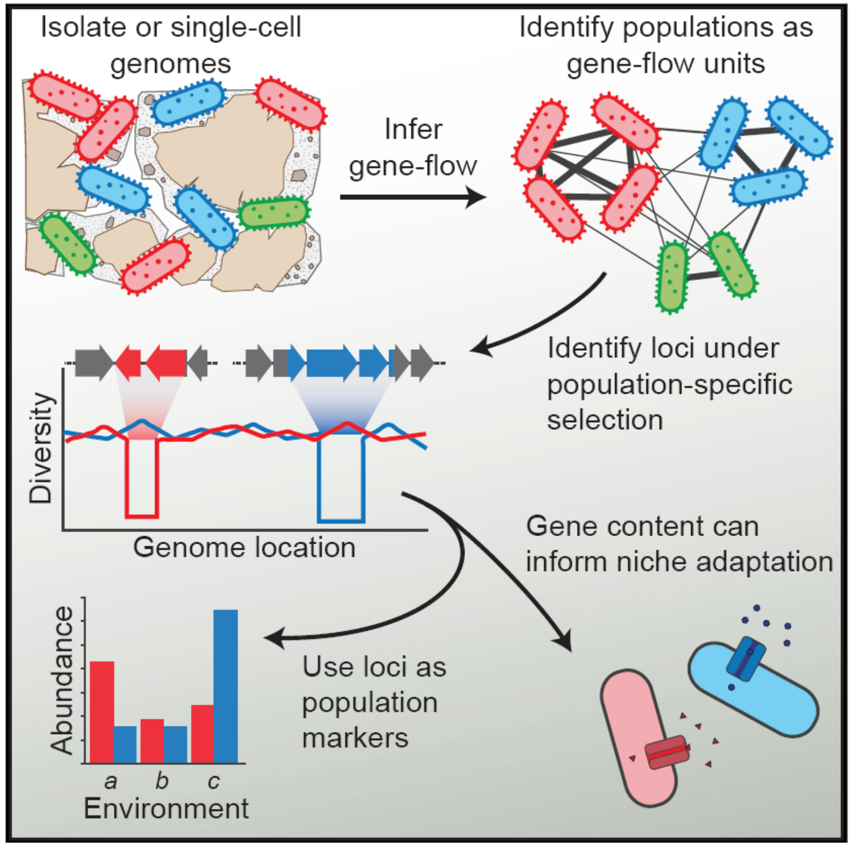
(¹ Microbiology Graduate Program, MIT)

Article

Cell 2019

A Reverse Ecology Approach Based on a Biological Definition of Microbial Populations

Graphical Abstract



Authors

Philip Arevalo, David VanInsberghe, Joseph Elsherbini, Jeff Gore, Martin F. Polz

Correspondence

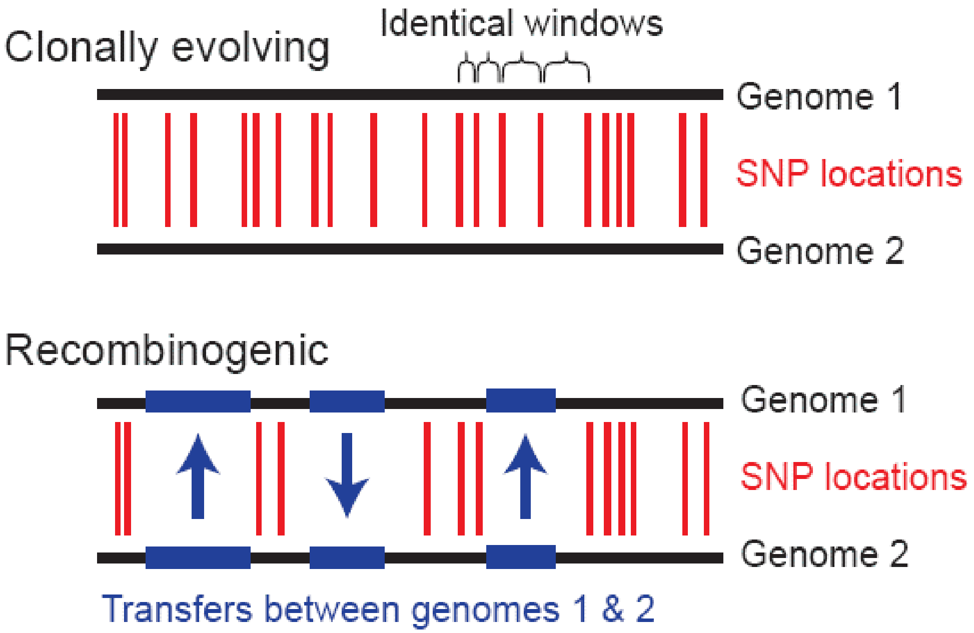
mpolz@mit.edu

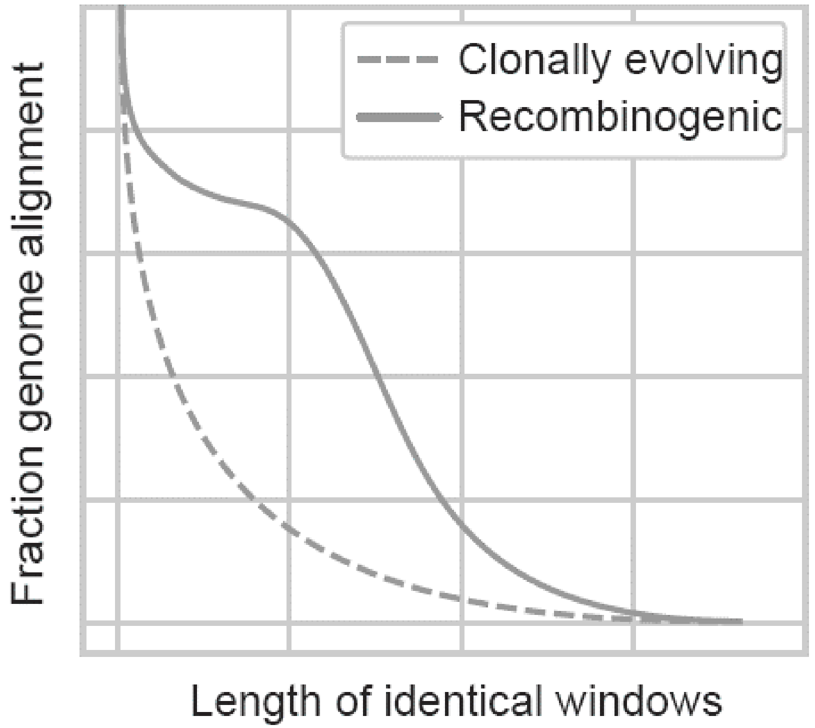
In Brief

Defining bacterial populations by examining recent horizontal gene transfer events illuminates ecologically meaningful interspecies relationships in the environment and in the human microbiome.

PopCOGenT

Amount of gene exchange should be reflected in fraction of identical genome windows

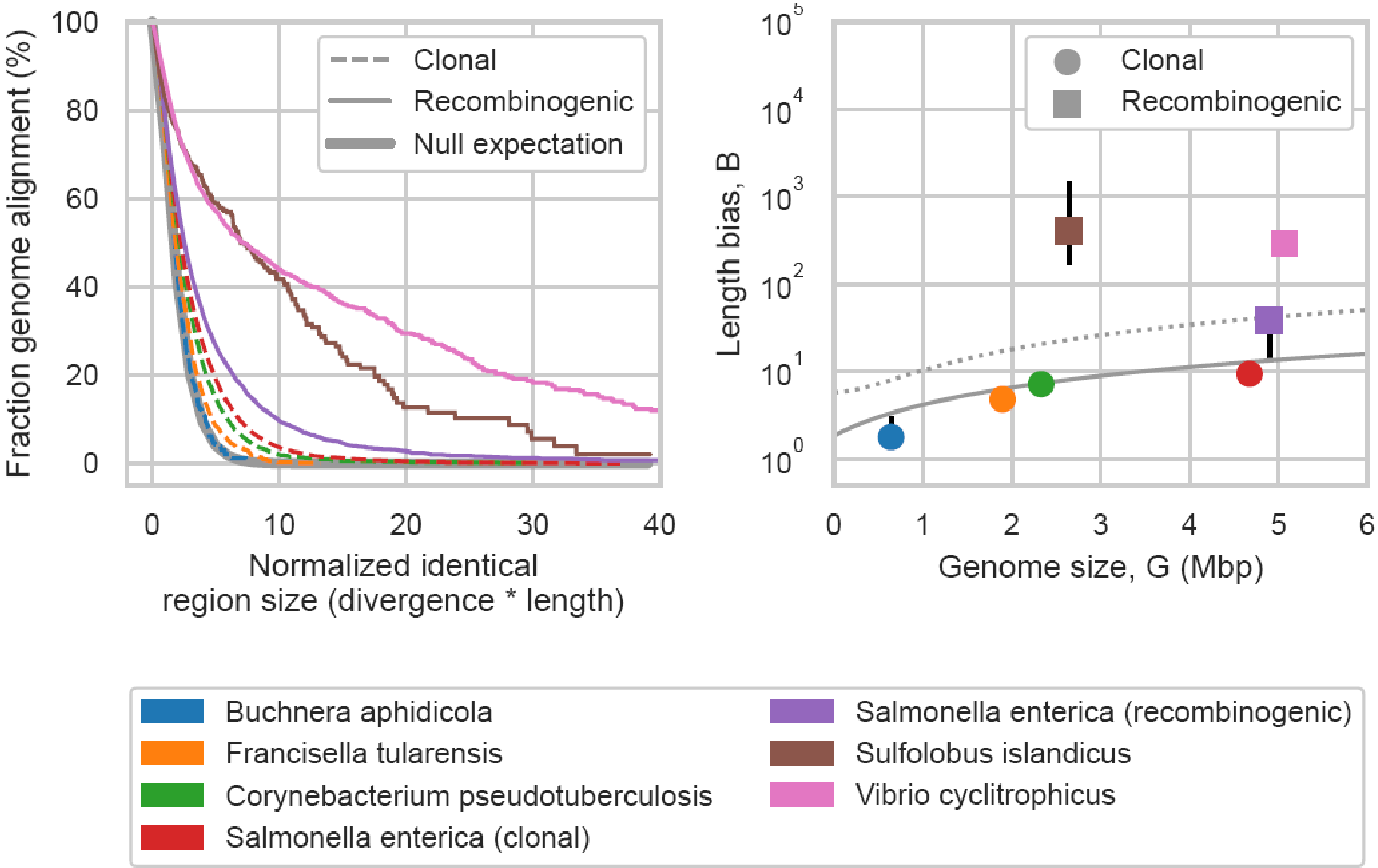




Null model of mutation distribution: identical regions between genomes should be exponentially distributed

