

A phylogeny-based taxonomic characterization of the human gut microbiome from shotgun sequence data

Analysis of HMP 454 WGS data with PhylOTU

Rebecca M. Lamb



**GLADSTONE INSTITUTE OF
CARDIOVASCULAR DISEASE**

University of California, San Francisco

HMP 454 WGS data

- Whole Genome Sequencing (WGS)
- Gut samples
- 14 individuals
- 37M sequences

PhylOTU Update

- Code more efficient and parallelized since publication
- Latest version available on github.com
- 24 hours of run time (60 parallel jobs)
- 59K 16S reads classified into OTUs
- 1775 OTUs @ equivalent 97% PID

Silva reference tree

Proteobacteria

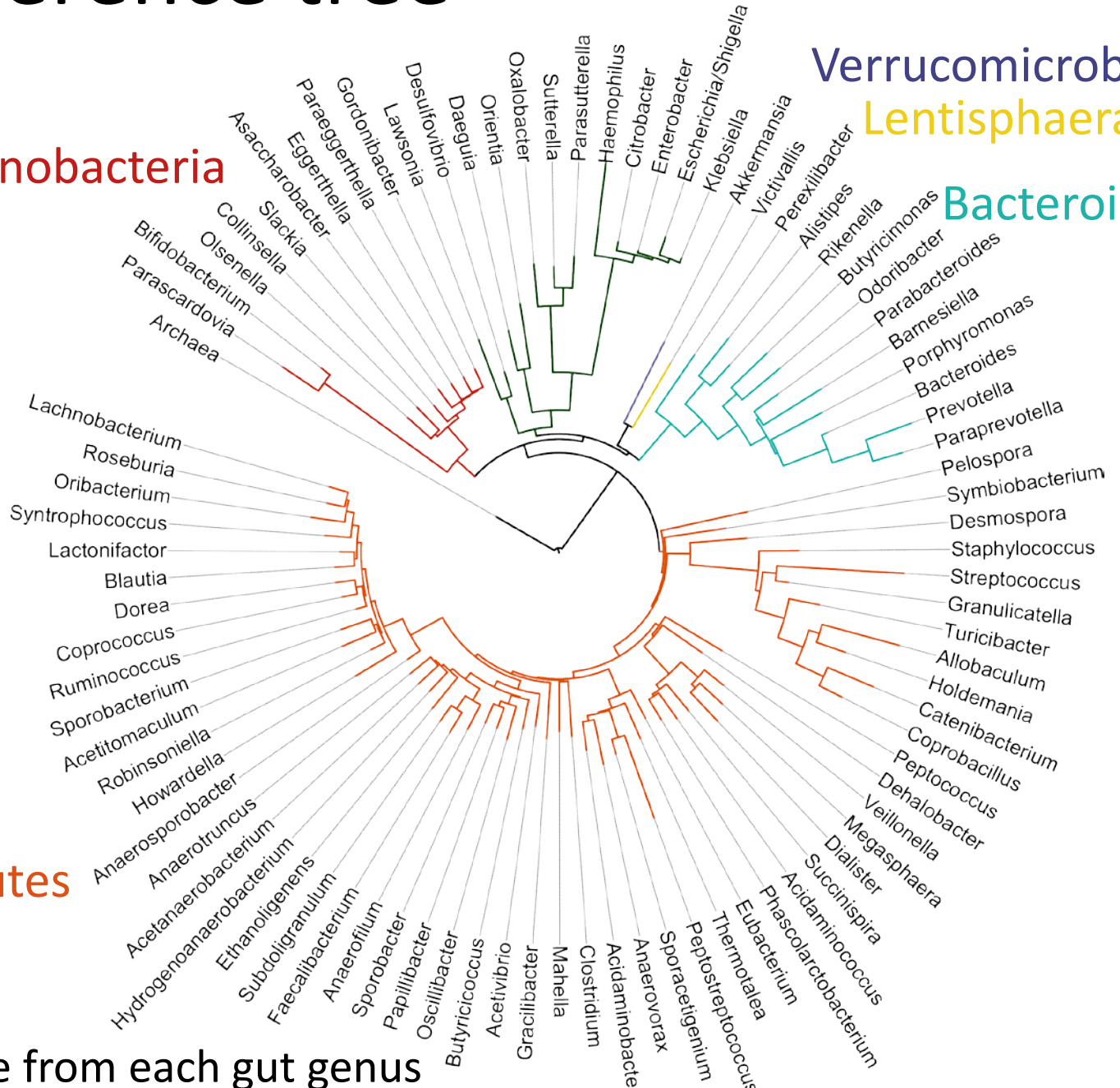
