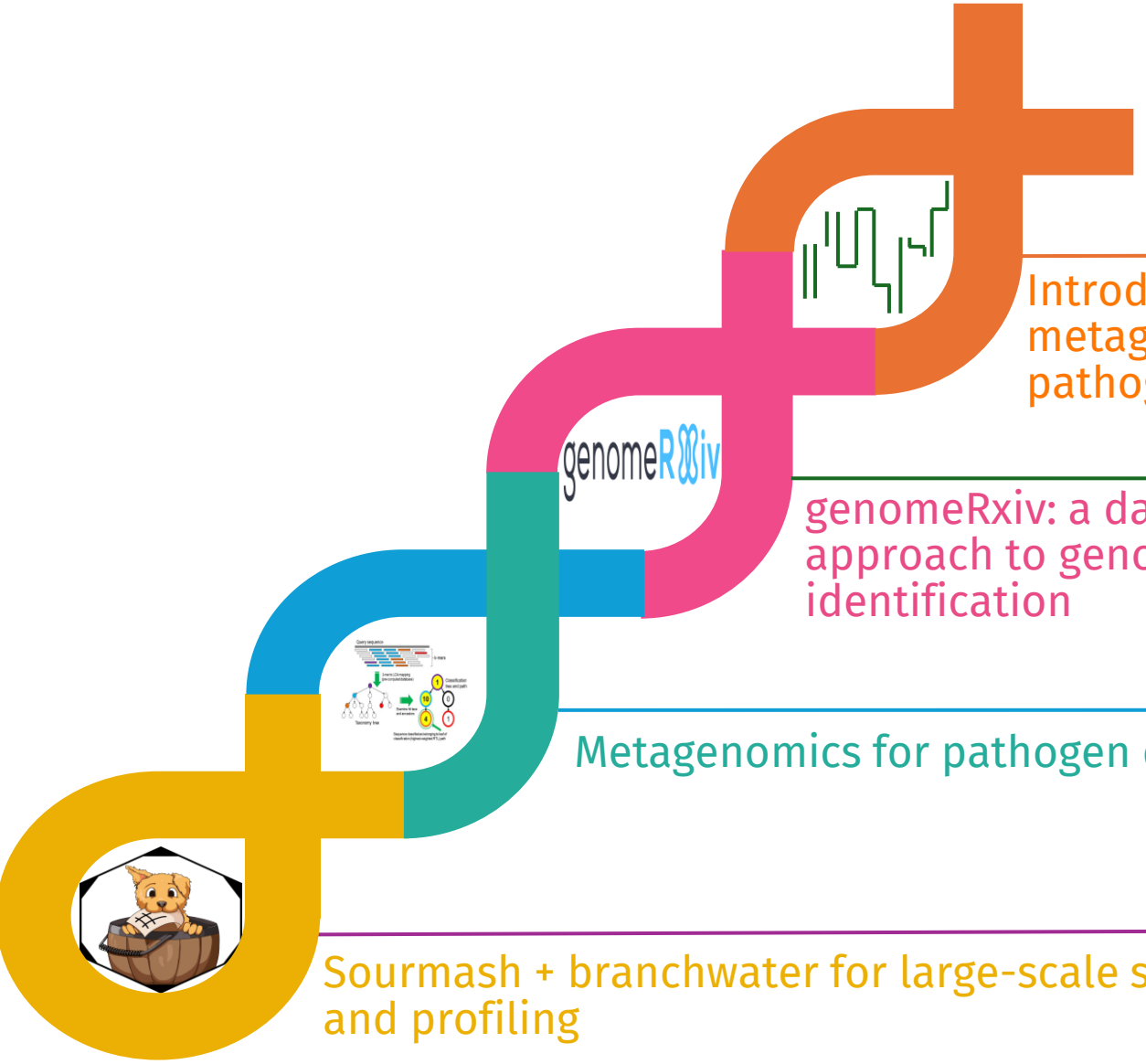
# genomeRxiv: a data driven approach to genome taxonomic identification