PopCOGenT

Method accurately identifies gene flow



Clonal bacteria conform to the null model and sum squared difference between curves can estimate gene flow

Cell, 2019

链霉菌物种形成研究材料

微黄白链霉菌

(*Streptomyces albidoflavus*)

- 产生苦霉素、几丁质酶及多种抗真菌活性物质；
- 是目前报道唯一产生胞外多向自调节信号蛋白--C因子的链霉菌种群。
- 41**株菌，ANI >98.15%**

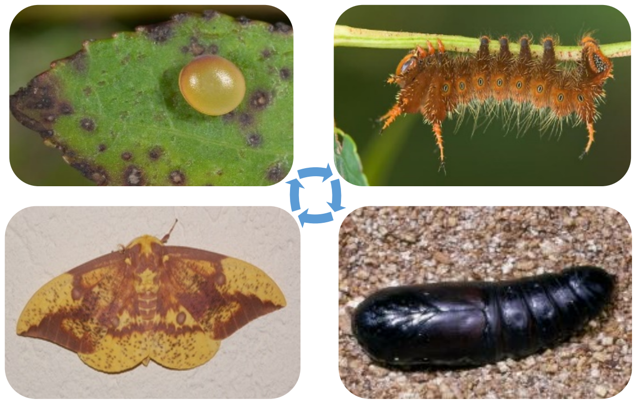
橄榄色链霉菌

(*Streptomyces olivaceus*)

- 产生Tetroazolemycis、Lobophorin、Bafilomycin A等；
- 降解纤维素、果胶等。
- 37**株菌，ANI >96.93%**

