

# An introduction of Integrated Prokaryotes Genome and pan-genome Analysis (IPGA) platform



中国科学院微生物研究所  
Institute of Microbiology, Chinese Academy of Sciences



国家微生物科学数据中心  
National Microbiology Data Center



Dongmei Liu, Yifei Zhang, Guomei Fan, *et al.* 2022. IPGA: a handy integrated prokaryotes genome and pan-genome analysis web service. *iMeta* 1: e55. <https://doi.org/10.1002/imt2.55>

# Background

In the fields of molecular biology and genetics, a pan-genome is the union of all the genomes of a clade.

Pan-genome analysis has been shown to be an effective approach to better understand a clade of pathogenic bacteria because it helps developing various and tailored therapeutic strategies on the basis of their biological similarities and differences.

Kim Y, et al. Current opinion in biotechnology, 2020

Perspective | Published: 20 April 2022

## The Human Pangenome Project: a global resource to map genomic diversity

[Ting Wang](#) ✉, [Lucinda Antonacci-Fulton](#), [Kerstin Howe](#), [Heather A. Lawson](#), [Julian K. Lucas](#), [Adam M. Phillippy](#), [Alice B. Popejoy](#), [Mobin Asri](#), [Caryn Carson](#), [Mark J. P. Chaisson](#), [Xian Chang](#), [Robert Cook-Deegan](#), [Adam L. Felsenfeld](#), [Robert S. Fulton](#), [Erik P. Garrison](#), [Nanibaa' A. Garrison](#), [Tina A. Graves-Lindsay](#), [Hanlee Ji](#), [Eimear E. Kenny](#), [Barbara A. Koenig](#), [Daofeng Li](#), [Tobias Marschall](#), [Joshua F. McMichael](#), [Adam M. Novak](#), [the Human Pangenome Reference Consortium](#) [+ Show authors](#)

*Nature* 604, 437–446 (2022) | [Cite this article](#)



Article | [Open Access](#) | Published: 15 January 2018

## Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice

[Qiang Zhao](#), [Qi Feng](#), [Hengyun Lu](#), [Yan Li](#), [Ahong Wang](#), [Qilin Tian](#), [Qilin Zhan](#), [Yiqi Lu](#), [Lei Zhang](#), [Tao Huang](#), [Yongchun Wang](#), [Danlin Fan](#), [Yan Zhao](#), [Ziqun Wang](#), [Congcong Zhou](#), [Jiaying Chen](#), [Chuanran Zhu](#), [Wenjun Li](#), [Qijun Weng](#), [Qun Xu](#), [Zi-Xuan Wang](#), [Xinghua Wei](#), [Bin Han](#) & [Xuehui Huang](#) ✉

*Nature Genetics* 50, 278–284 (2018) | [Cite this article](#)



Article | Published: 13 May 2019

## The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor

[Lei Gao](#), [Itay Gonda](#), [Honghe Sun](#), [Qiyue Ma](#), [Kan Bao](#), [Denise M. Tieman](#), [Elizabeth A. Burzynski-Chang](#), [Tara L. Fish](#), [Kaitlin A. Stromberg](#), [Gavin L. Sacks](#), [Theodore W. Thannhauser](#), [Majid R. Foolad](#), [Maria Jose Diez](#), [Jose Blanca](#), [Joaquin Canizares](#), [Yimin Xu](#), [Esther van der Knaap](#), [Sanwen Huang](#), [Harry J. Klee](#), [James J. Giovannoni](#) ✉ & [Zhangjun Fei](#) ✉

*Nature Genetics* 51, 1044–1051 (2019) | [Cite this article](#)