Reza Mazloom
ICPPB & Biocontrol
July 2024

# Genome/metagenome-based pathogen identification




Introduction to metagenome-based pathogen identification

1

**Sahar Abdelrazek, Ph.D.**

*VA-MD College of Veterinary Medicine, Virginia Tech*

genomeRxiv: a data driven approach to genome taxonomic identification

2

**Reza Mazloom**

*Department of Computer Science, Virginia Tech, VA*

Metagenomics for pathogen detection

3

**Parul Sharma, Ph.D.**

*School of Medicine, Division of infectious diseases*
*Emory University*

Sourmash + branchwater for large-scale sequence search and profiling

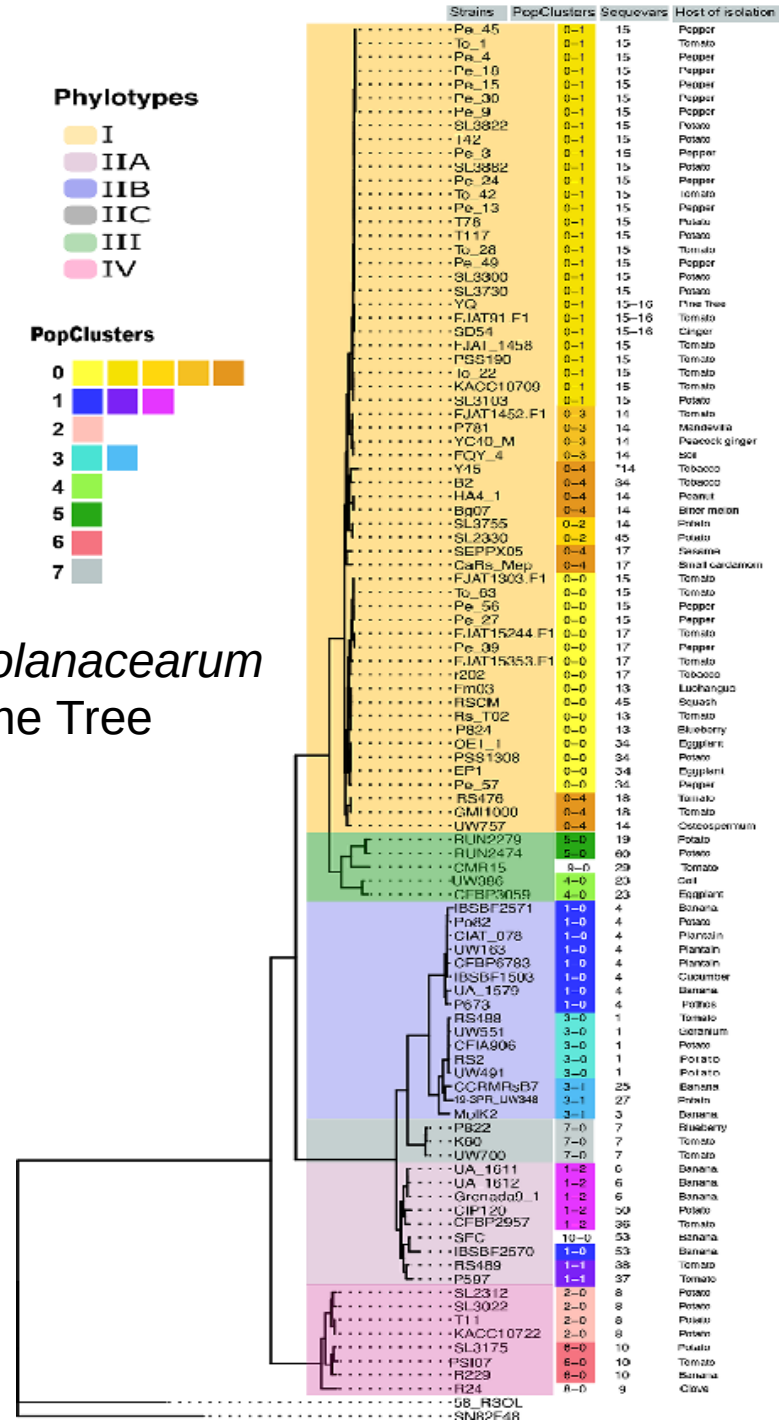
4

**Tessa Ward, Ph.D.**

*Department of Population Health and Reproduction*
*University of California, Davis*

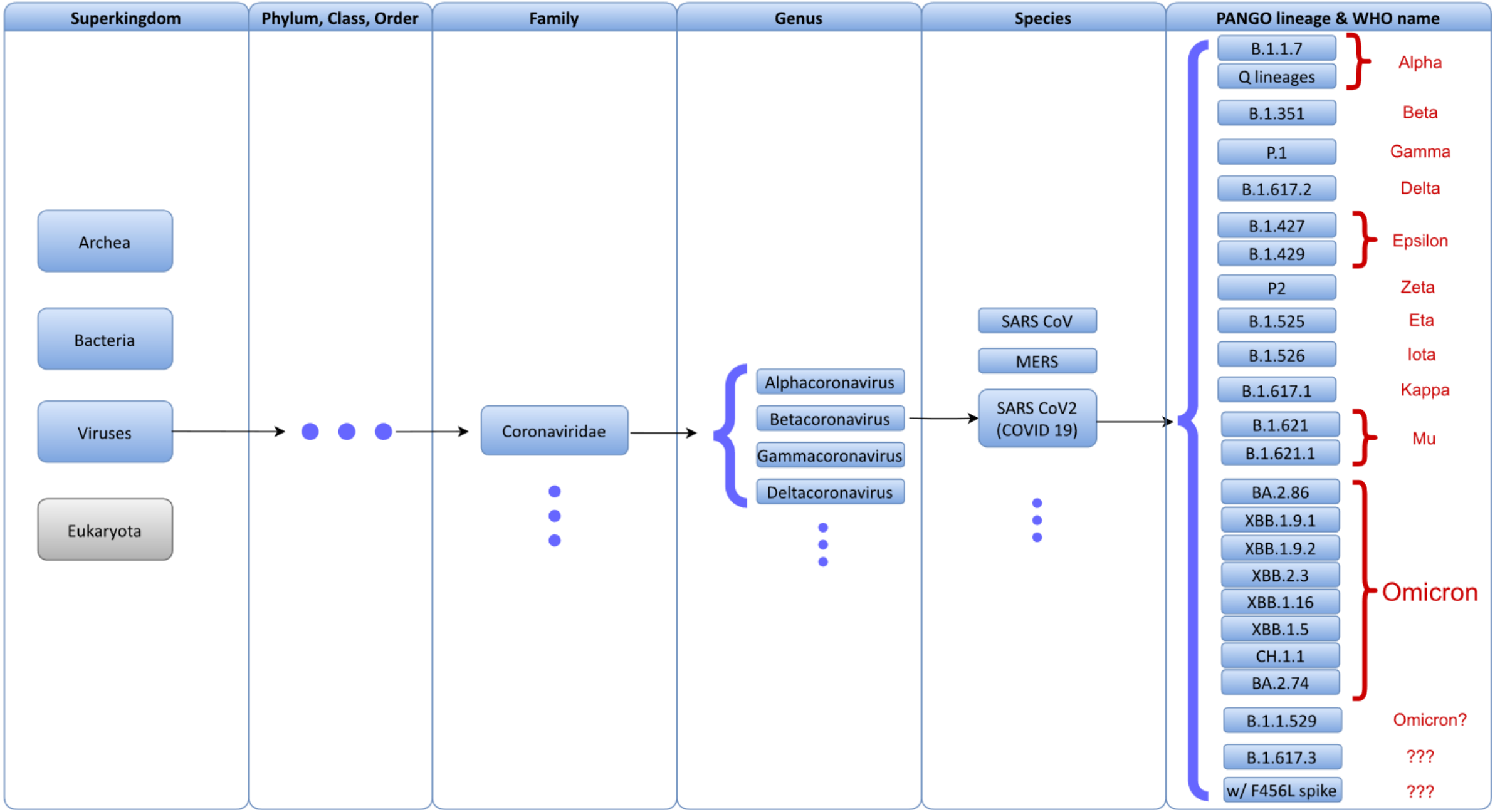
# The Problem!

- Great organism classification up to species
- Within species classification is left mostly untouched
- Unless the species has been of interest such as SARS-CoV-2
- What do we mean by different species?
- What is different between each species?
- We have mainly focused on Prokaryotes, Fungi, and Viruses until now