土壤

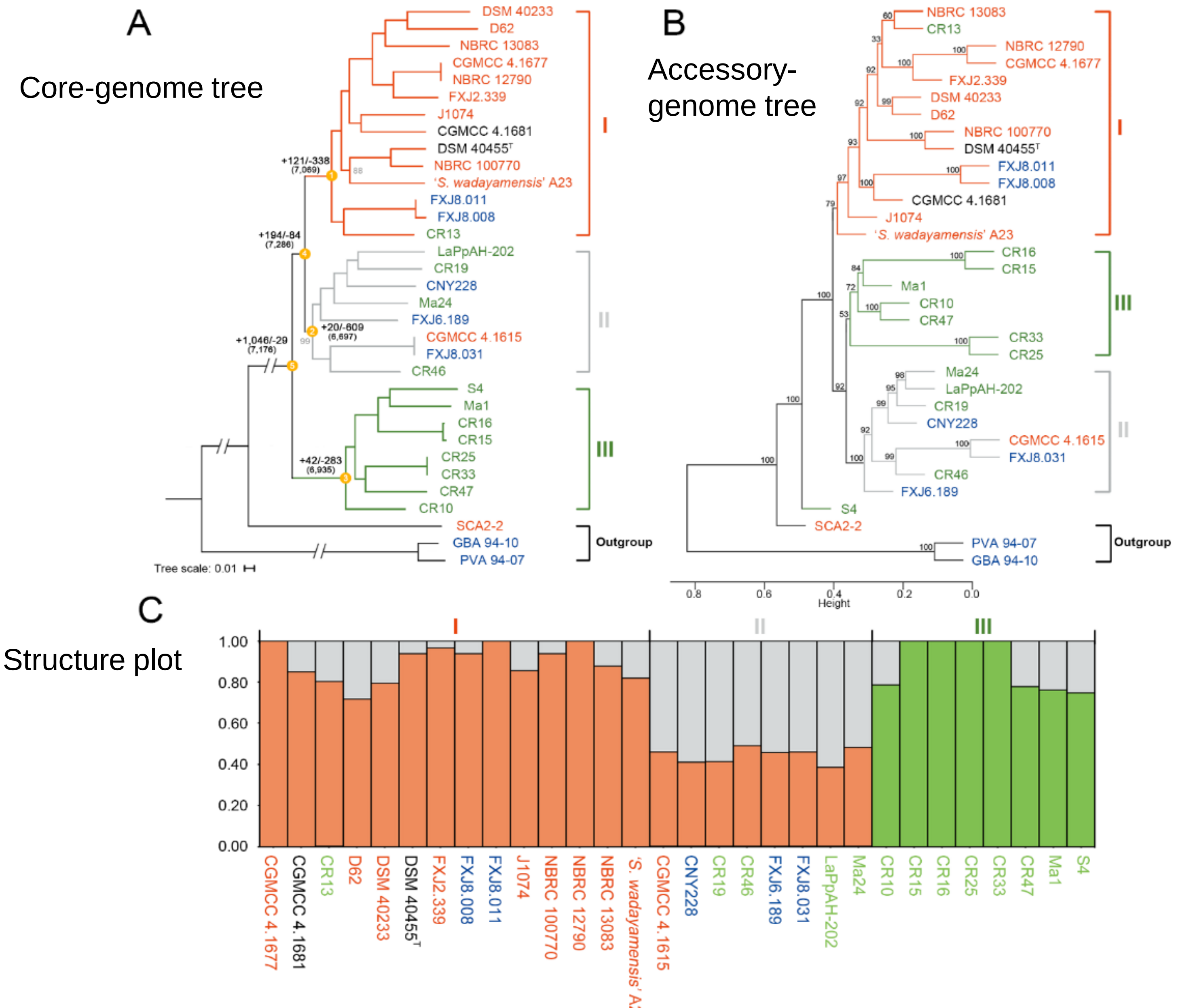
海洋



昆虫

菌株分离自：
土壤、海洋、昆虫
全球多个位点

微黄白链霉菌基因组系统发育和种群结构



◆ 生境相关的三个种群：

I、II、III

◆ 两类生境：

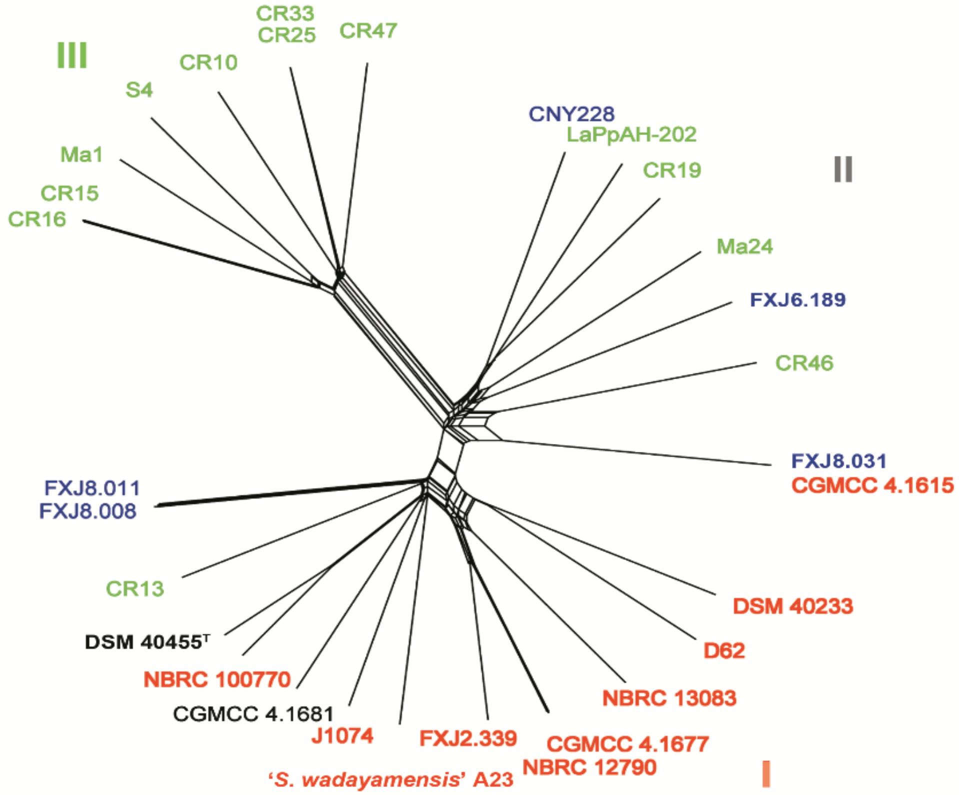
Free-living

Insect-associated

Li et al. Appl Environ Microbiol, 2019

微黄白链霉菌的同源重组较频繁

ρ/θ: mean, 3.36; median, 2.42



Phylogenetic network

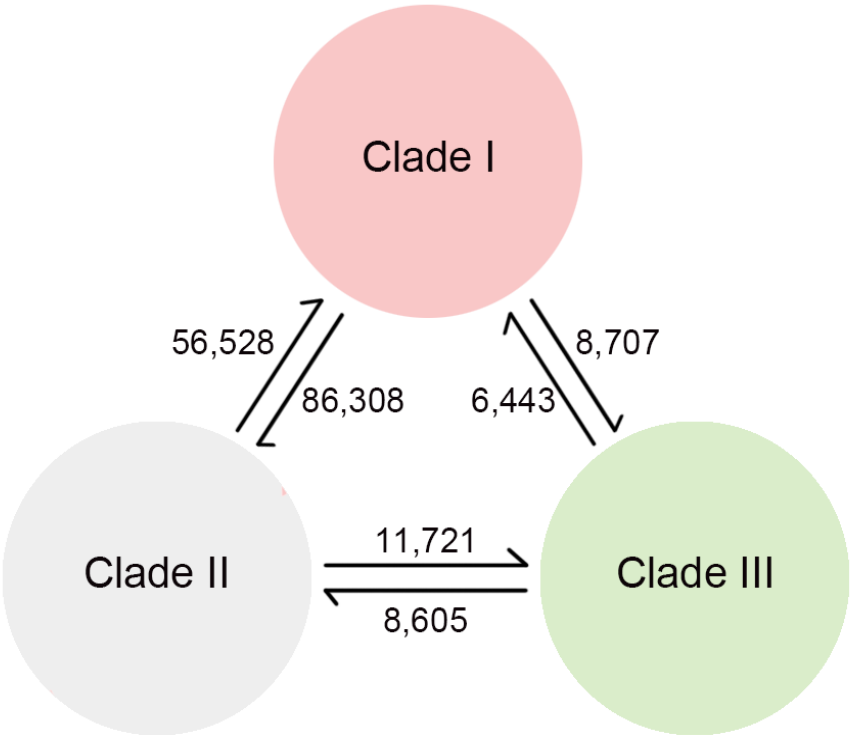
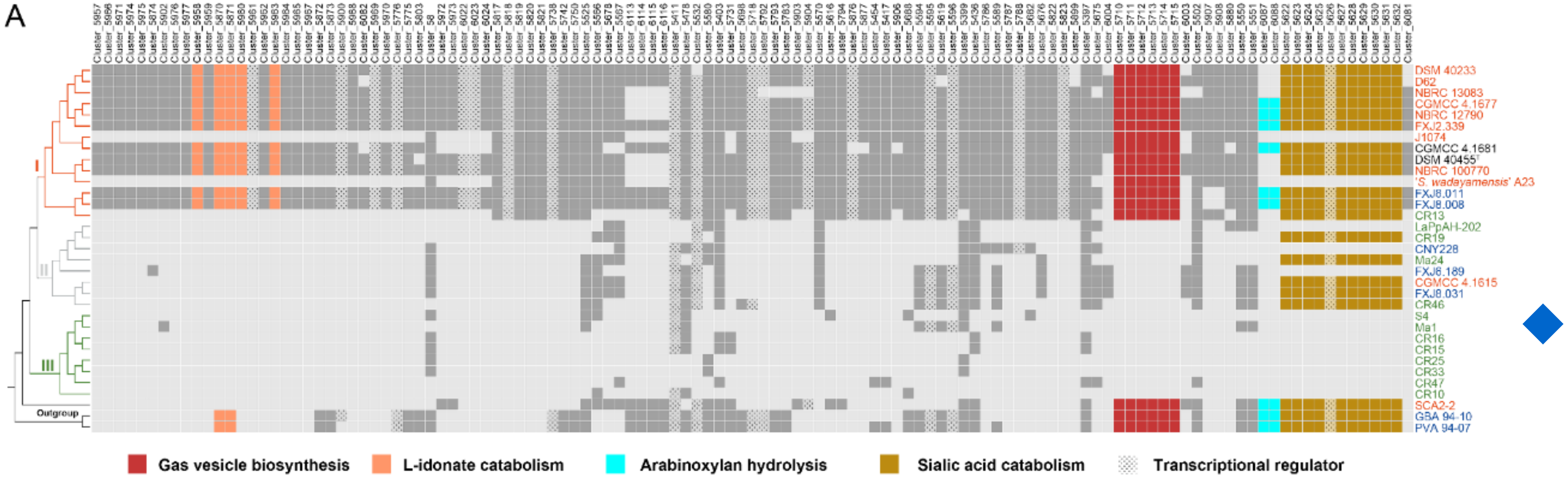