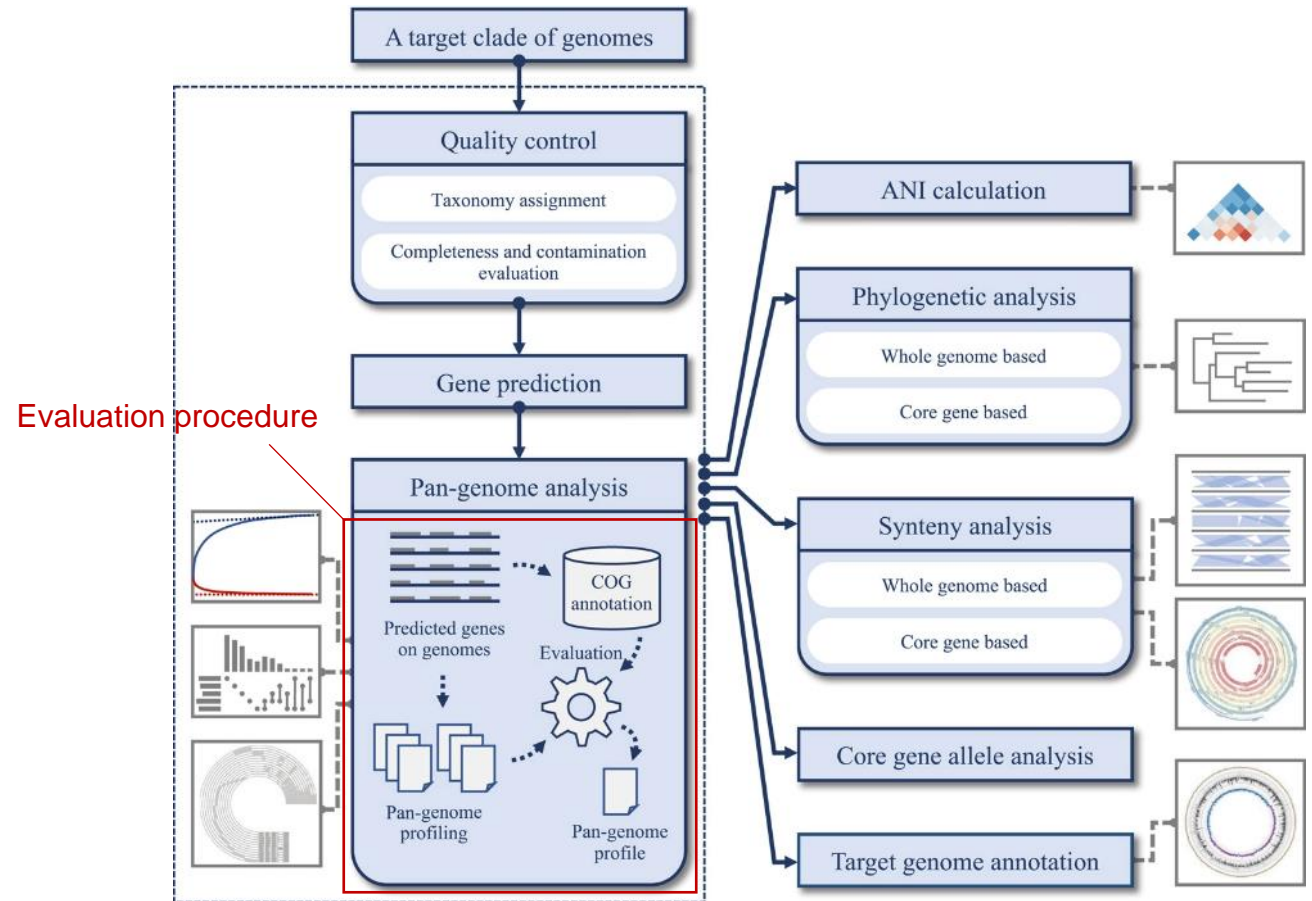


# Methods

SOFTWARE	ORTHOLOGY ANALYSIS	REFERENCES
BPGA	USEARCH, CD-HIT and OrthoMCL	Chaudhari et al <sup>48</sup>
EDGAR and EDGAR 2.0	Score ratio values	Blom et al <sup>53</sup>
GET_HOMOLOGUES	Bidirectional best-hit, COGtriangles, or OrthoMCL	Contreras-Moreira and Vinuesa <sup>26</sup>
PanDelos	Dictionary-based method	Bonnici et al <sup>54</sup>
Panseq	MUMmer and BLASTn	Laing et al <sup>55</sup>
PanWeb	PGAP	Pantoja et al <sup>62</sup>
PanX	Diamond and MCL	Ding et al <sup>58</sup>
PGAP	Inparanoid, MultiParanoid and Gene Family	Zhao et al <sup>61</sup>
PGAT	BLASTp (sequence alignment of >80% and sequence identity >91%-92%)	Brittnacher et al <sup>64</sup>
PGAweb	PGAP and PGAP-x modules	Chen et al <sup>63</sup>
Piggy <sup>b</sup>	Roary	Thorpe et al <sup>66</sup>
Roary	CD-HIT, BLAST and MCL	Page et al <sup>65</sup>

<sup>a</sup>Not mentioned in manuscript.