Verrucomicrobia

Lentisphaerae

Bacteroidetes

Actinobacteria

Firmicutes

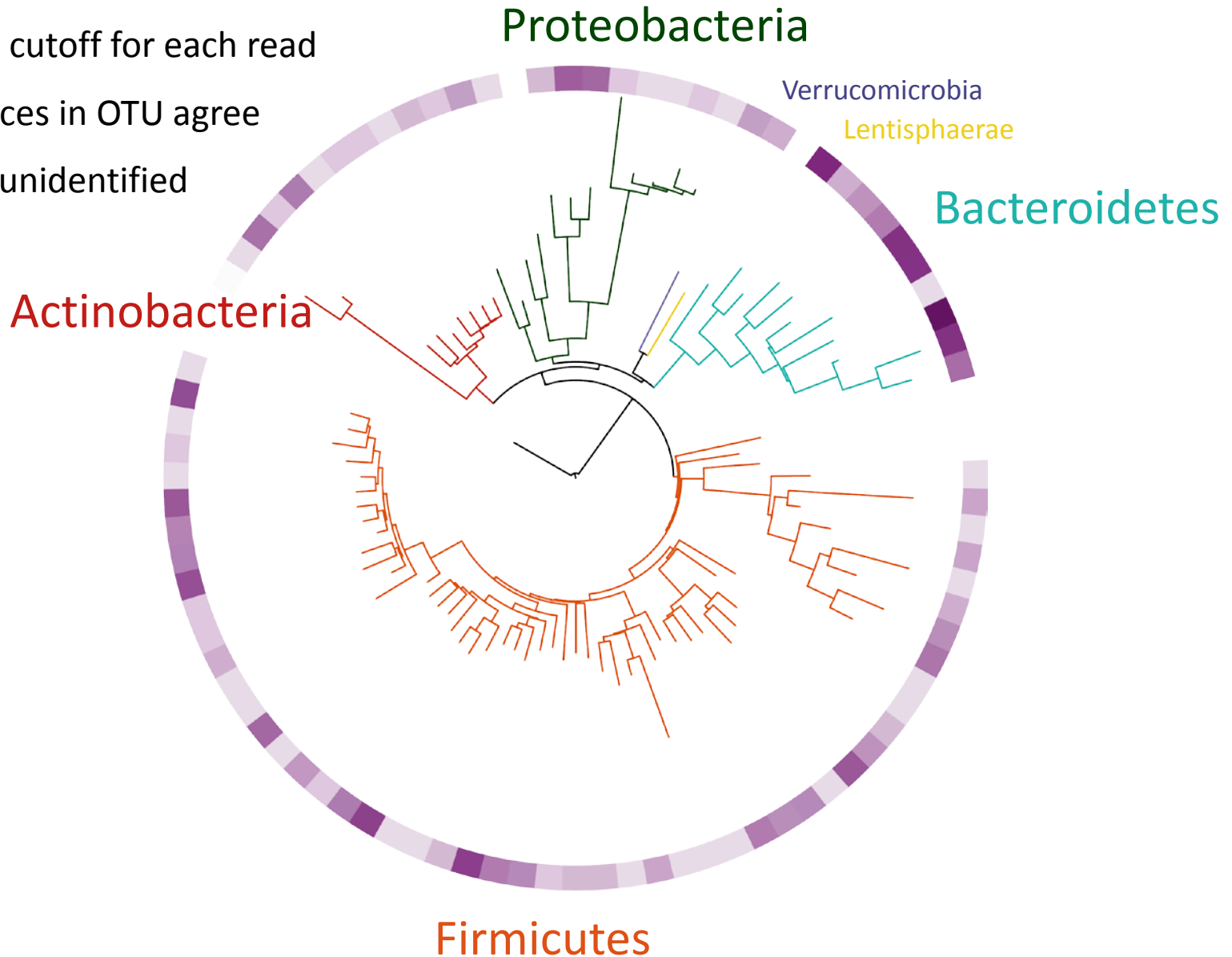


1 representative from each gut genus

Number of OTUs

RDP classification

- 0.5 bootstrap cutoff for each read
- >50% sequences in OTU agree
- 40% of OTUs unidentified



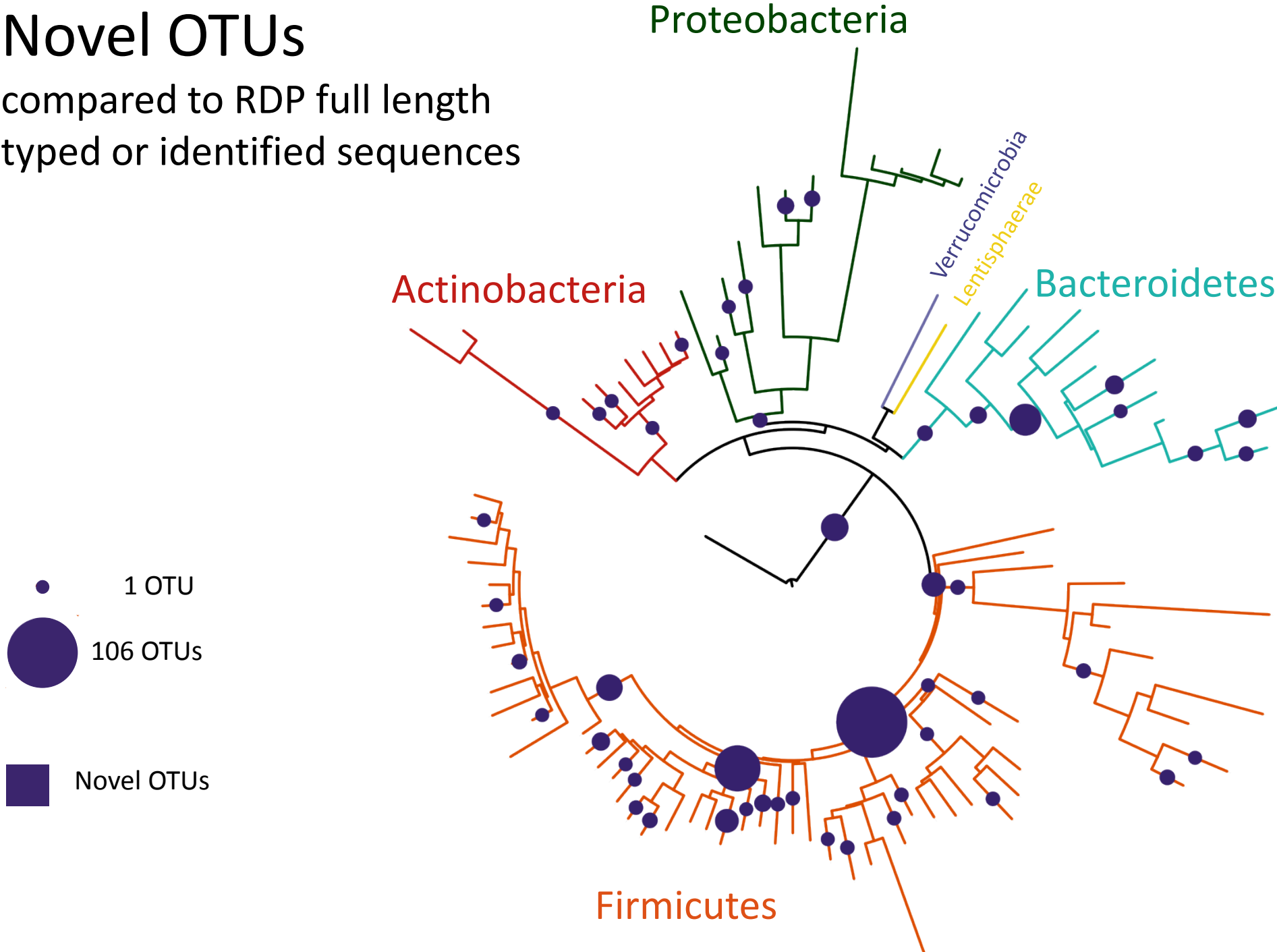
Numbers of novel OTUs

	Definition	Full length typed or identified RDP DB	Full RDP DB
Novel OTUs	No RDP match @ 95% coverage, 95% ID to any read in the OTU	410	49
Novel + Illumina Support	Illumina read match to 454 read 80bp @ 99% ID, 0 gaps	230	23
Novel + Core	454 or Illumina reads present in ≥7 (50%) individuals	37	2