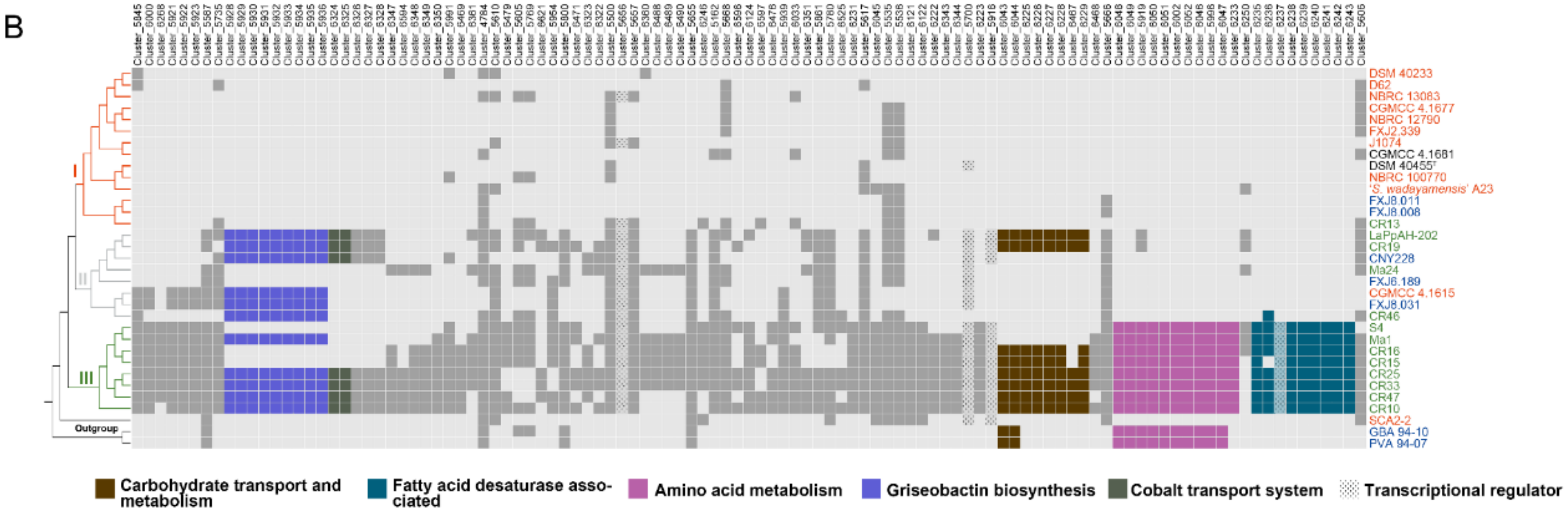
Diagram of interclade recombination

GWAS揭示生境相关的附属基因

Free-living



Entomic



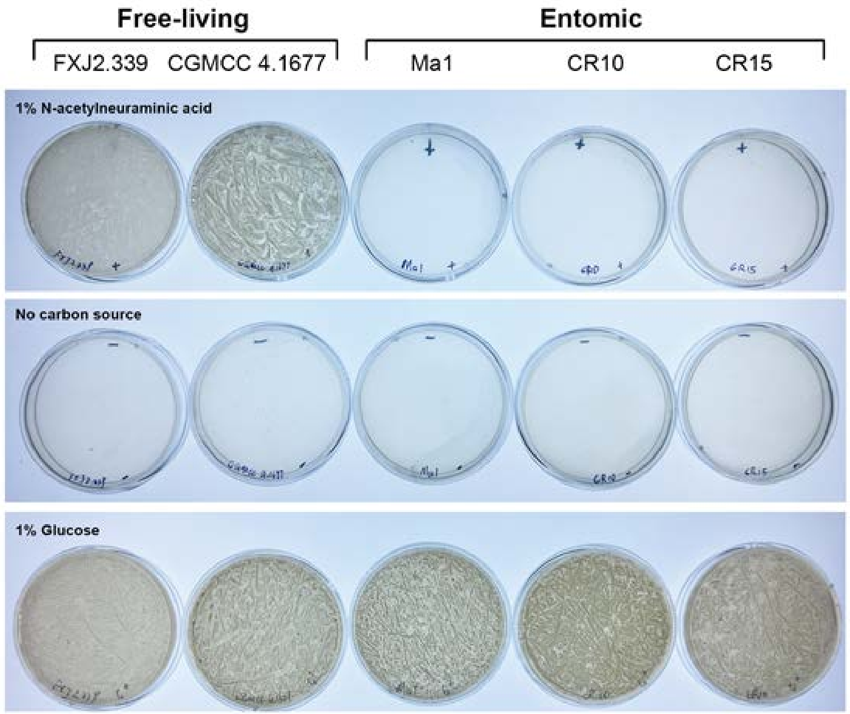
◆ 226 个与生境相关的附属基因

All habitat-associated OGs

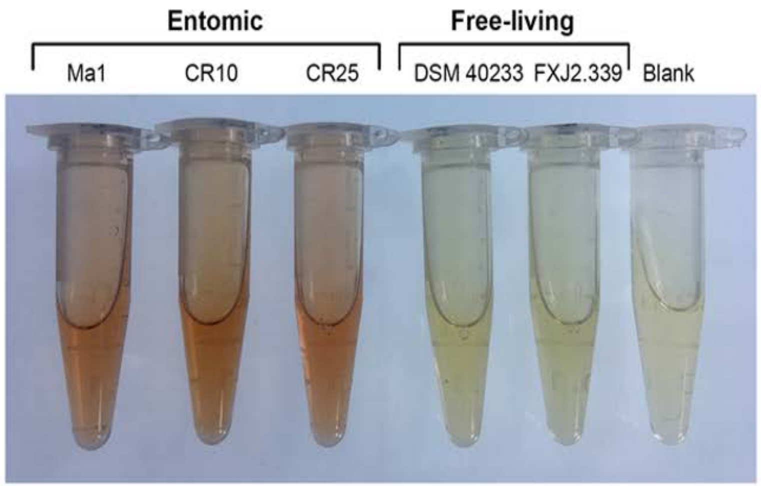
生境相关附属基因带来差异性适应功能

唾液酸分解代谢

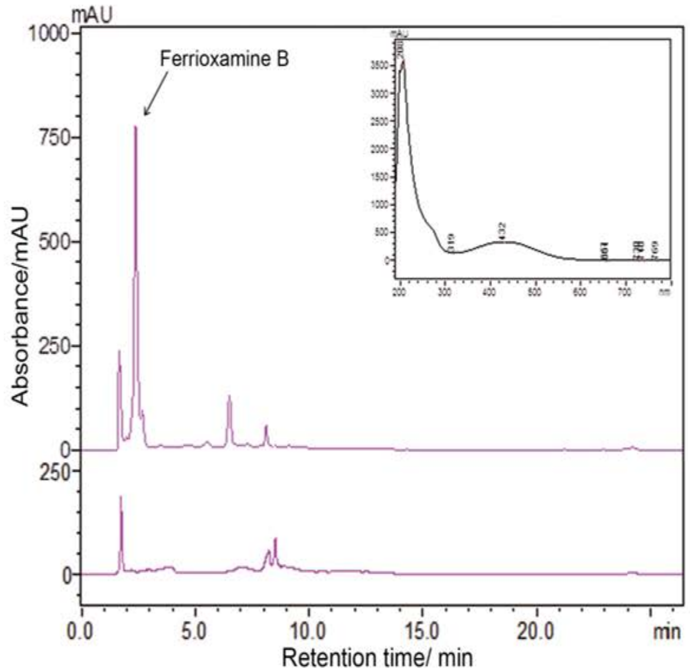
A



唾液酸分解代谢的功能验证



Griseobactin 生物合成基因簇的功能验证



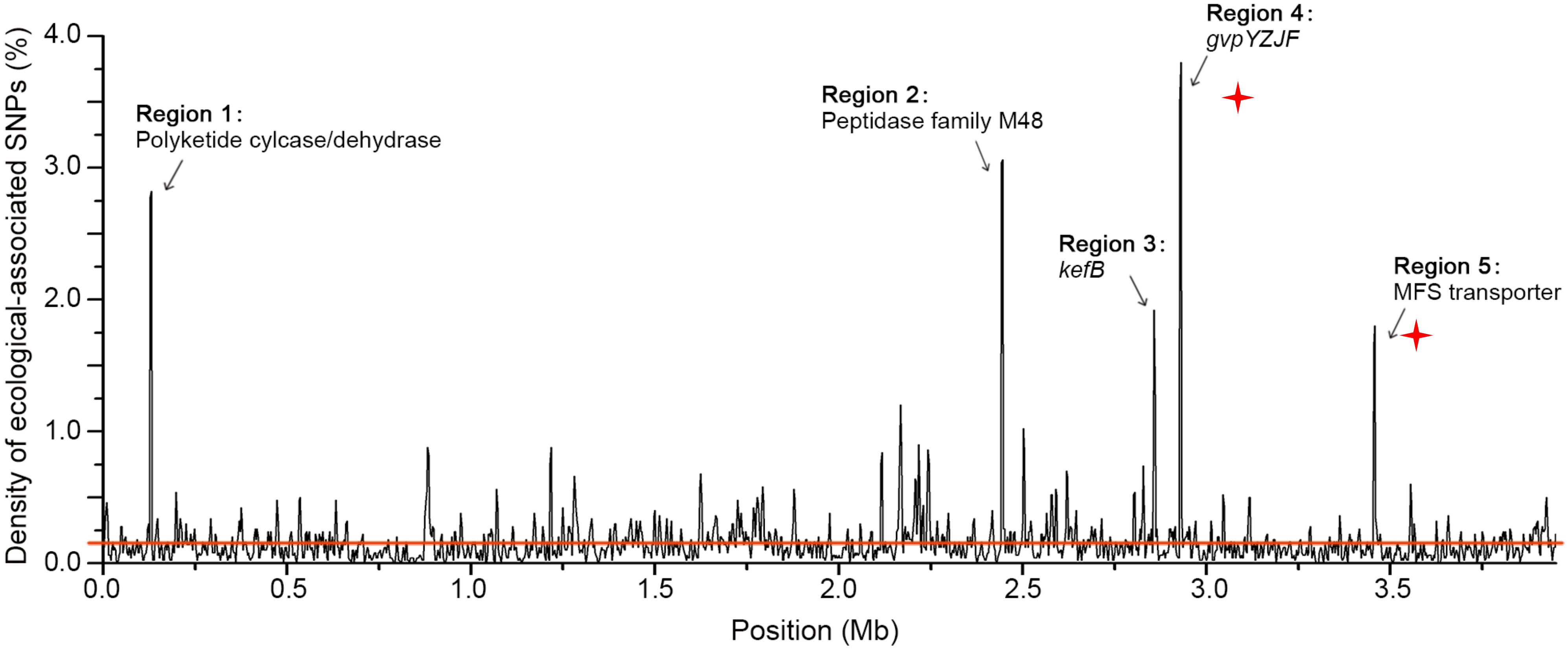
Desferrioxamine B 的检测

🟥 NRPS/NRPS related
🟩 DHBA biosynthesis
🟦 Siderophore export
🟫 Siderophore uptake/utilization
🟦 Cobalt transporter
☐ Hypothetical protein
⬜ Unrelated and conserved

Li et al. Appl Environ Microbiol, 2019

生境相关的SNPs形成基因组岛：生态分歧进行中

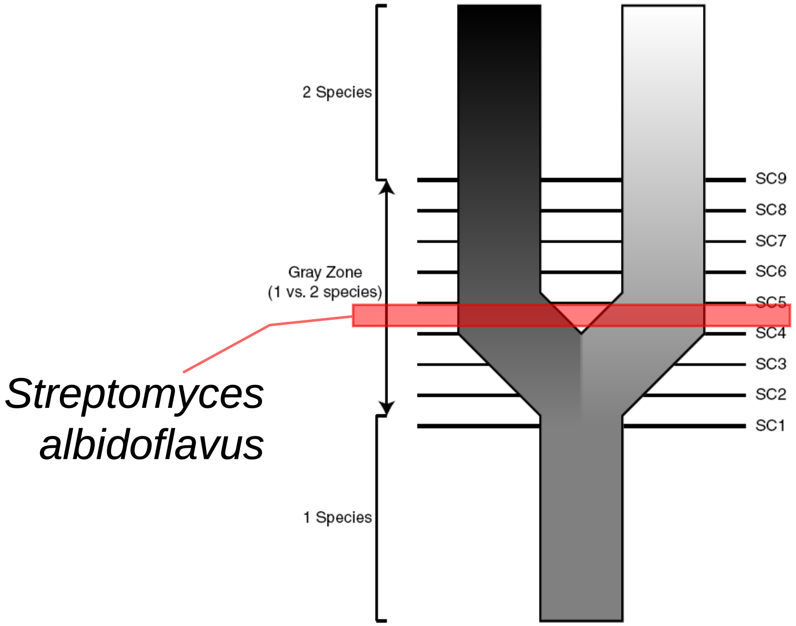
- 总SNPs: 169,594；生境相关SNPs: 6,511 (3.8%)
- 5个高度分歧的区域: gene-specific sweeps