<sup>b</sup>Pan-genome analysis of intergenic regions.

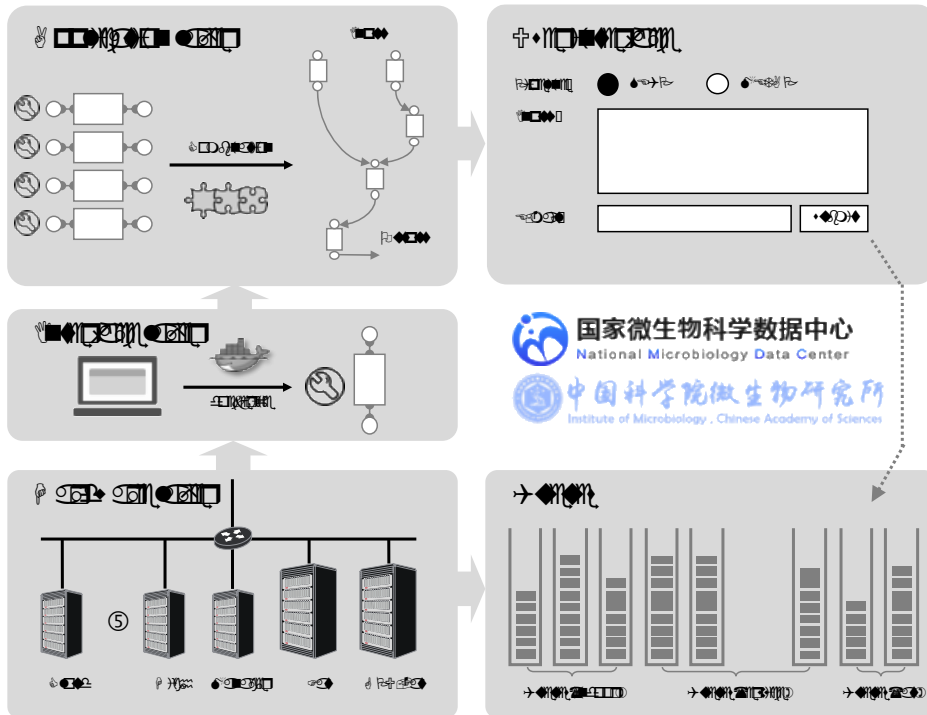


IPGA workflow

# Introduction

Service status: **RUNNING** Running tasks: 0 Total tasks: 295

Service status on page header



High-performance computing Cloud Services Framework



Service status: **RUNNING** Running tasks: 0 Total tasks: 323

Home page Analysis Output query Download **Manual** Video Q&A

Load demo input files and arguments

Public genome selection

Public genome selection interface with a search box and a 'Check' button.

Private genome submission

Private genome submission interface with a 'Drag & Drop the FASTA file to here!' area and a 'Confirm' button.