- **3.5M** 16S reads from Illumina WGS of 12 of the same individuals

Novel OTUs

compared to RDP full length
typed or identified sequences

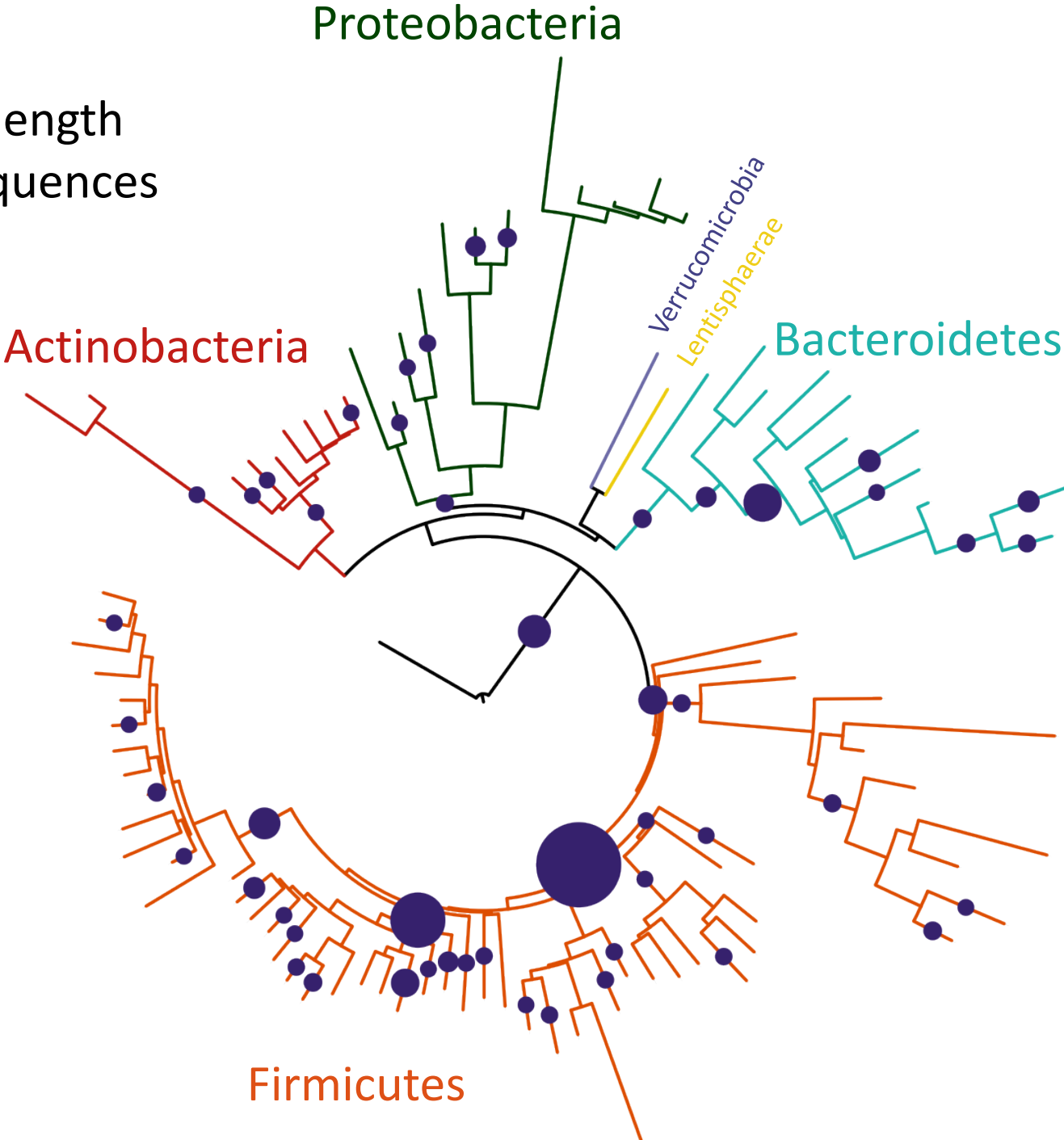