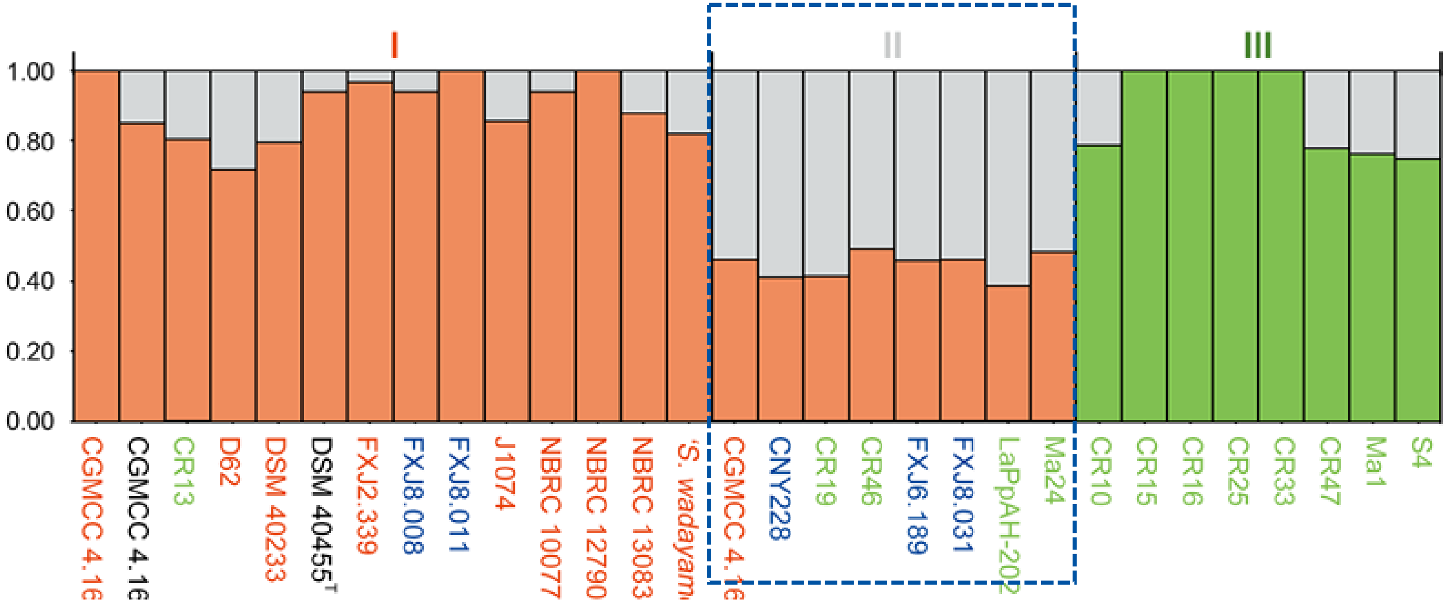


Li et al. Appl Environ Microbiol, 2019

微黄白链霉菌处于物种形成后期

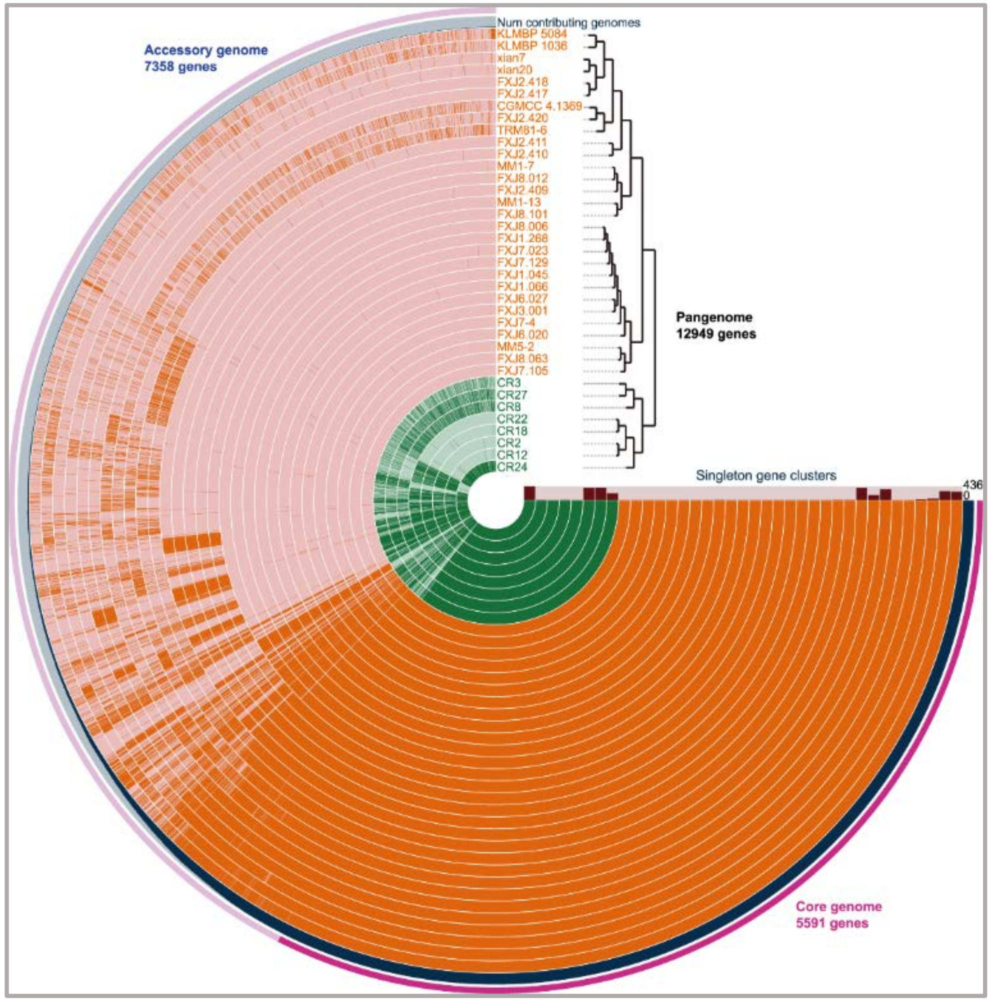
- 一系列与自由生活和昆虫生境显著相关的基因和位点，增强菌株适应性，导致基因组分化；
- 重组进一步塑造了微黄白链霉菌的种群结构；
- 物种分化过程已进入中后期，但仍未完成，分支 III 可能逐渐分化为更适应昆虫生境的新物种。