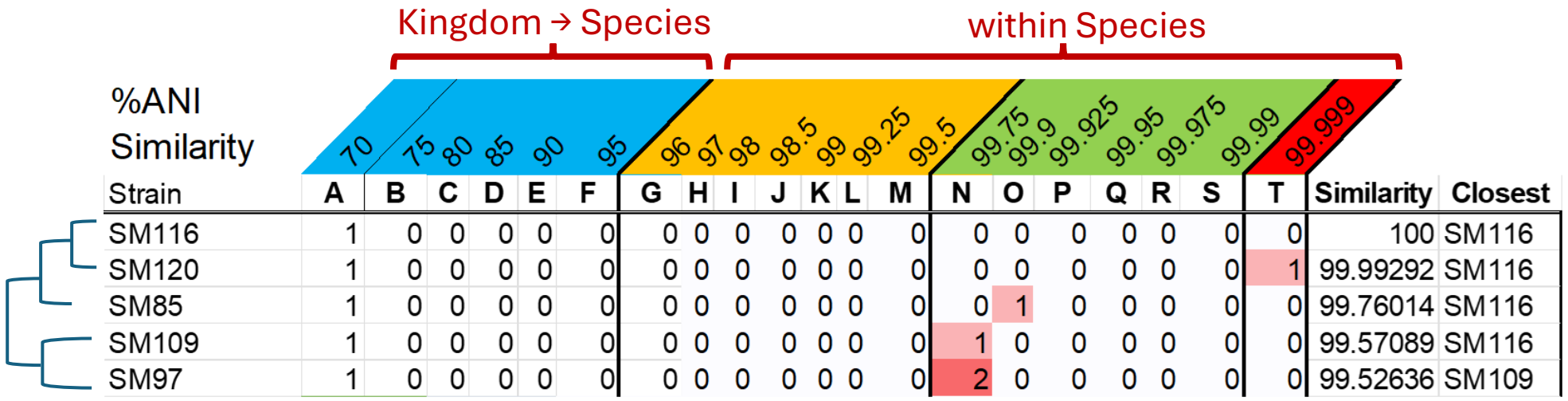

*Ralstonia solanacearum*
Core-genome Tree

# Segregating taxonomies



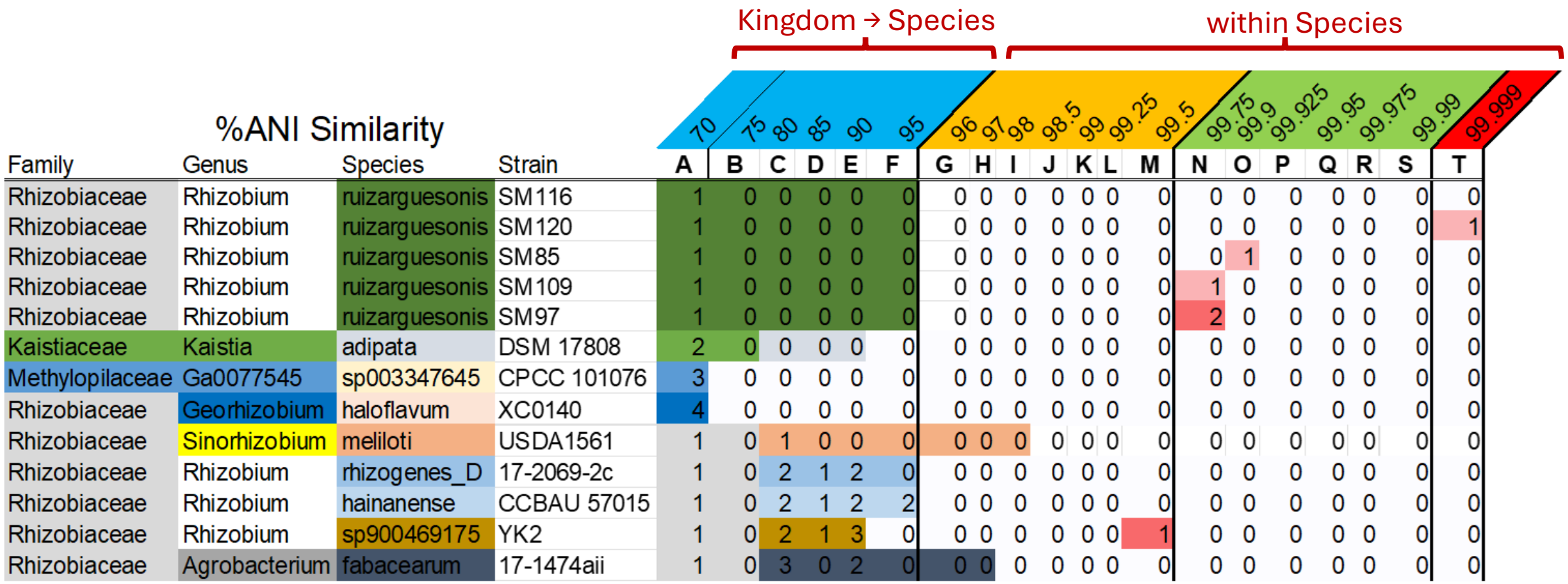
# LINs and Genome Similarity

- Assembled genome input
- Numbers are not values but symbols (class labels)
- LIN assignment is based on the k-nearest neighbors
- The assignment can be dependent on more than one criteria e.g., ANI similarity and alignment coverage




| Strain | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | Similarity | Closest |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SM116 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | SM116 |
| SM120 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 99.99292 | SM116 |
| SM85 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 99.76014 | SM116 |
| SM109 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 99.57089 | SM116 |
| SM97 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 99.52636 | SM109 |

# LINs and Genome Similarity



# Prefixes and LINgroups

- Prefixes are partial LINs
- Genomes starting with the same values as a Prefix are its members
- LINgroups are taxonomy circumsriptions based on LINs
- Each LINgroup describes one or more Prefixes

The Species *Ralstonia Solanacearum* includes all genome LINs starting with “864,0,0,1,[0,1,2]” in Blue

The Genus *Ralstonia* includes all genomes LINs starting with “864,0” in Grey

Kingdom → Species | within Species

%ANI Similarity: 70 | 75 80 85 90 95 | 96 97 98 98.5 99 99.25 99.5 | 99.75 99.9 99.925 99.95 99.975 99.99 | 99.999

| Genus | Species | Phylotype | Strain | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ralstonia | Solanacearum | IIC | K60 | 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Ralstonia | Solanacearum | IIC | Rs5 | 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ralstonia | Solanacearum | IIC | NCPPB325 | 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ralstonia | Solanacearum | IIC | UW700 | 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ralstonia | Solanacearum | IIA | UA-1611 | 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ralstonia | Solanacearum | IIA | UA-1612 | 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Ralstonia | Solanacearum | | GEO_304 | 864 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ralstonia | Solanacearum | | GEO_96 | 864 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Ralstonia | Solanacearum | | CFBP3059 | 864 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ralstonia | Solanacearum | III | CMR15 | 864 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ralstonia | Solanacearum | IV | T98 | 864 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ralstonia | Solanacearum | I | SRS89 | 864 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ralstonia | picketti | | LMG24248 | 864 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