Novel OTUs

compared to RDP full length typed or identified sequences

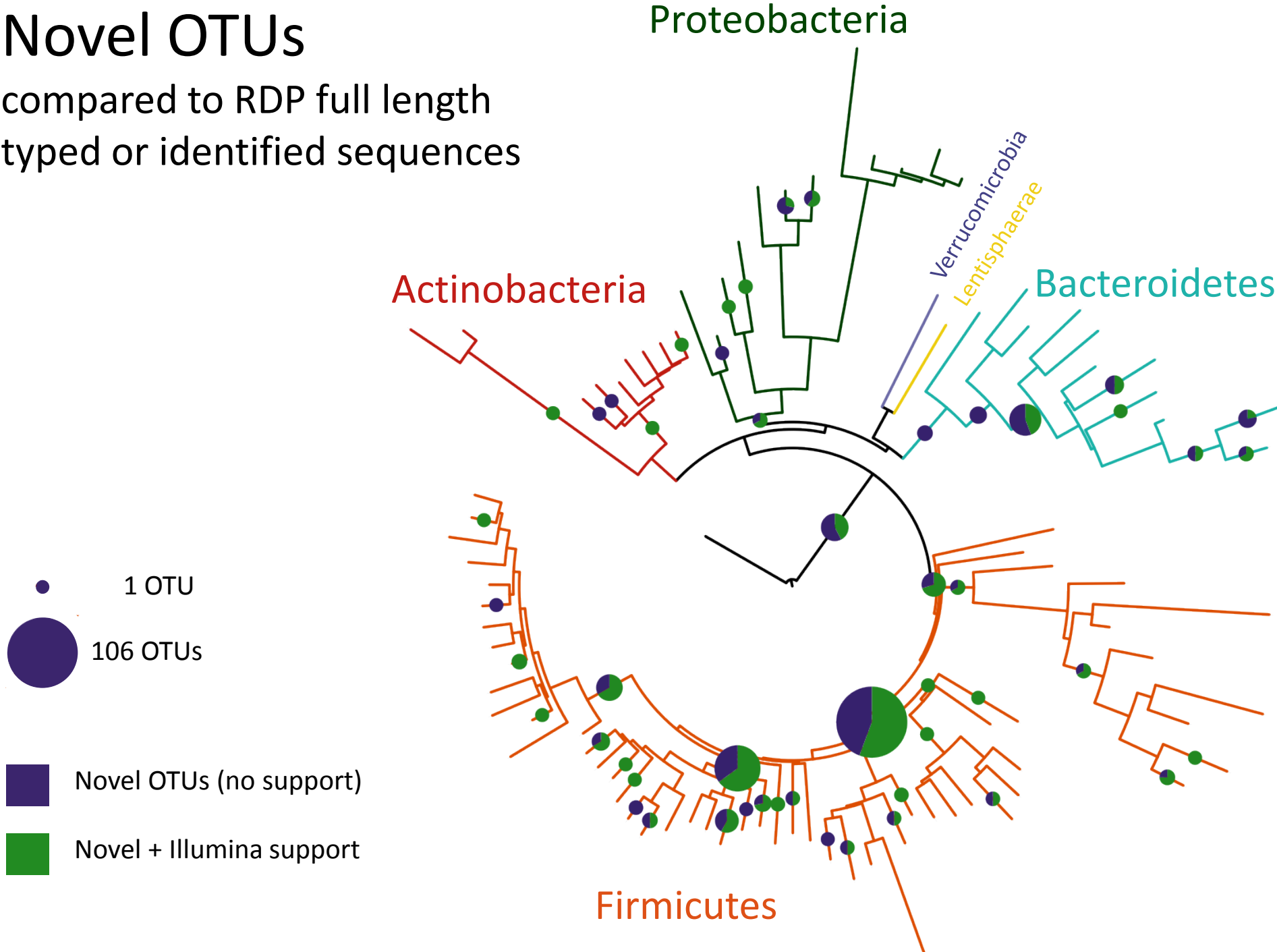
Novel OTUs are distributed across the tree