## 0. Quick start

### 1. Overview of IPGA

#### 2. Instructions

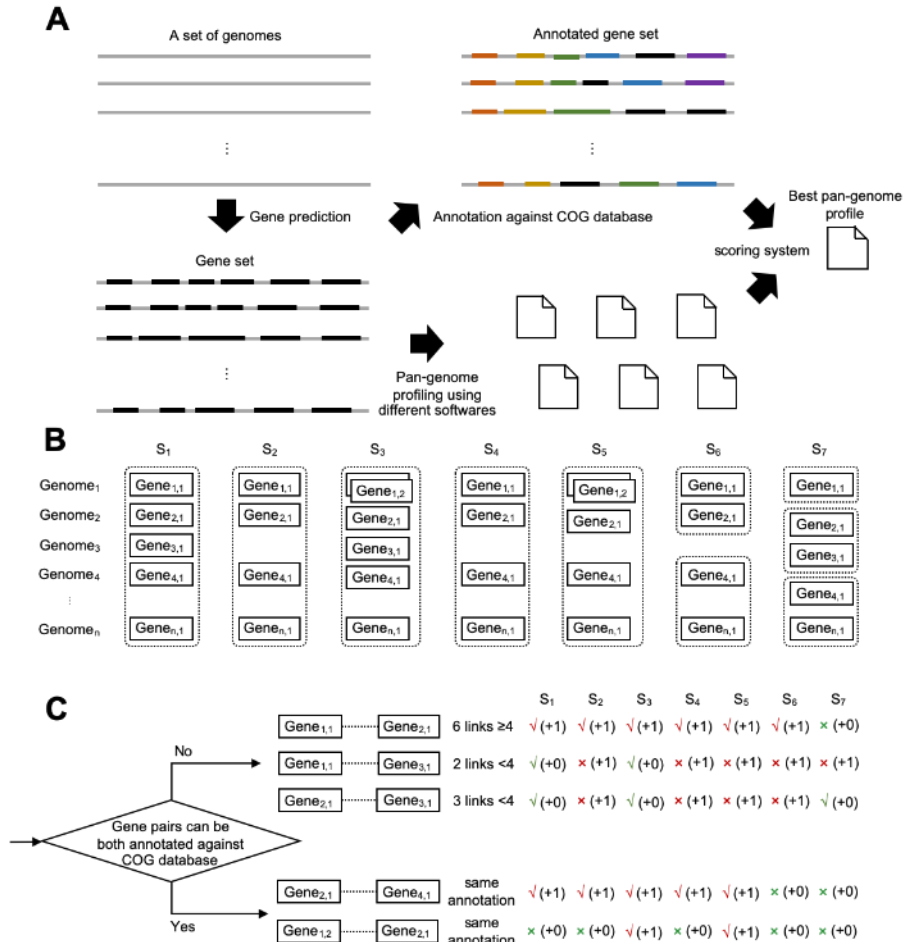
- 2.1 Input requirements
  - 2.1.1 Public genome selection
  - 2.1.2 Private genome selection
  - 2.1.3 Confirm the data
- 2.2 Analysis module and parameter selection
  - 2.2.1 Quality control module
  - 2.2.2 Pan-genome analysis module
  - 2.2.3 Target genome selection
  - 2.2.4 Genome annotation module
  - 2.2.5 Submission

#### 3. Outputs

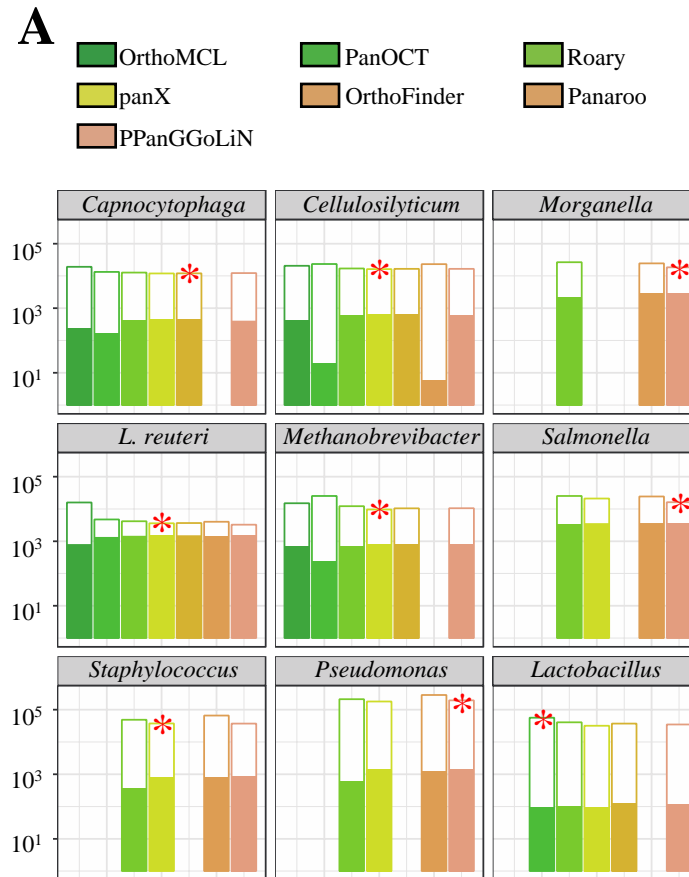
- 3.1 Output description
- 3.2 Analysis result data structure