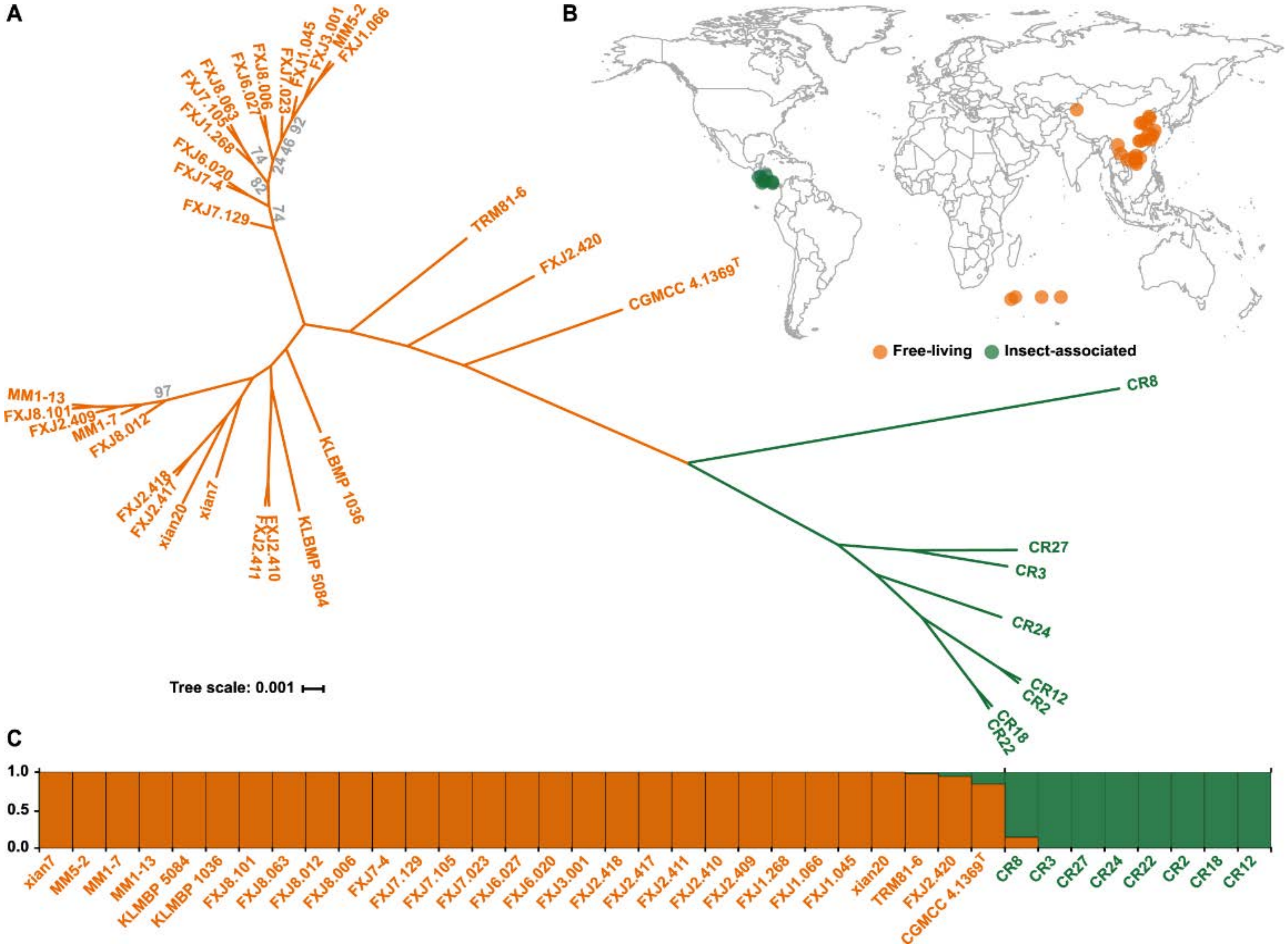




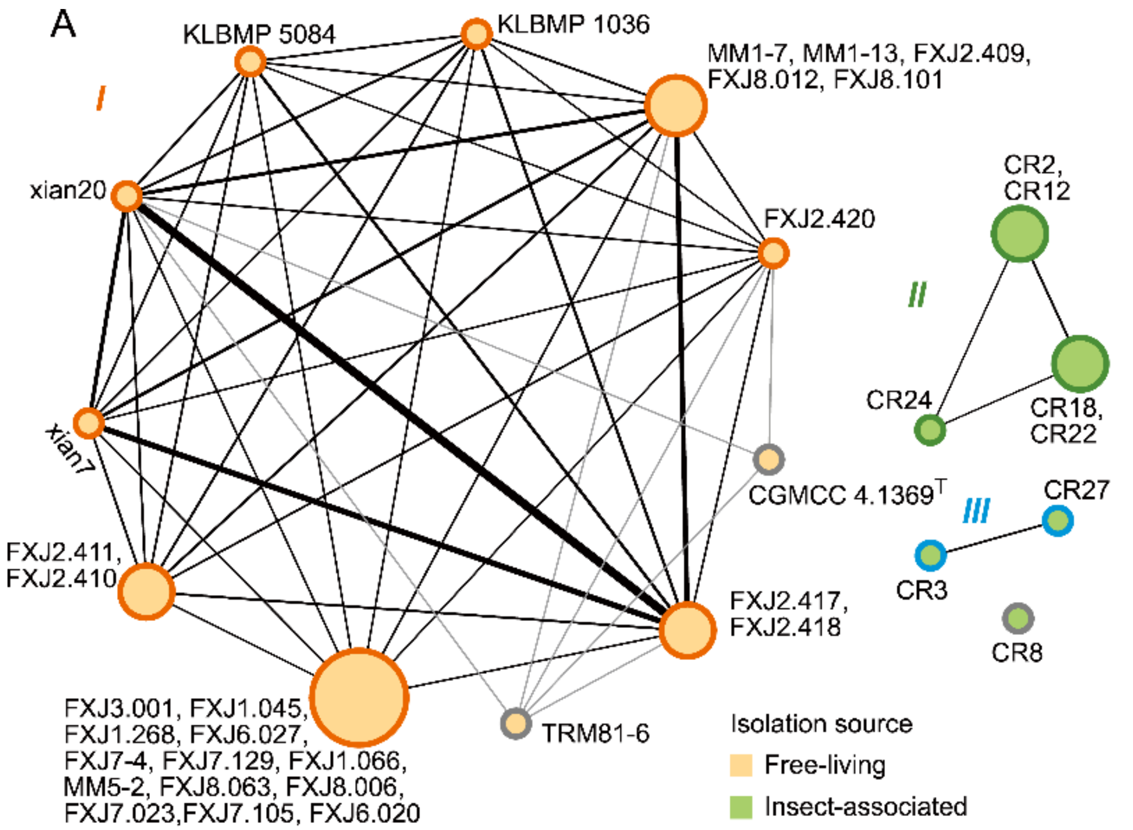
橄榄色链霉菌的系统发育与种群结构

- 形成两个稳健的进化分支，代表两个进化谱系，与**生境**和**地理**位置都能很好对应。

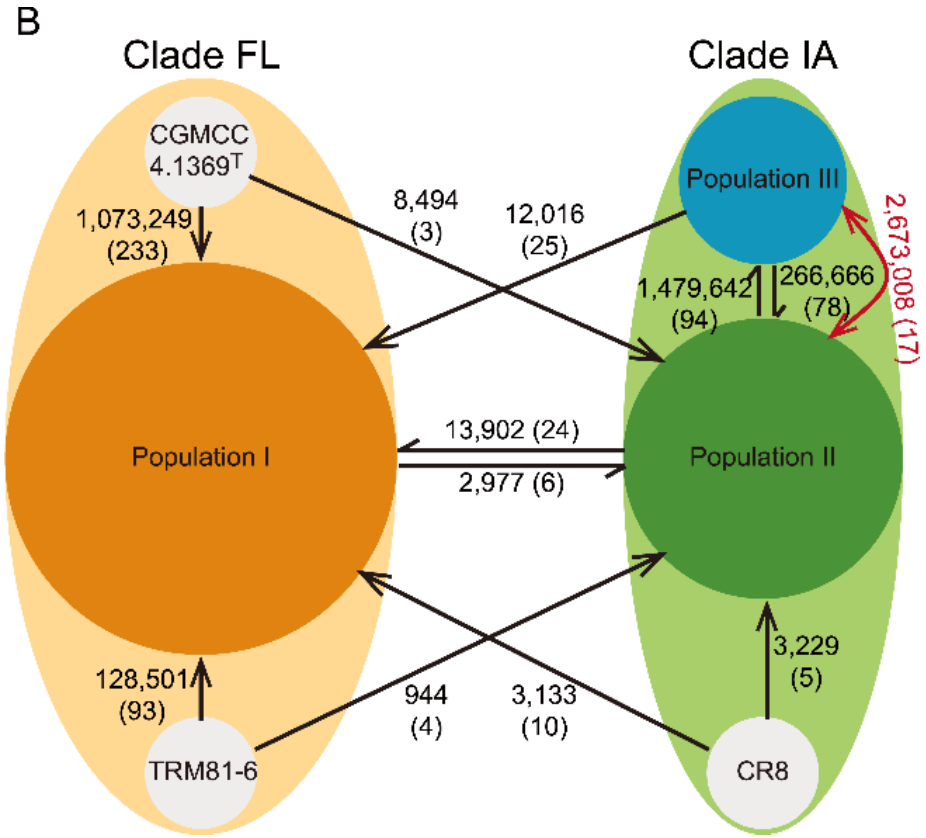




橄榄色链霉菌的基因交流模式



A. contemporary recombination pattern identified by **PopCOGenT**

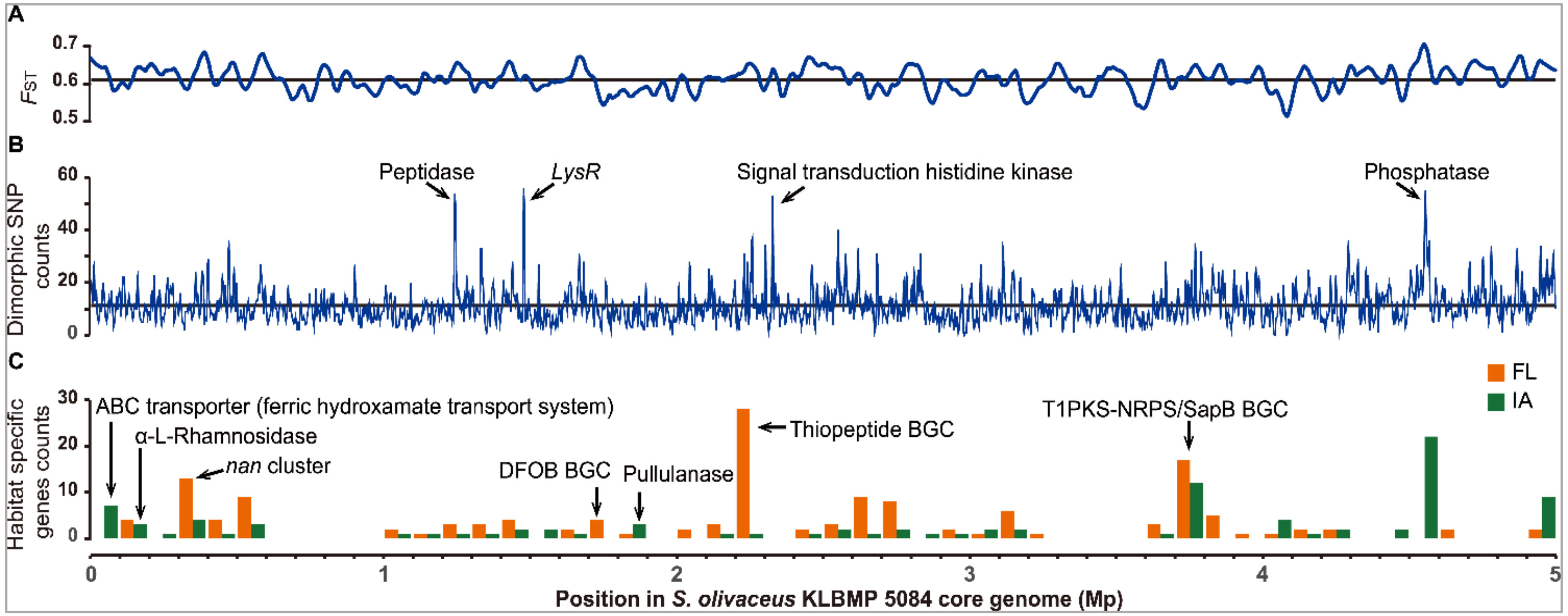


B. Diagram of recombination identified by fastGEAR

- 谱系内部的重组水平明显高于谱系间，且绝大多数重组事件为非祖先事件；
- 谱系间已经存在基因交流障碍。

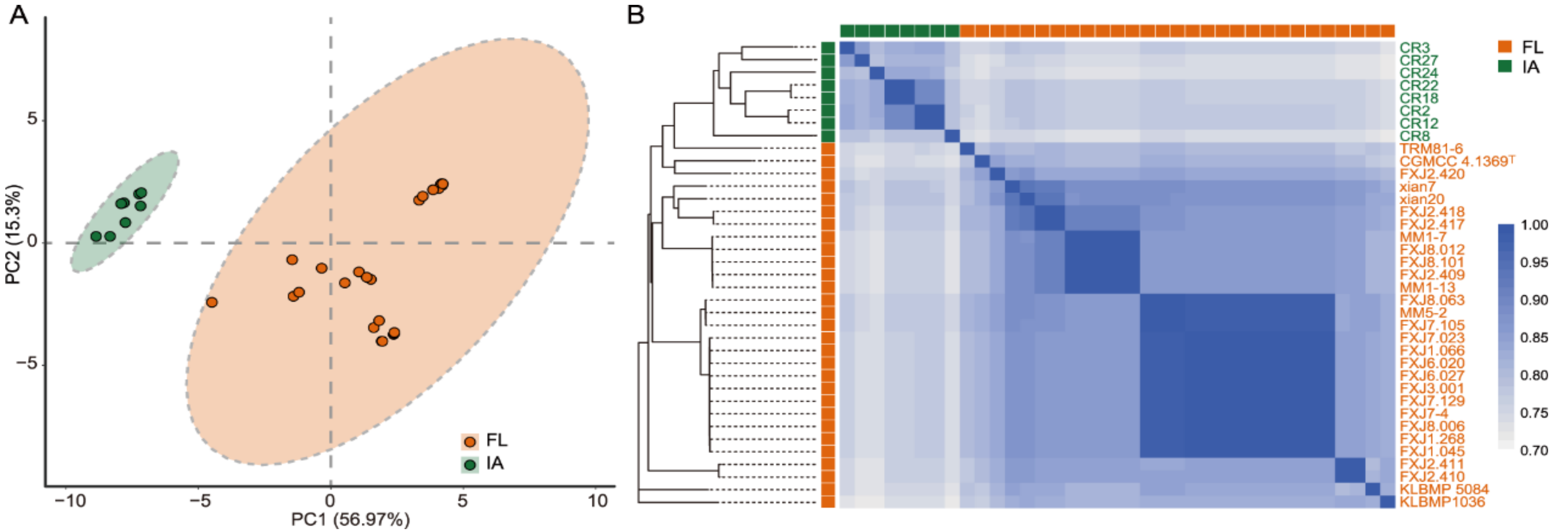
Wang et al. mBio, 2022

橄榄色链霉菌的基因组分化



核心基因组
的分化

◆ 两个谱系间已产生
全基因组分化