- 1 OTU
- 106 OTUs
- Novel OTUs



Novel OTUs

compared to RDP full length typed or identified sequences

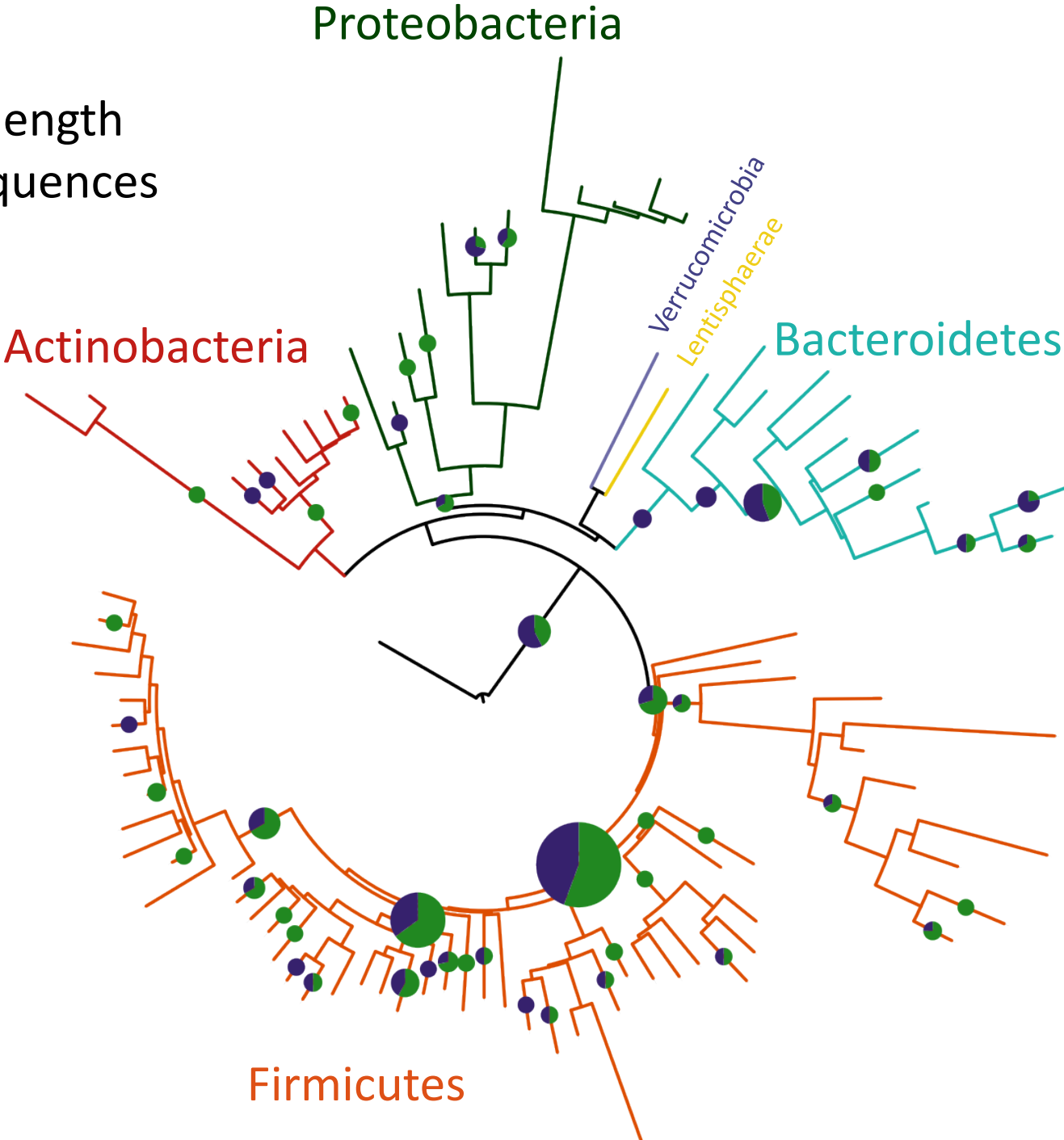


Novel OTUs

compared to RDP full length typed or identified sequences

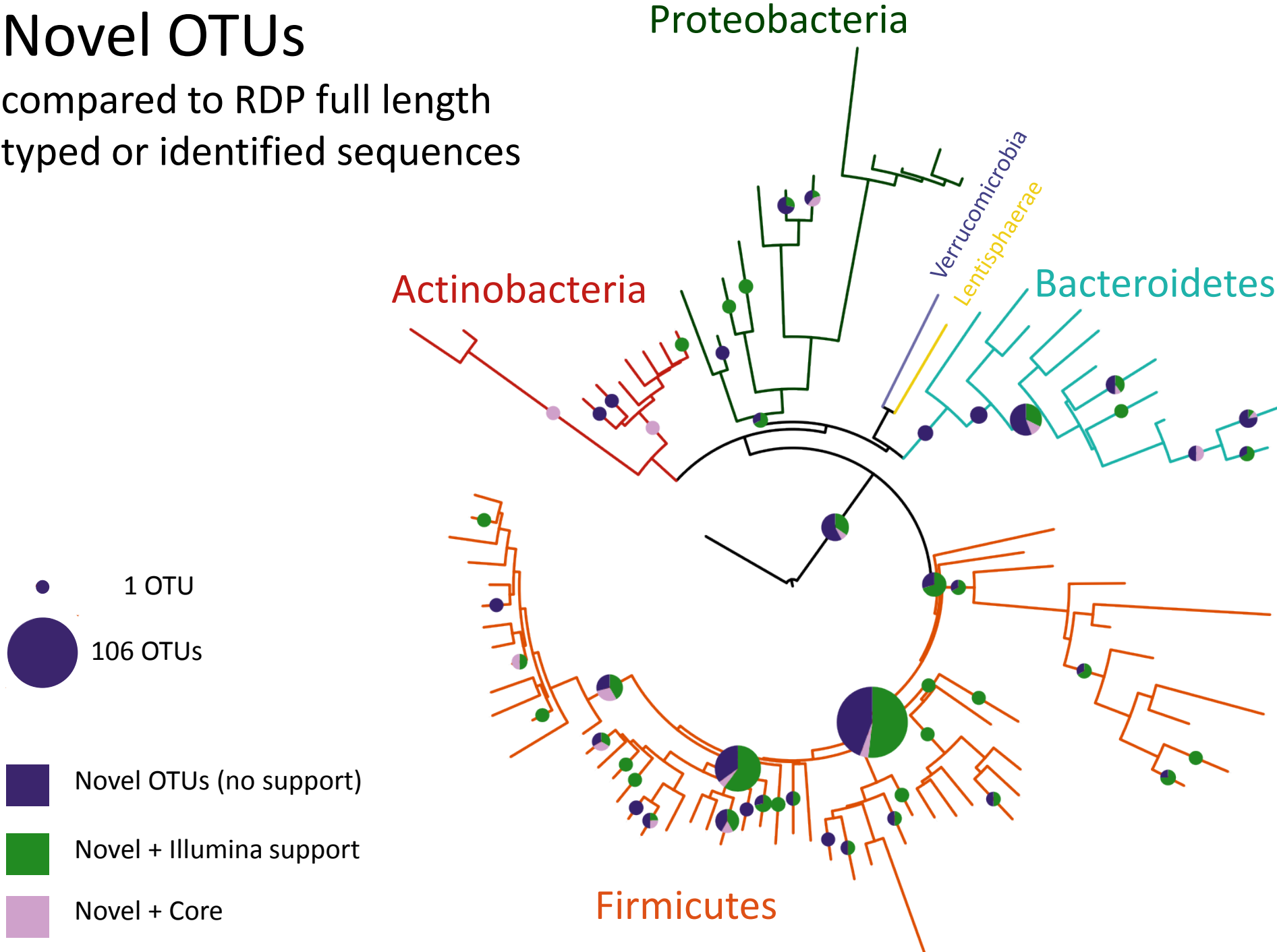
High confidence novel OTUs are distributed across the tree