User interface of IPGA

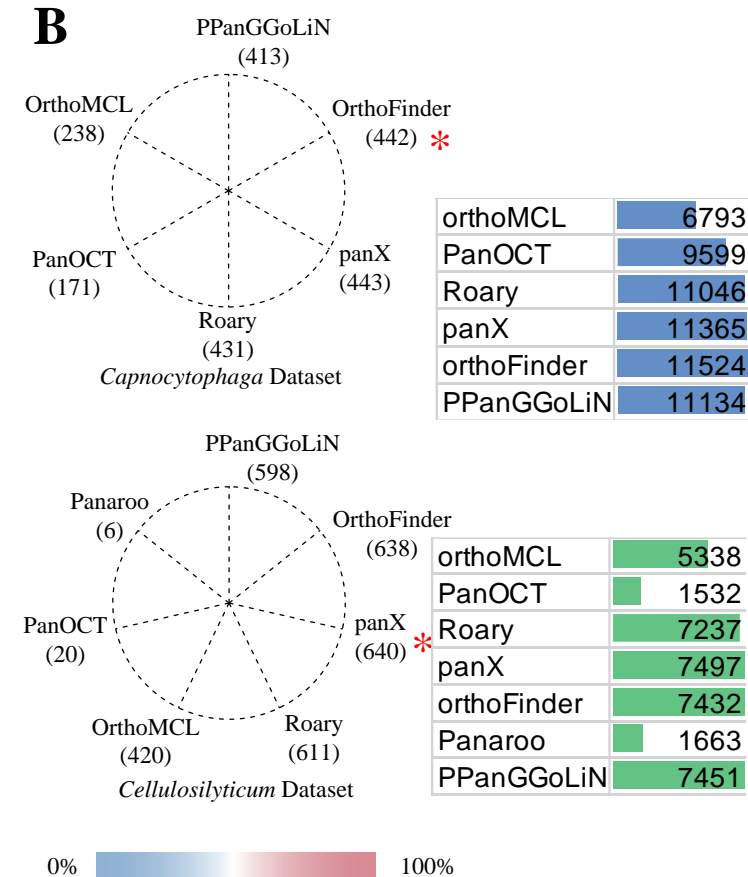
# Features



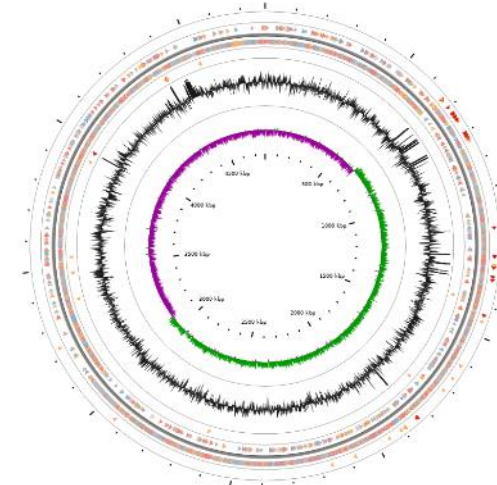
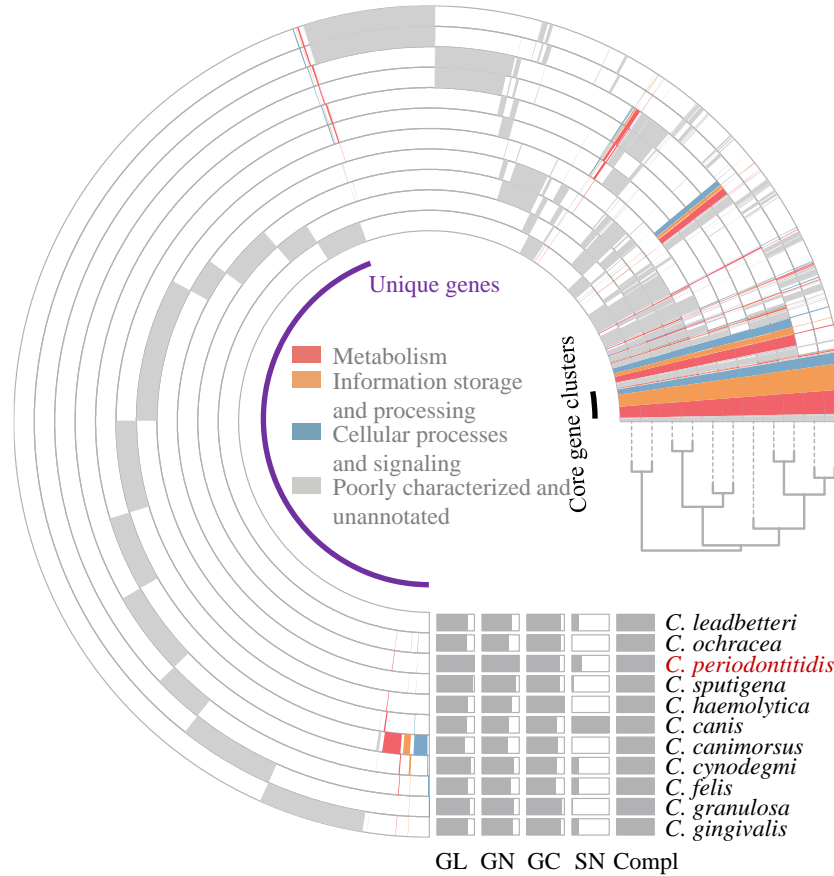