# genomeRxiv Demo
# genomerxiv.cs.vt.edu

Assembly file from sample

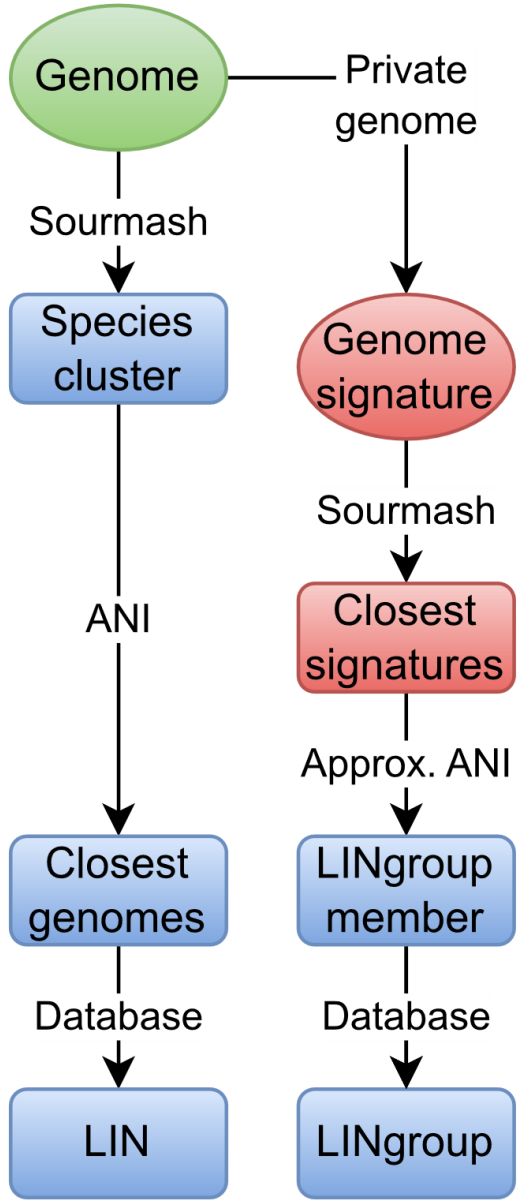
Identify: http://genomerxiv.cs.vt.edu/index.php/result/668ac204c9bd7

Closest genome:
http://genomerxiv.cs.vt.edu/index.php/genome/864,0,0,1,0,0,0,0,0,0,0,0,0,3,0,0,0,0,0,0

Species *Ralstonia Solanacearum* LINgroup:
http://genomerxiv.cs.vt.edu/index.php/lingroup/864,0,0,1,0,0

Different *Ralstonia* LINgroup Prefix: http://genomerxiv.cs.vt.edu/index.php/lingroup/864,0,0,1

*Ralstonia Solanacearum* members: http://genomerxiv.cs.vt.edu/index.php/result/668ac36f2a7bf



Menu

genomeRxiv

Identify using a FASTA file | Identify using an accession number

Accession or assembly number.

[Download example genome] [Use example accession number] [Download workshop assembly]

Identify

The name of the file you upload will be used as the title for your identification job.

Identification can be done with a FASTA file, a publicly available NCBI accession or assembly number, or a Sourmash signature

# Identify ralstonia_example.fna

| | | | |
|---|---|---|---|
| **Job UUID** | 668ac2427581e | **Submit time** | 2024-07-07 12:28:50.481 |
| **Job name** | ident_genome | **Start time** | 2024-07-07 12:28:50.497 |
| **Submitter** | guest | **Terminate time** | 2024-07-07 12:29:06.525 |
| **Status** | success | | |

This genome was already found in our database

## Assigned LIN

LIN

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | GenomeID | Scaffolds | Length |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 1 | 0 | 0 | 0 | 277860 | 69 | 5745432 |

## Closest Genome

99.9431% ANI to Target

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | NCBI organism name | Strain | Type strain | Length | Scaffolds |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | Ralstonia solanacearum | UW700 | not type material | 5487557 | 11 |

## Member LINgroups

2 prefixes

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | Rank | Title | Source |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 864 | 0 | 0 | 1 | 0 | 0 | | | | | | | | | | | | | | | Species | Ralstonia solanacearum | GTDB220 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | | | | phylotype | iic | GENOMERXIV |

# Genome

Submitted by genomeRxiv on 2024-03-06 21:36:58

## LIN

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |

## Isolation Information

| | |
|---|---|
| **BioProject** | PRJNA381968 |
| **BioSample** | SAMN06697628 |
| **Completeness %** | 99.46 |
| **Contamination %** | 1.83 |
| **Country** | USA |
| **Date of isolation** | 2010-01-01 |
| **GPS Coordinates** | 37.59 -75.78 |
| **Host of isolation** | tomato |
| **NCBI Accession Number** | ASM225160v3 |
| **NCBI Assembly** | GCF_002251605.2 |
| **NCBI Taxonomy ID** | 305 |
| **NCBI ftp link** | FASTA |
| **NCBI organism name** | Ralstonia solanacearum |
| **Number of contigs** | 11.0 |
| **Region** | Virginia |
| **Scaffolds** | 11 |
| **Sequence Length** | 5,487,557 |
| **Source of isolation** | agricultural field |
| **Strain** | UW700 |
| **Type strain** | not type material |

## Related LINgroups

2 prefixes

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | Rank | Title | Source |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 864 | 0 | 0 | 1 | 0 | 0 | | | | | | | | | | | | | | | Species | Ralstonia solanacearum | GTDB220 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | | | | phylotype | iic | GENOMERXIV |

# Selected LINgroup

## Described Prefixes

1 prefix

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 76 | 6 | 2 | | | | | | | | | | | | | | | | | |
| 76 | 6 | 5 | | | | | | | | | | | | | | | | | |
| 76 | 6 | 13 | | | | | | | | | | | | | | | | | |

## Properties

| | |
|---|---|
| **Source** | GENOMERXIV |
| **Rank** | Genus |
| **Name** | Pantoea |
| **Described by** | guest |
| **URL** | |

## Taxonomy

| Rank | Taxon |
|---|---|
| Superkingdom | Bacteria |
| Phylum | Pseudomonadota |
| Class | Gammaproteobacteria |
| Order | Enterobacterales |
| Family | Enterobacteriaceae |
| Genus | Pantoea |

## Description

This LINgroup is an ani subgroup of the Genus pantoea.