- 1 OTU
- 106 OTUs
- Novel OTUs (no support)
- Novel + Illumina support



Novel OTUs

compared to RDP full length typed or identified sequences

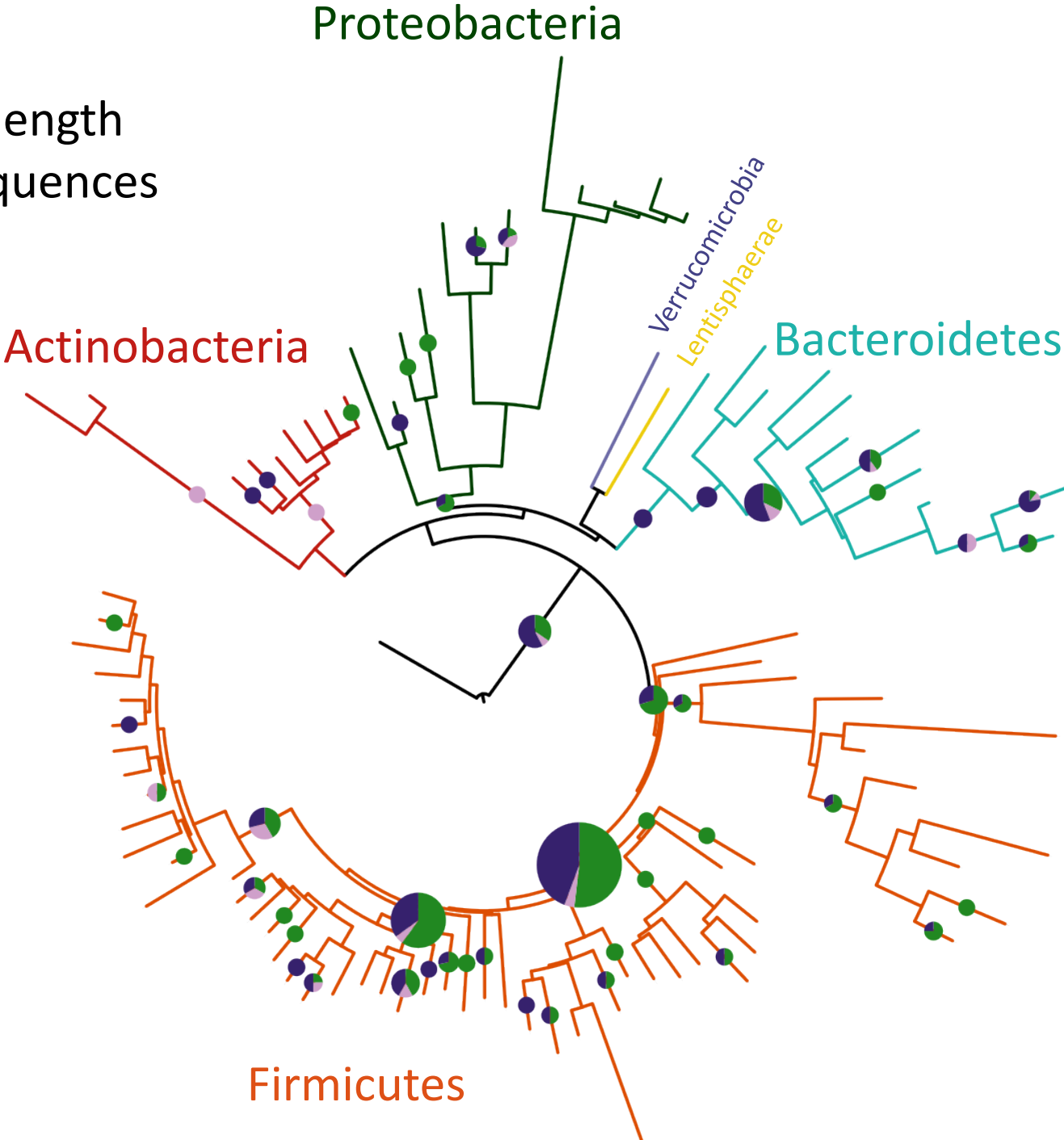


Novel OTUs

compared to RDP full length typed or identified sequences

Some core OTUs are very poorly identified

- 1 OTU
- 106 OTUs
- Novel OTUs (no support)
- Novel + Illumina support
- Novel + Core