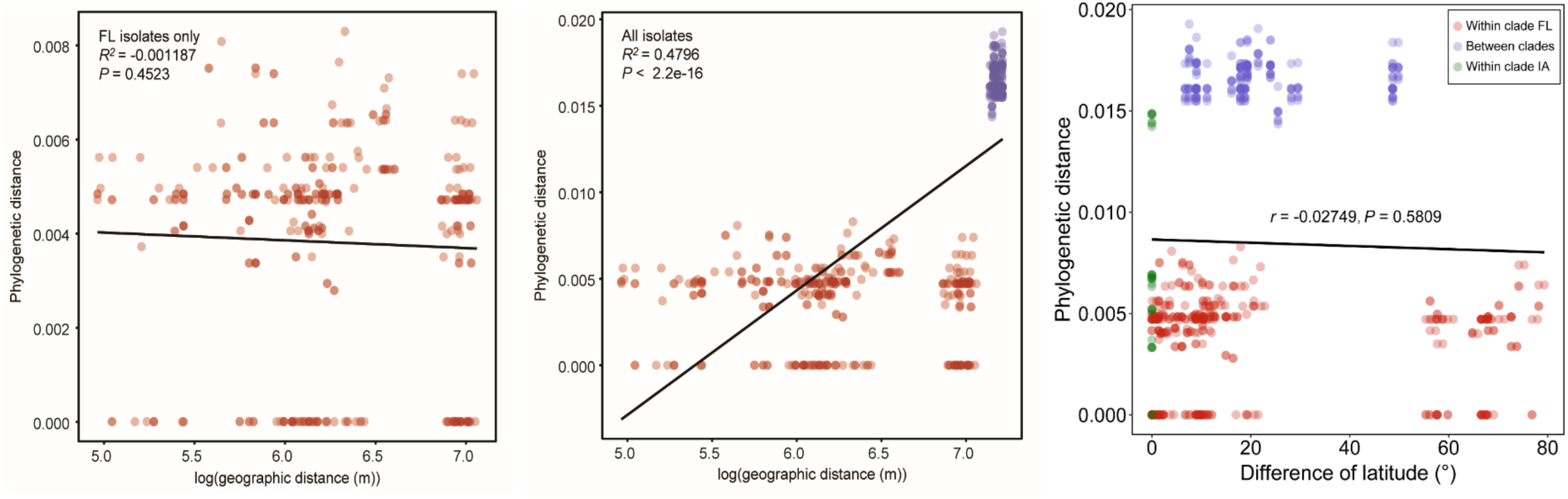


功能基因和附属
基因的分化

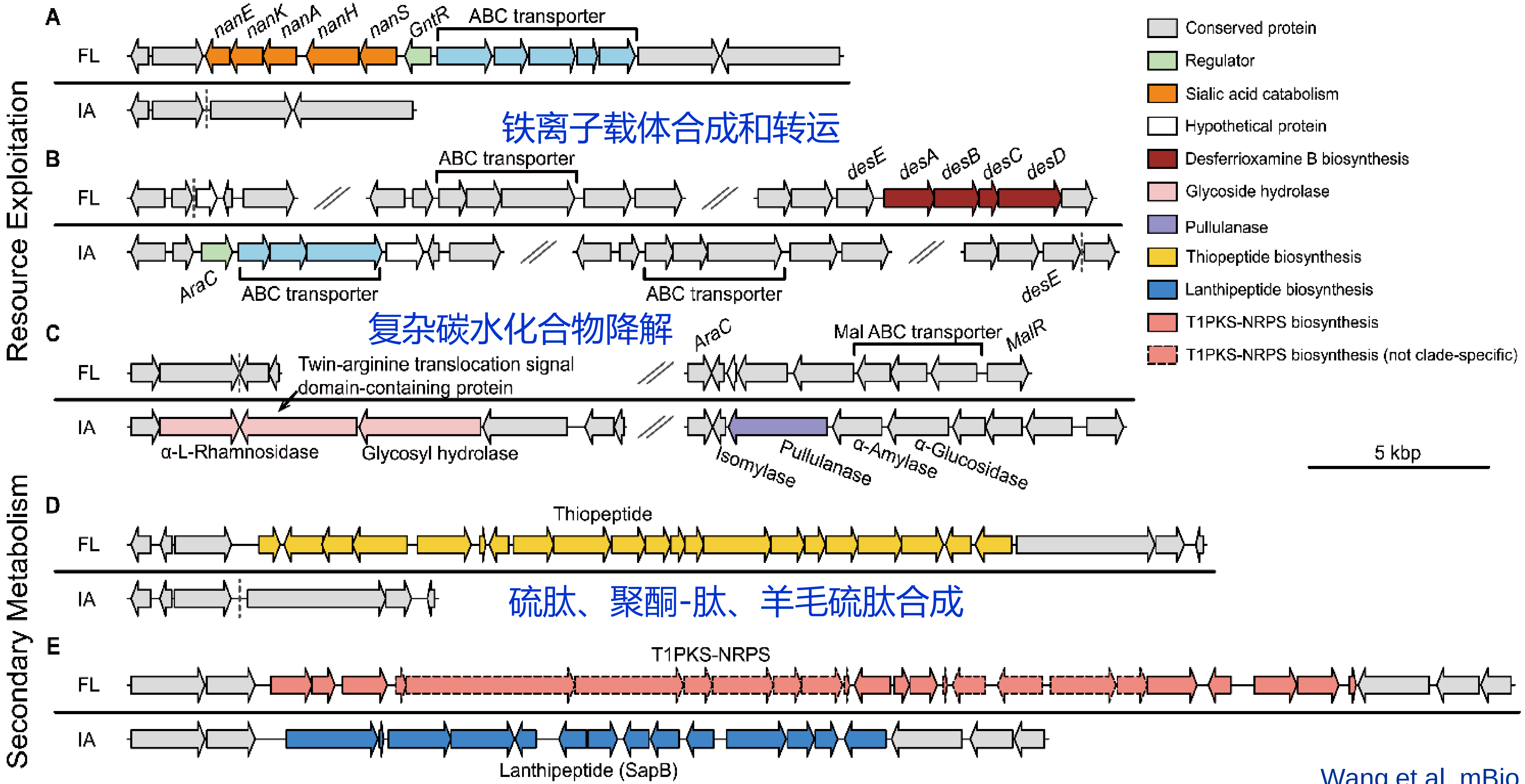
Wang et al. mBio, 2022

地理隔离和生态隔离对基因组分化的贡献

- 自由生活菌株：系统发育距离和地理距离之间无显著相关性， 菌株在洲际尺度的扩散不受限制；
- 所有菌株：**系统发育距离与生境类型显著相关(r = 0.935，P < 0.001)** ，而与地理距离和纬度均无关。



谱系/生境特异的功能基因(簇)



唾液酸分解代谢

铁离子载体合成和转运

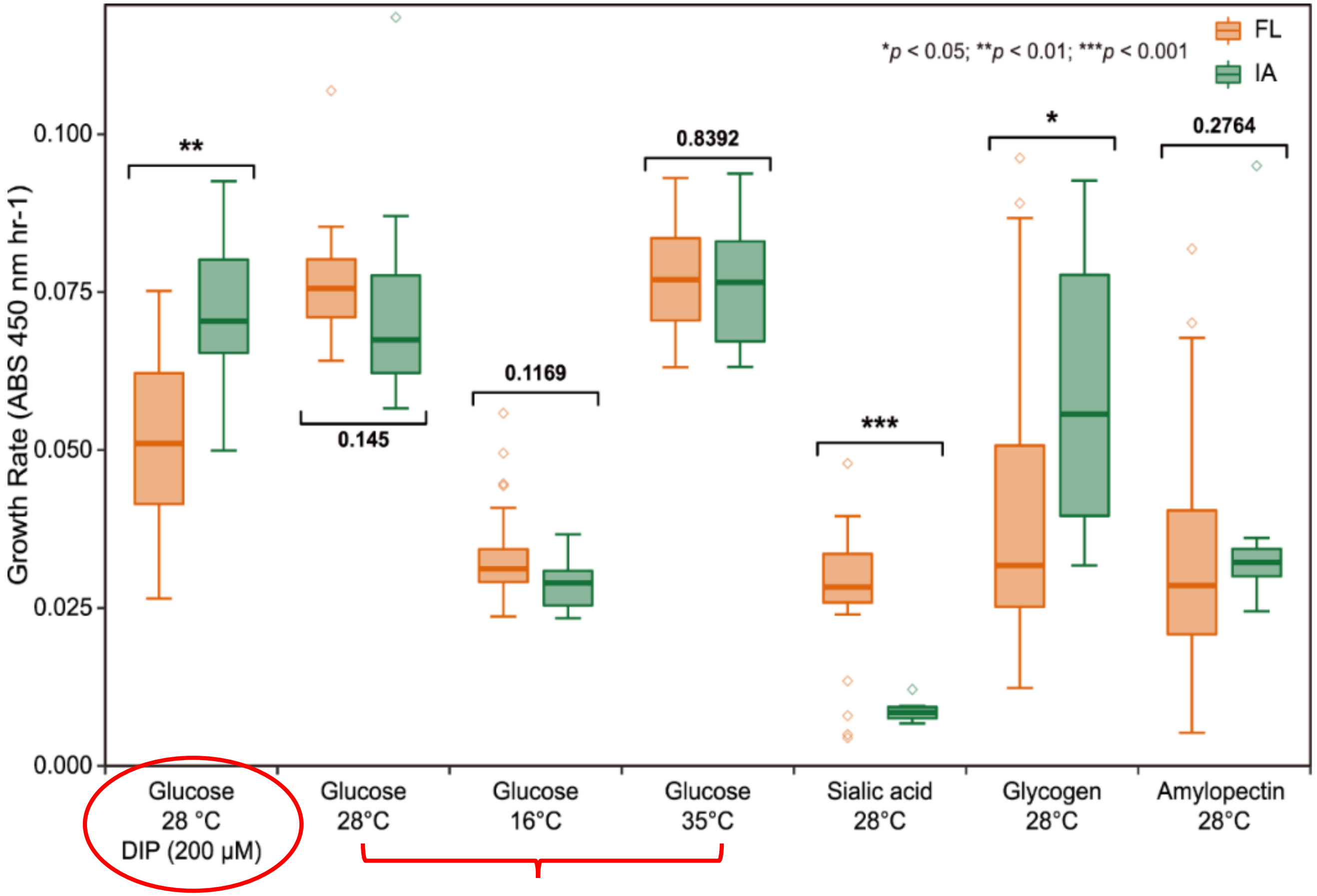
复杂碳水化合物降解

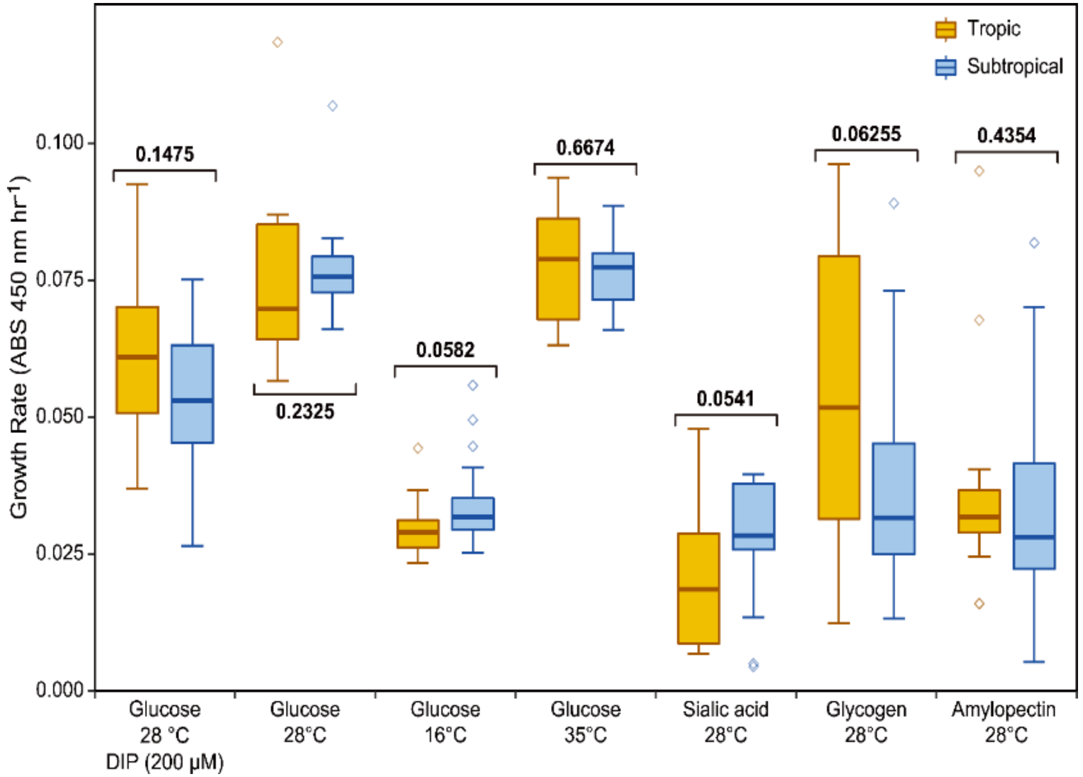
硫肽、聚酮-肽、羊毛硫肽合成

Wang et al. mBio, 2022

生境相关的生理特征分化

- 游离铁匮乏的条件下，昆虫谱系比自由生活谱系具有更高的适合度（fitness）
- 两个谱系在不同温度下的适合度没有显著区别，而唾液酸和糖原利用可以很好区分两个谱系；纬度分组无显著差异