# Related LINgroups

## Described Overarching Prefixes

no prefixes found

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | Rank | Title | Source |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

## Described Sub-dividing Prefixes

41 prefixes

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | Rank | Title | Source |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 76 | 6 | 2 | 0 | | | | | | | | | | | | | | | | | clade | pantoea agglomerans & astica lineage | GENOMERXIV |
| 76 | 6 | 2 | 0 | 0 | 0 | | | | | | | | | | | | | | | Species | Pantoea alfalfae | GENOMERXIV |
| 76 | 6 | 2 | 0 | 0 | 1 | | | | | | | | | | | | | | | Species | Pantoea vagans | GENOMERXIV |
| 76 | 6 | 2 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | | | | | | | | | | | clade | 1 | GENOMERXIV |
| 76 | 6 | 2 | 0 | 0 | 2 | | | | | | | | | | | | | | | Species | Pantoea gossypiicola | GENOMERXIV |
| 76 | 6 | 2 | 0 | 0 | 3 | | | | | | | | | | | | | | | Species | Pantoea agglomerans | GENOMERXIV |
| 76 | 6 | 2 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | | | | | | | | | | | clade | 1 | GENOMERXIV |
| 76 | 6 | 2 | 0 | 0 | 3 | 0 | 0 | 1 | | | | | | | | | | | | clade | 2 | GENOMERXIV |
| 76 | 6 | 2 | 0 | 0 | 3 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | clade | 3 | GENOMERXIV |

# Selected LINgroup

## Described Prefixes

1 prefix

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 864 | 0 | 0 | 1 | 0 | | | | | | | | | | | | | | | |

## Properties

| | |
|---|---|
| **Source** | GTDB220 |
| **Rank** | Species |
| **Name** | Ralstonia solanacearum |
| **Described by** | guest |
| **URL** | |

## Taxonomy

| Rank | Taxon |
|---|---|
| Superkingdom | Bacteria |
| Phylum | Pseudomonadota |
| Class | Gammaproteobacteria |
| Order | Burkholderiales |
| Family | Burkholderiaceae |
| Genus | Ralstonia |
| Species | Ralstonia solanacearum |

## Description

# Related LINgroups

## Described Overarching Prefixes

no prefixes found

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | Rank | Title | Source |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

## Described Sub-dividing Prefixes

1 prefix

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | Rank | Title | Source |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | | | | phylotype | iic | GENOMERXIV |

View genomes in this LINgroup

# Selected LINgroup

## Described Prefixes

1 prefix

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | | | |

## Properties

| | |
|---|---|
| **Source** | GENOMERXIV |
| **Rank** | phylotype |
| **Name** | iic |
| **Described by** | guest |
| **URL** | |

## Taxonomy

| Rank | Taxon |
|---|---|
| Superkingdom | Bacteria |
| Phylum | Pseudomonadota |
| Class | Gammaproteobacteria |
| Order | Burkholderiales |
| Family | Burkholderiaceae |
| Genus | Ralstonia |
| Species | Ralstonia solanacearum |
| Phylotype | iic |

## Description

This LINgroup corresponds to phylotype iic

# Related LINgroups

## Described Overarching Prefixes

1 prefix

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | Rank | Title | Source |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 864 | 0 | 0 | 1 | 0 | 0 | | | | | | | | | | | | | | | Species | Ralstonia solanacearum | GTDB220 |

## Described Sub-dividing Prefixes

no prefixes found

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | Rank | Title | Source |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

View genomes in this LINgroup

genomeRxiv

Menu

# Genomes in LINgroup: 864,0,0,1,0,0

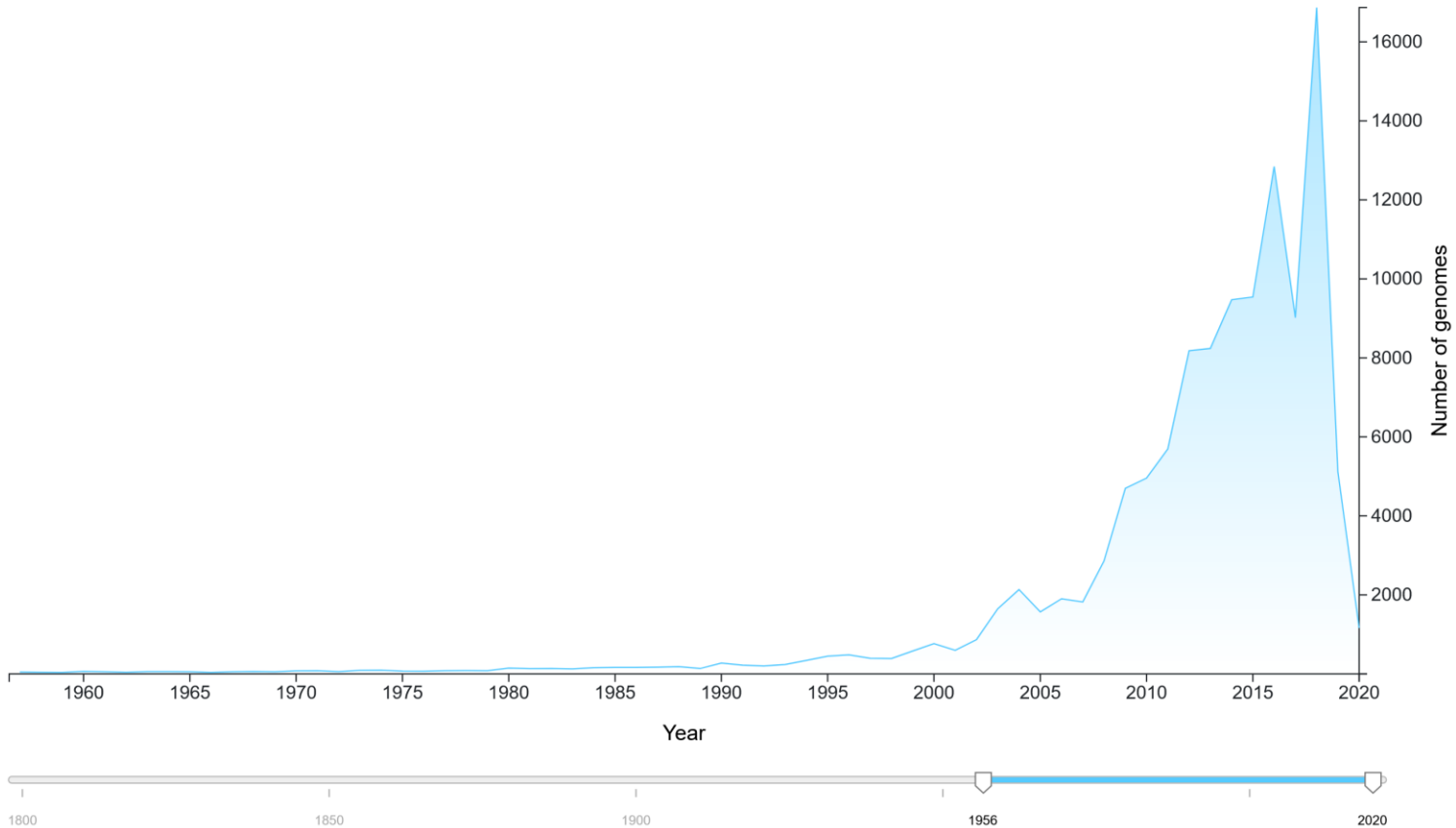
| | | | |
|---|---|---|---|
| **Job UUID** | 668ac36f2a7bf | **Submit time** | 2024-07-07 12:33:51.174 |
| **Job name** | prefix_genome_members | **Start time** | 2024-07-07 12:33:51.218 |
| **Submitter** | guest | **Terminate time** | 2024-07-07 12:33:51.879 |
| **Status** | success | | |