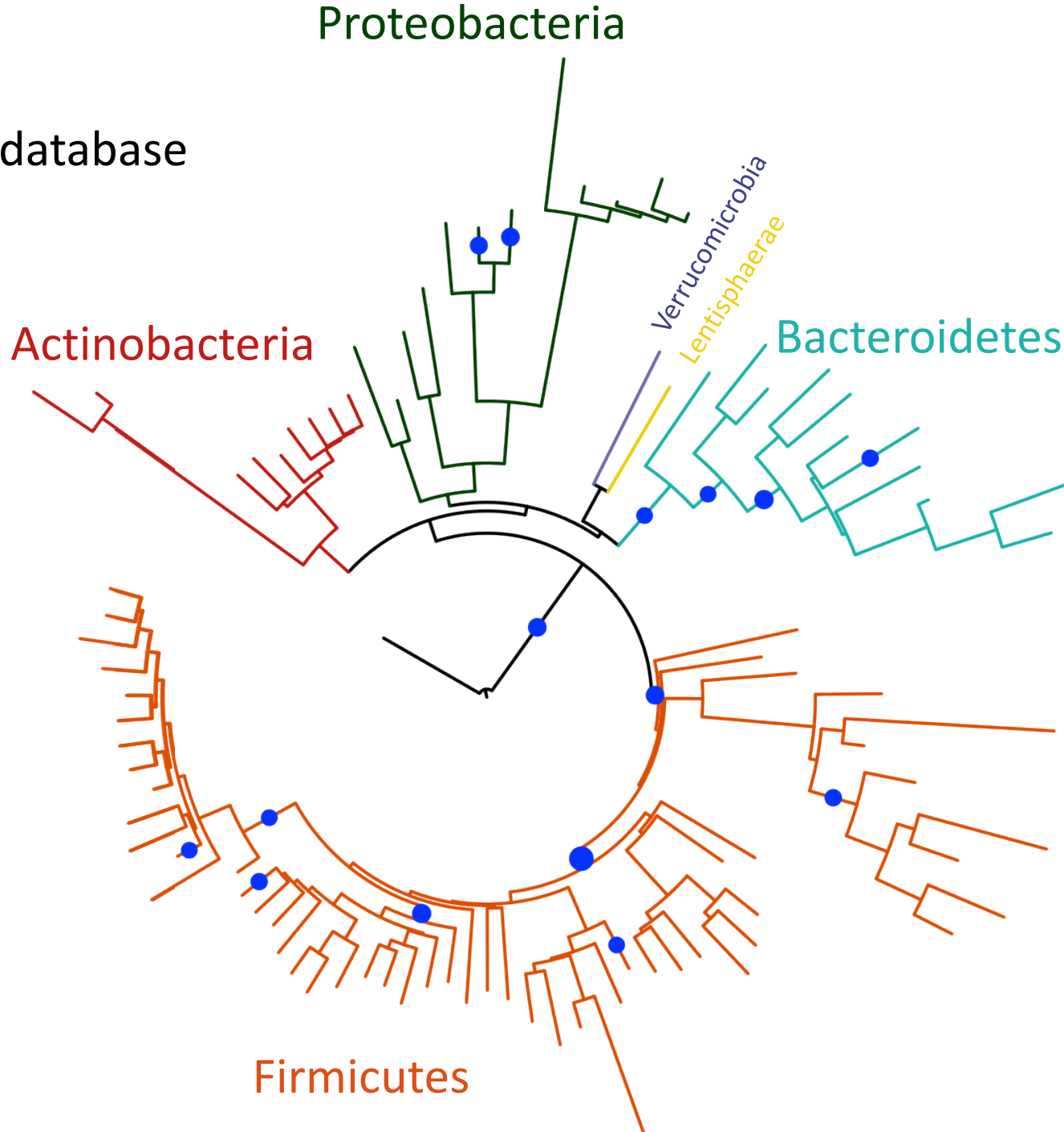


Novel OTUs

compared to full RDP database

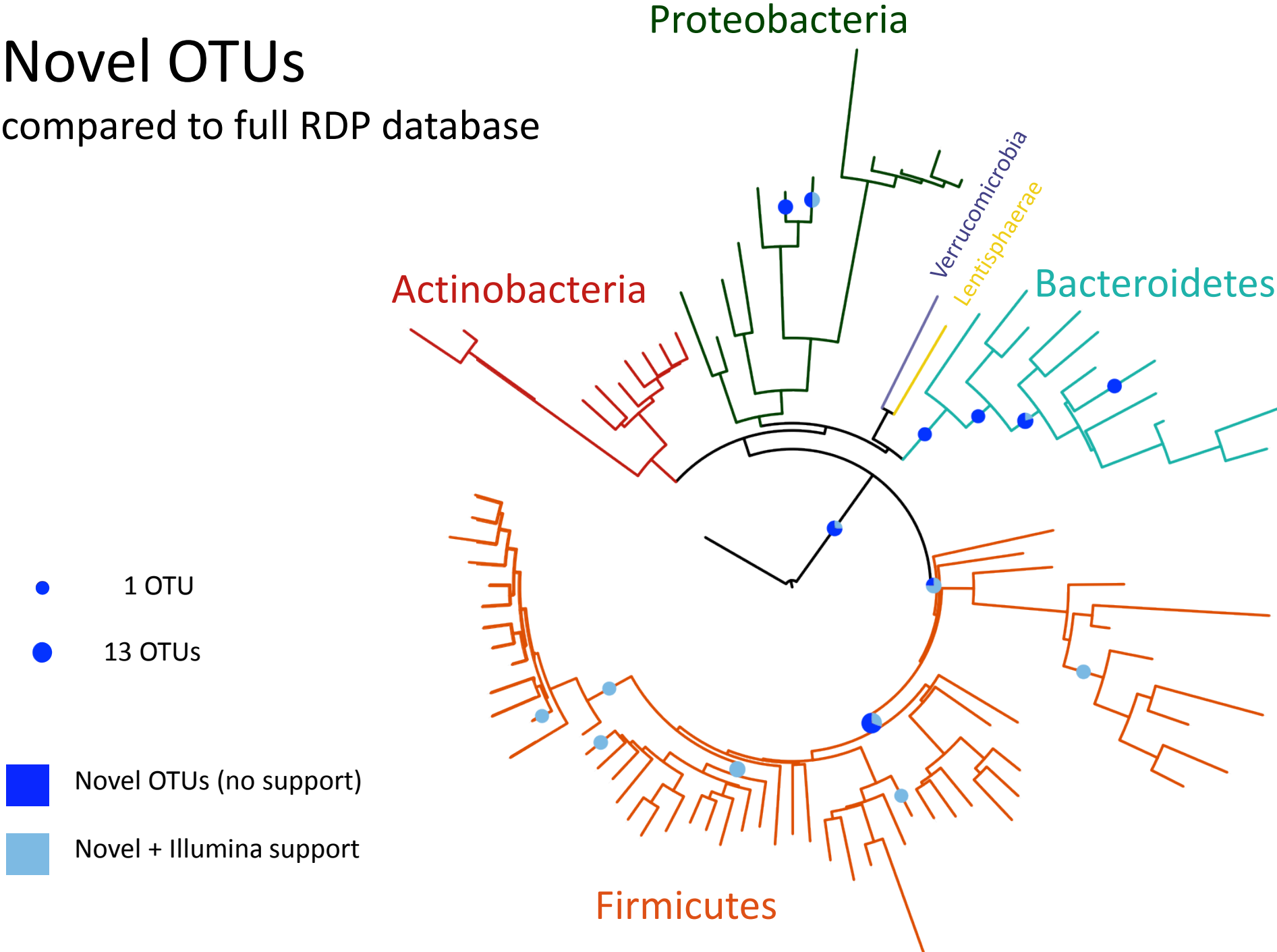
Strictly defined novel OTUs are relatively phylogenetically restricted