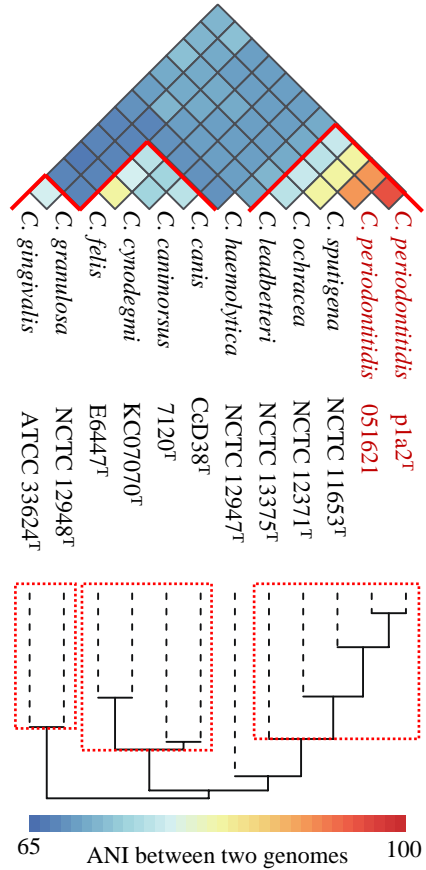
Pan-genome analysis module and the scoring system in IPGA for selecting the best pan-genome profile.