## LINgroup Members

262 genomes found

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | NCBI organism name | Strain | Type strain | Length | Scaffolds |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | Ralstonia solanacearum K60 | K60 | type strain of species | 3675578 | 347 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | Ralstonia solanacearum | Rs5 | not type material | 5430180 | 2 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | Ralstonia solanacearum | NCPPB 325 | type strain of species | 5652708 | 19 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | Ralstonia solanacearum | UW700 | not type material | 5487557 | 11 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 1 | 0 | 0 | 0 | | | | 5745432 | 69 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Ralstonia solanacearum | UA-1611 | not type material | 5195693 | 1 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | Ralstonia solanacearum | UA-1612 | not type material | 5003259 | 1 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Ralstonia solanacearum | CRMRs218 | not type material | 5580300 | 2 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Ralstonia solanacearum | SFC | not type material | 5684563 | 2 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | Ralstonia solanacearum | IBSBF 2570 | not type material | 5693730 | 2 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Ralstonia solanacearum CFBP2957 | CFBP2957 | not type material | 3417386 | 1 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Ralstonia solanacearum | CCRMRs277 | not type material | 5630998 | 2 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | Ralstonia solanacearum | CCRMRs304 | not type material | 5624722 | 2 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Ralstonia solanacearum | RS 489 | not type material | 5411504 | 2 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Ralstonia solanacearum | GEO_304 | not type material | 5117074 | 110 |
| 864 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | | | | 5007920 | 131 |

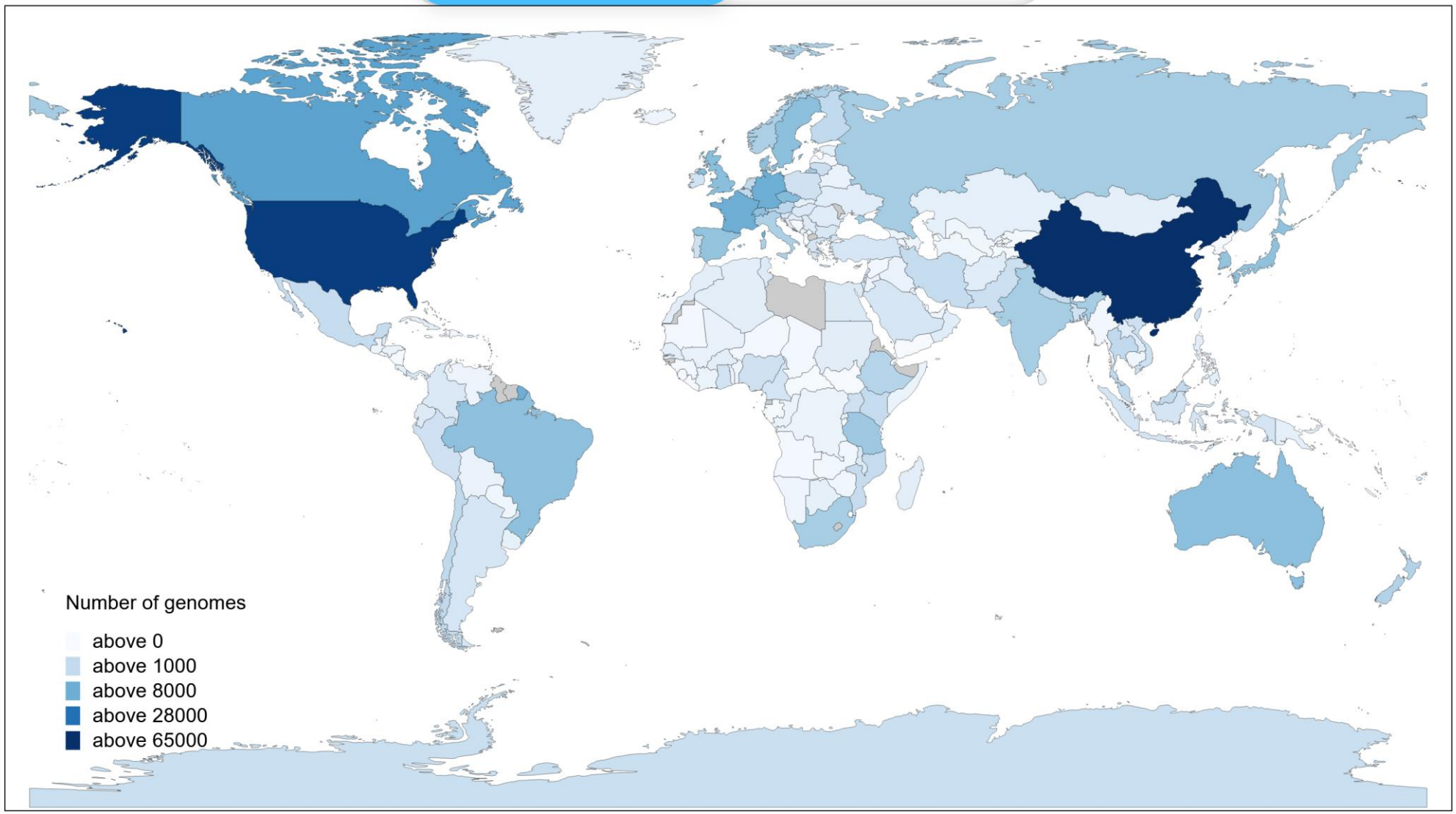
Distribution of Genomes Over Time




The collection dates are retrieved from NCBI Assembly or Biosample database based on the genome AssemblyID.