- 1 OTU
- 13 OTUs
- Novel OTUs



Novel OTUs

compared to full RDP database



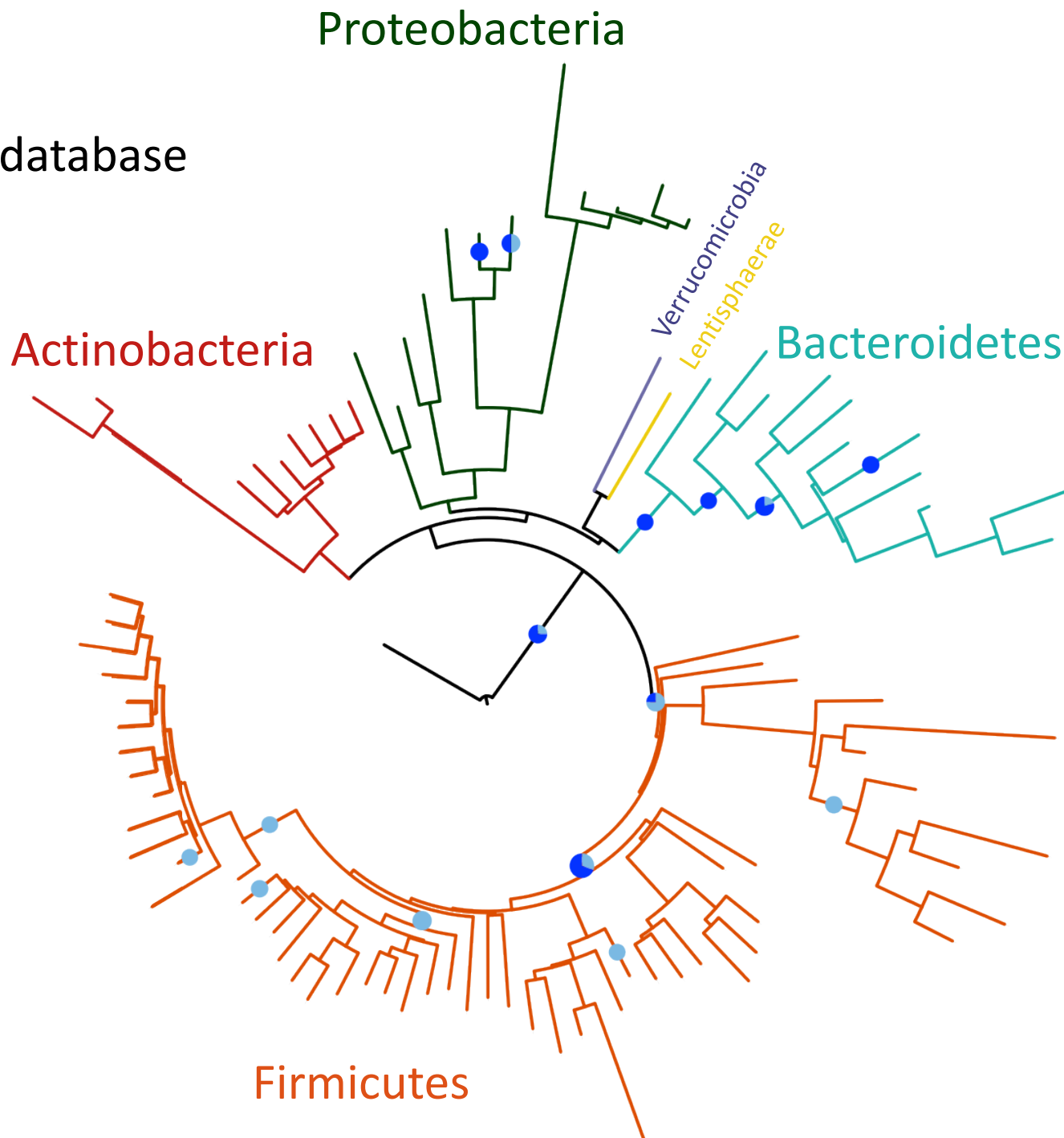
Novel OTUs

compared to full RDP database

High confidence novel OTUs in Firmicutes, Proteobacteria, Bacteroidetes

- 1 OTU
- 13 OTUs

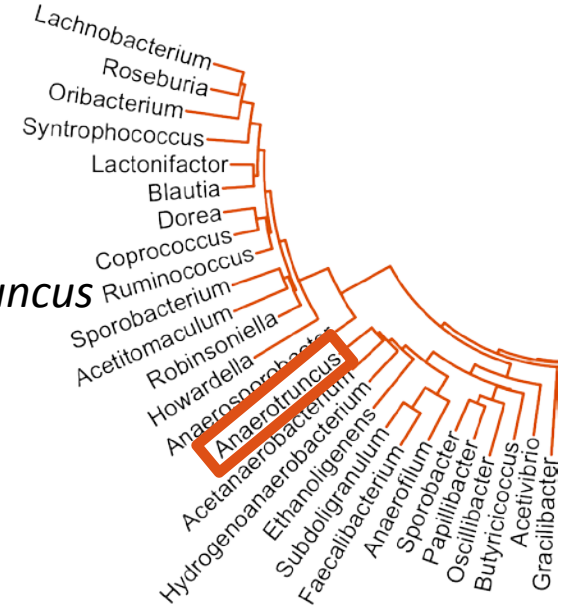
- Novel OTUs (no support)
- Novel + Illumina support



Novel OTUs

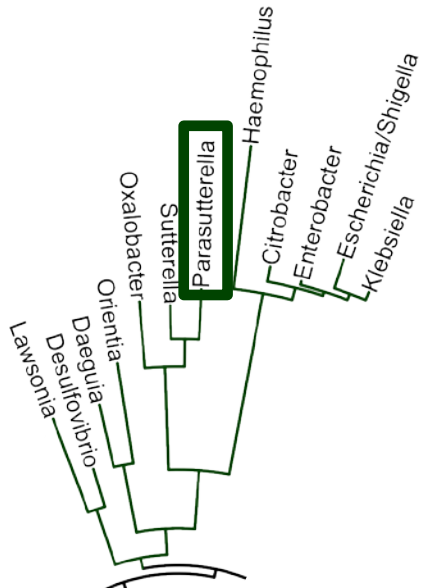
compared to full RDP database

Firmicutes – Clostridia - Clostridiales – Ruminococcaceae – *Anaerotruncus*



- **1** 454 Read, 0.65 RDP bootstrap
- **20** Illumina reads from **7** individuals
- *Anaerotruncus* (2004)
 - 21 other OTUs identified as this genus
 - Gut associated
 - Associated with bacteraemia

Proteobacteria – Betaproteobacteria – Burkholderiales – Alcaligenaceae – *Parasutterella*



- **1** 454 Read, 1.0 RDP bootstrap
- **243** Illumina reads from **7** individuals
- *Parasutterella* (2009)
 - 21 other OTUs identified as this genus
 - Gut associated

Conclusions

- Identified gut microbiome diversity below the genus level
- Novel OTUs identified
 - ~50% supported with Illumina reads
 - ~9% are present in multiple individuals
- Come visit my poster

Acknowledgements



**GLADSTONE INSTITUTE OF
CARDIOVASCULAR DISEASE**

Katherine S. Pollard
Thomas J. Sharpton



**HUMAN
MICROBIOME
PROJECT**

THE
Genome
CENTER
AT WASHINGTON UNIVERSITY

Makedonka Mitreva
George Weinstock
Erica Sodergren
Hongyu Gao



Anthony Fodor



Backup

Number of OTUs

RDP classification

- 0.5 bootstrap cutoff for each read
- >50% sequences in OTU agree
- 40% of OTUs unidentified (inner ring)