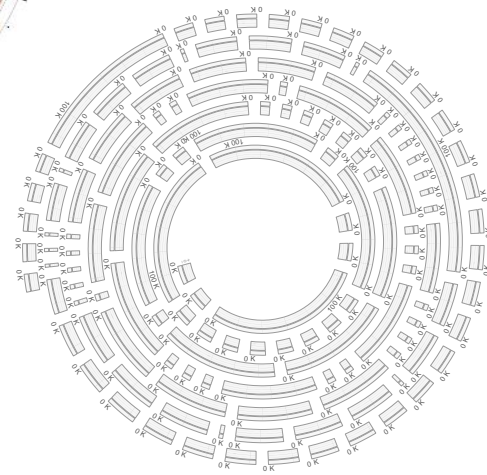
(A) The performance evaluation on nine different data sets. (B) The number of shared core gene links among four different softwares..



# Results



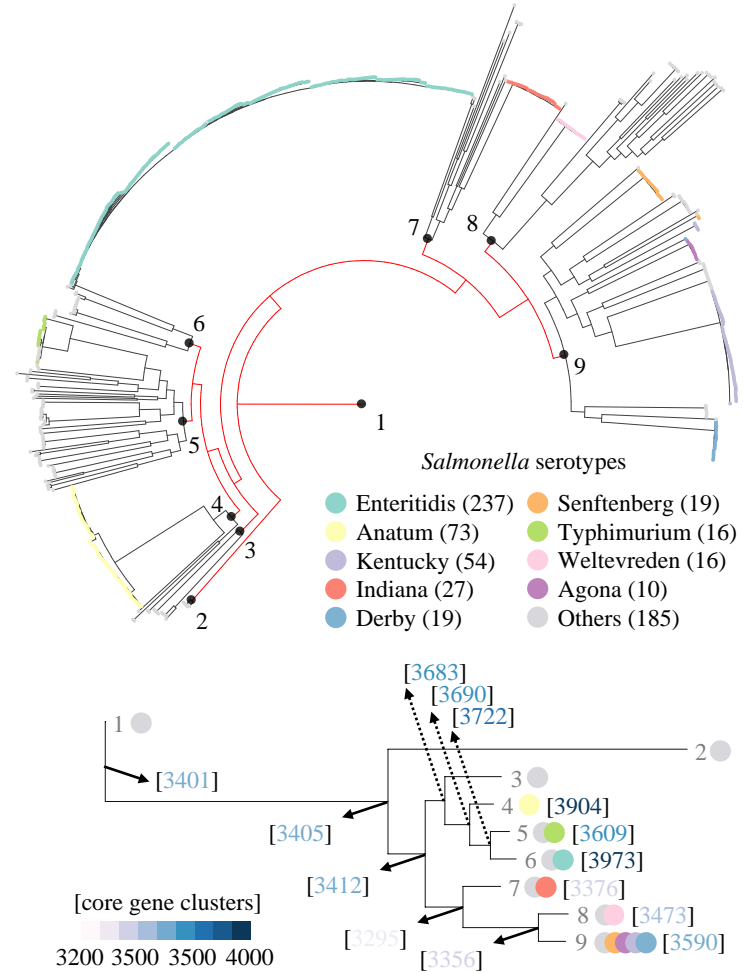
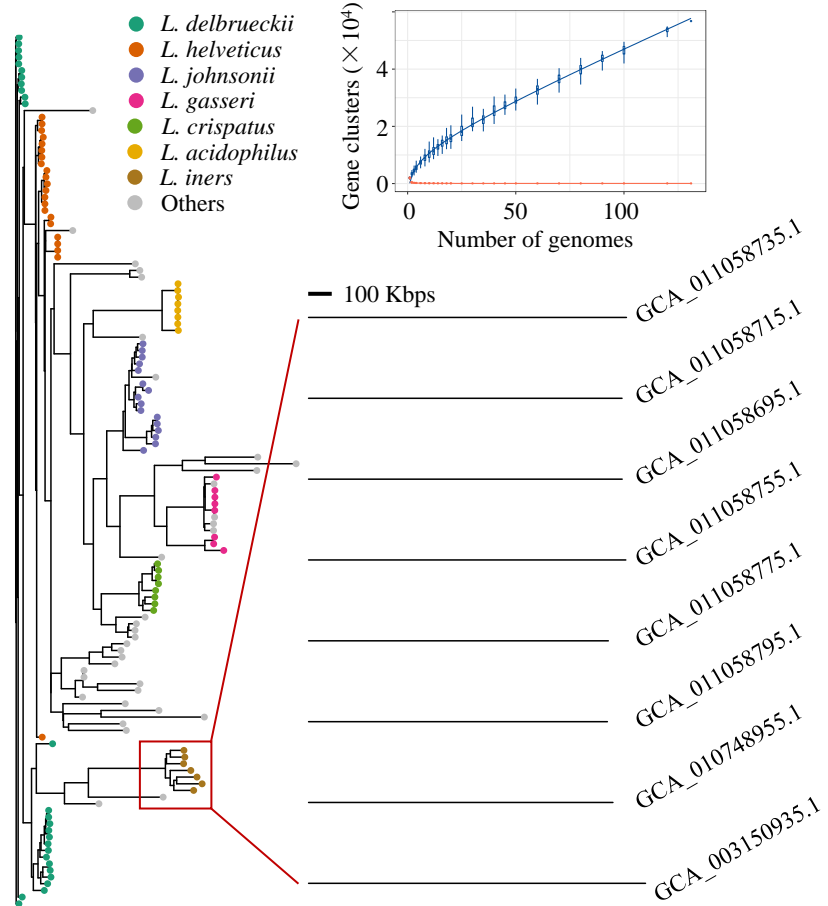
*Cellulosilyticum* sp. (strain WCF-2)  
isolated from cow feces  
(GCA\_003990395.1)



*Cellulosilyticum* sp. (MAG SIG270)  
isolated from goat feces  
(GCA\_015058045.1)



# Results



Thanks!

Contacts:  
shiw@cau.edu.cn  
ma@im.ac.cn



中国科学院微生物研究所  
Institute of Microbiology, Chinese Academy of Sciences



国家微生物科学数据中心  
National Microbiology Data Center



Dongmei Liu, Yifei Zhang, Guomei Fan, *et al.* 2022. IPGA: a handy integrated prokaryotes genome and pan-genome analysis web service. *iMeta* 1: e55. <https://doi.org/10.1002/imt2.55>