# Geographic location of genome-sequenced samples in genomeRxiv

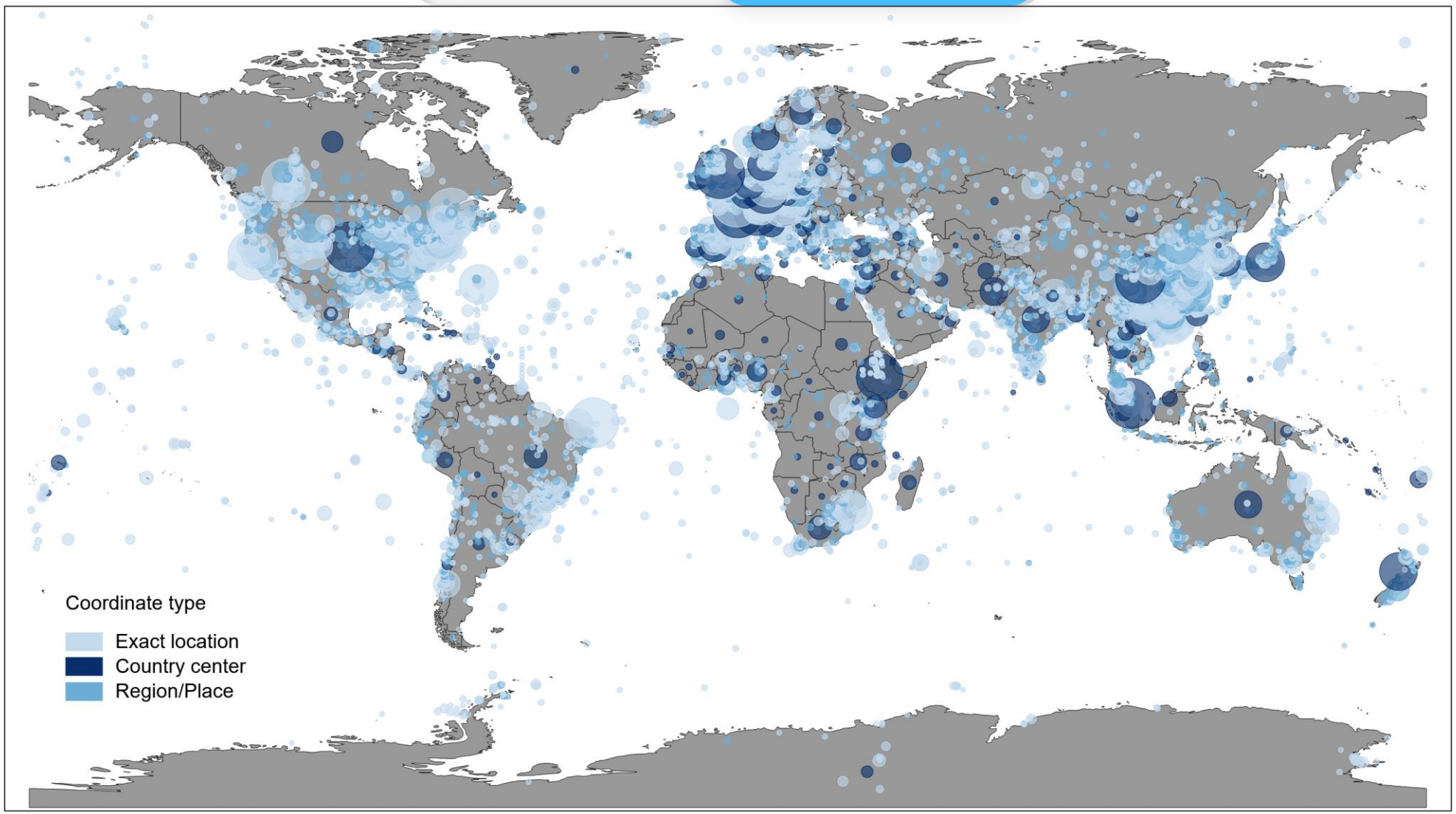
Toggle to change | HEAT MAP




The location are retrieved from NCBI Assembly or Biosample database based on the genome AssemblyID.

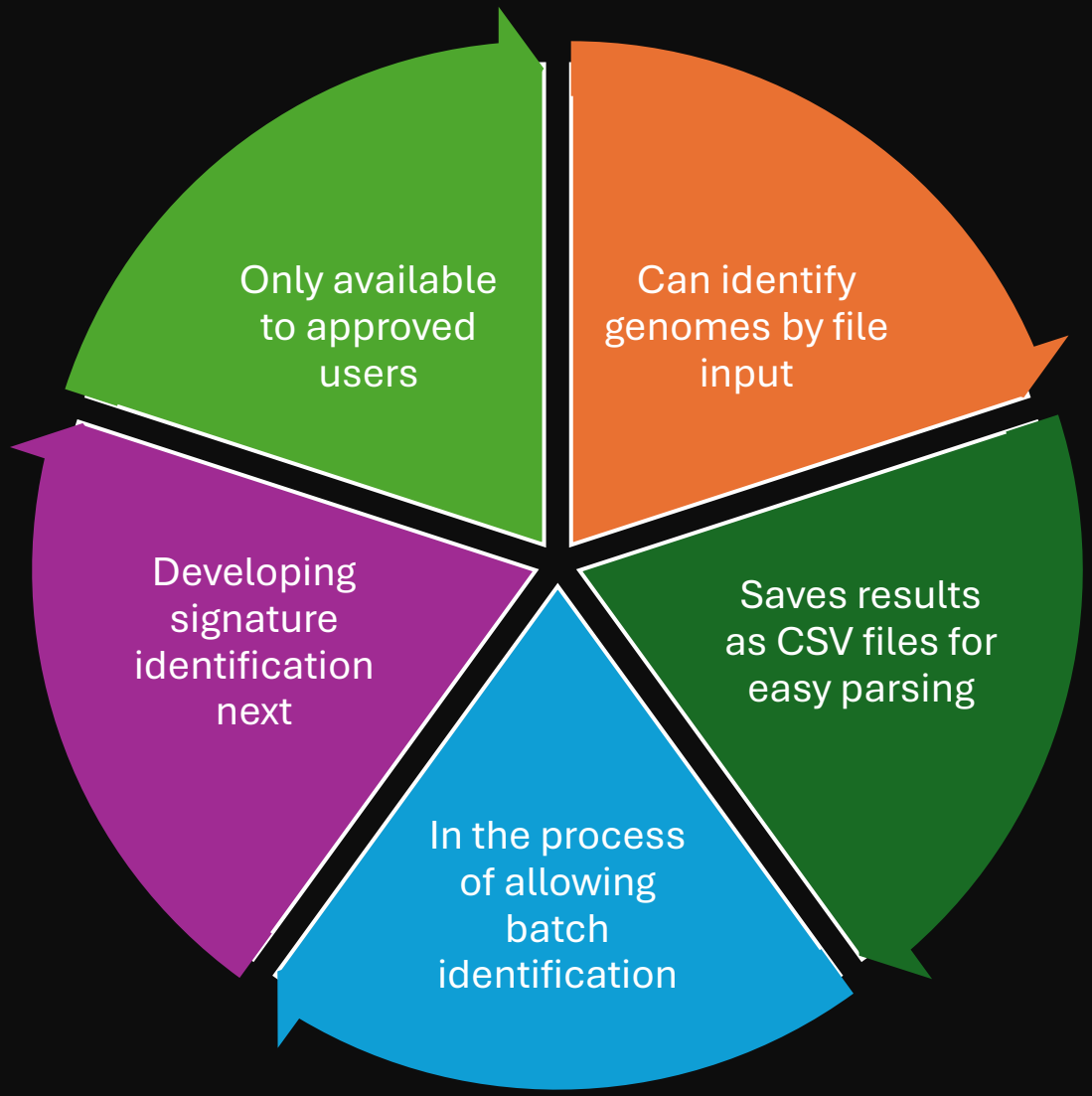
# Geographic location of genome-sequenced samples in genomeRxiv

BUBBLE PLOT | Toggle to change




The location are retrieved from NCBI Assembly or Biosample database based on the genome AssemblyID.