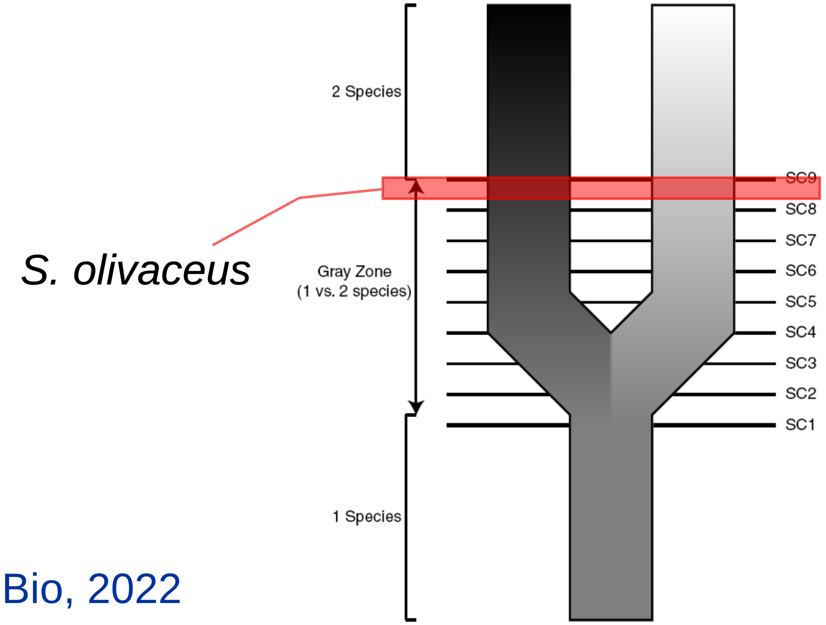


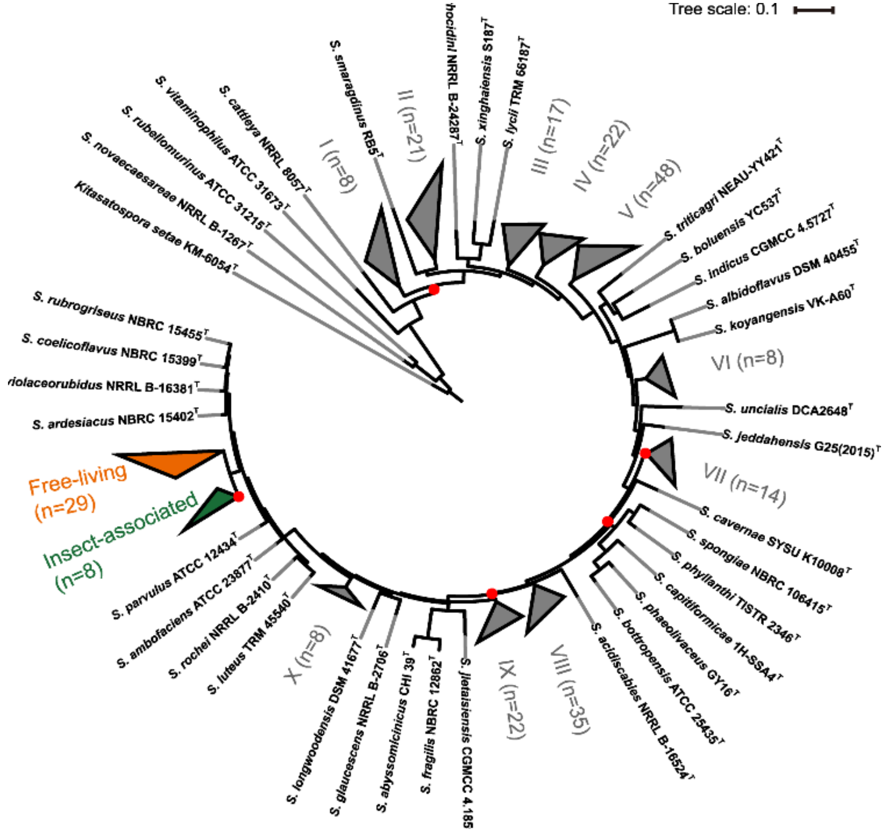


Wang et al. mBio, 2022

橄榄色链霉菌已完成近期的物种形成

- ◆ 生境适应性过程，而非大尺度的地理隔离，驱动了亲缘关系密切的链霉菌之间出现遗传分化；
- ◆ 种群无法跨生境扩散和定殖，从而阻碍了它们之间的基因交流，最终导致了新物种形成；
- ◆ 橄榄色链霉菌包括两个新近分化的隐蔽种。



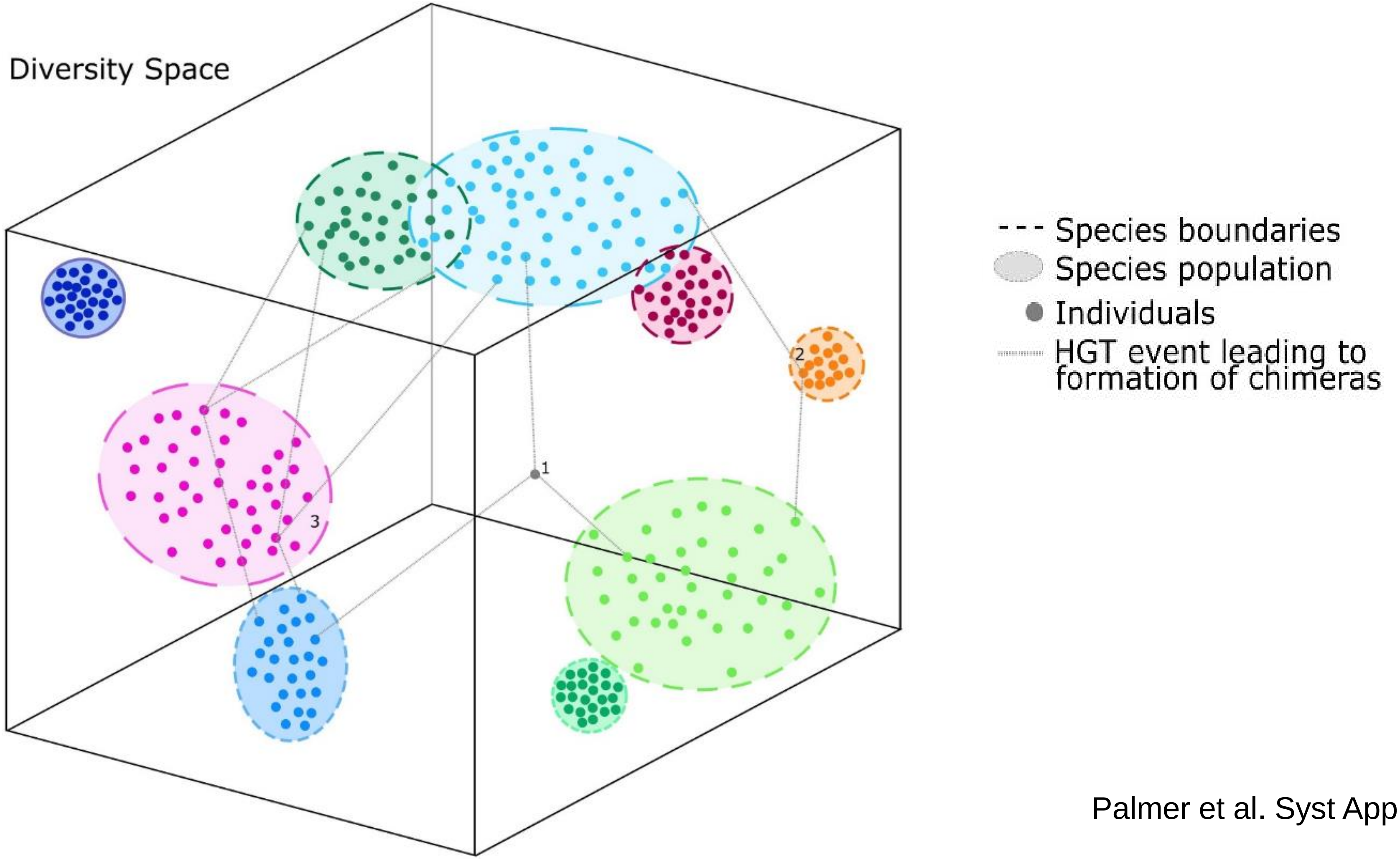


Wang et al. mBio, 2022

微生物物种形成模型

◆ **物种形成的集成模型**

描述微生物物种多样性空间的结构；物种是具有凝聚力的独立进化群体



Palmer et al. Syst Appl Microbiol, 2019

谢谢！

NSFC

中国科学院
CHINESE ACADEMY OF SCIENCES

中國大洋協會
COMRA