# BrookLIN genomeRxiv python API client



# Running the Client

./brooklin.py genome-identify
-u auth.conf
-f ralstonia_example.fna
-o ralstonia

./brooklin_batch.py batch genome-identify
-u auth.conf
-f ralstonia_genome_files_paths.csv
-o ralstonia_genomes

>see results at
http://genomerxiv.cs.vt.edu/index.php/result/668ac43e17122

ralstonia _Tentative_LIN.csv

| Input File | Tentative ID | Tentative LIN | Scaffold Count | Sequence Length | ANI Similarity |
|---|---|---|---|---|---|
| ralstonia_example.fna | 277860 | 864, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 0, 1, 0, 0, 0 | 69 | 5745432 | 99.9431 |

ralstonia _Closest_LIN.csv

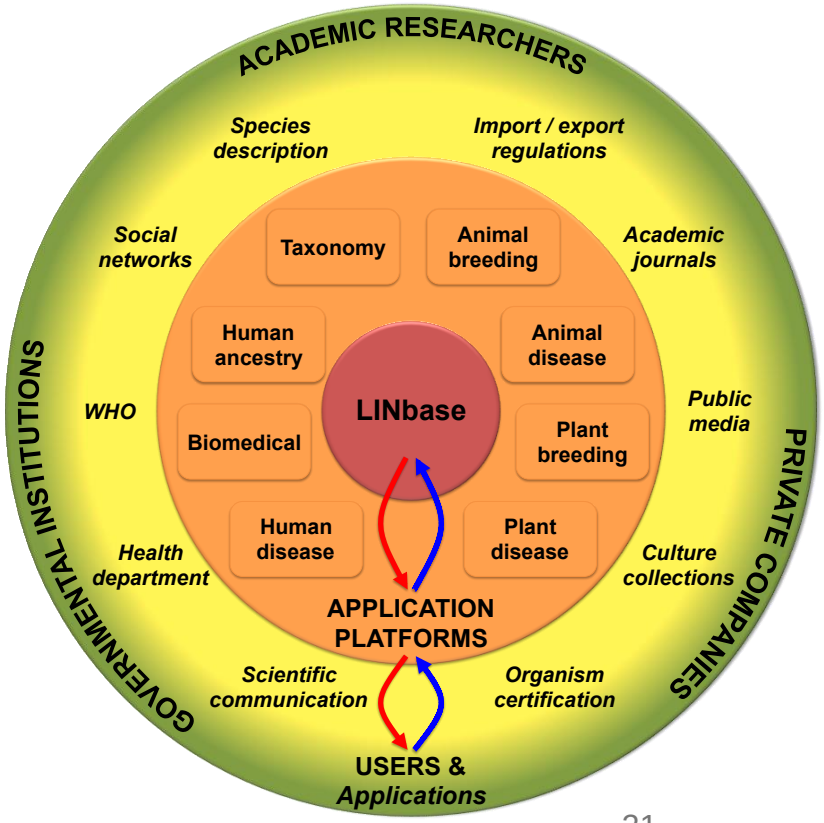
| Input File | Neighbor ID | Closest LIN | Organism Name | Strain Name | SequenceLength |
|---|---|---|---|---|---|
| ralstonia_example.fna | 5923 | 864, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 0, 0 | Ralstonia solanacearum | UW700 | 5487557 |

ralstonia _LINgroups.csv

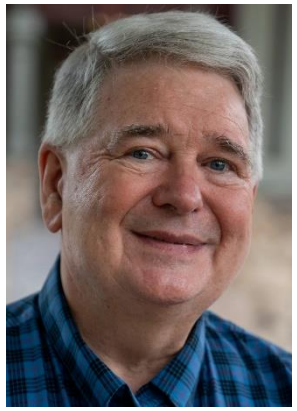
| Input File | LIN | Organism Name | Source | Rank Name |
|---|---|---|---|---|
| ralstonia_example.fna | 864,0 | Ralstonia | GENOMERXIV | Genus |
| ralstonia_example.fna | 864,0,0,1,0,0 | Ralstonia solanacearum | GENOMERXIV | Species |
| ralstonia_example.fna | 864,0,0,1,0,0,0,0,0,0,0,0,0 | iic | GENOMERXIV | phylotype |

# How to assign LINs to your genomes

- **GenomeRxiv** is our web-service that clusters and categorizes most known Bacteria and Archaea using hashed genomes http://genomerxiv.cs.vt.edu [https://www.nsf.gov/awardsearch/showAward?AWD_ID=2018522]
- **LINflow** is our flexible python implementation of LINs to generate unique codes and evolutionary trees https://code.vt.edu/linbaseproject/LINflow [https://doi.org/10.7717/peerj.10906]
- **BrookLIN** is a command line client for genomeRxiv to automate the identification process https://code.vt.edu/linbaseproject/linbase_api
- Our GitLab repository is available at: https://code.vt.edu/linbaseproject

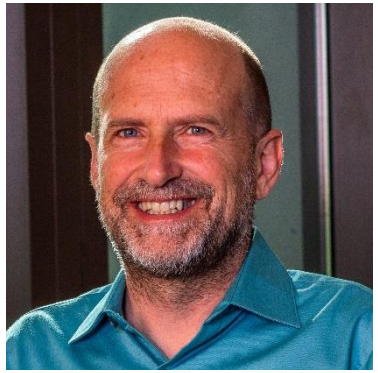


# The LINproject team


Lenwood S. Heath


Sehgeet Kaur


Kassaye Belay


Boris A. Vinatzer


Parul Sharma


Mitchell Gercken

Special thanks to all our collaborators

**C. Titus Brown**, Tessa P. Ward, Mohamed Abuelanin Hussein, & Luiz Irber (UC Davis)

**Leighton Pritchard**, Angelika Kiepas, & Bailey Harrington (Univ. of Strathclyde UK)

Tiffany Lowe-Power (UC Davis)

Caitilyn Allen (University of Wisconsin-Madison)

Long Tian, Marcela A. Johnson, & Song Li (Virginia Tech)

Alexandra J. Weisberg & Jeff H. Chang (Oregon State Univ.)

Undergraduate researchers from Virginia Tech:

Abhilash Chauhan, Aman Kothari, Atul Bharadwaj, Chandra Sekhar Nerella, Matthew Russell, Pradyumann Singhal, Rahul Ramarao, Rituraj Sharma, & Suvasish Pant