

# **NFDI4Microbiota** Use Case GUT – "Crowd-sourcing high-quality descriptions of novel taxa" by Dr. Thomas Hitch





## How many prokaryotic species are there?

Parks et al. 2021. Nuc. Acid Res.: GTDB Release 202

	Bacteria	Archaea	Total
Phylum	127	19	146
Class	360	47	407
Order	1,163	116	1,279
Family	2,886	336	3,222
Genus	12,037	851	12,888
Species	45,555	2,339	47,894





Exclusively inclusion and/or and/or and/or Exclusively isolates Isolate and environmental genomes



#### Cultured fraction of the human gut (MAGs)

3





# Protologger: Automated analysis and comprehensive description

7 lines of taxonomic evidence for accurate placement
4 pieces of functional information for description of predicted physiology and niche occupation
2 lines of ecological information to study prevalence across ecosystems



- >3,000 jobs run on the web-server
- Paper downloaded >2,000 times
- Second most cited paper in ISME comms for 2021



#### Tindall. 1999 IJSB

#### The introduction of a 'protologue'

Descriptions of new taxa appear in a wide range of publications, and the requirements of the journals are diverse. This has led to various problems with locating information relating to the description of the new taxon. In some cases the name of the taxon may be included in the text, without clear reference to its properties. It is therefore, suggested that all descriptions of new names or combinations should adhere to a uniform format (c.f. the BioCode; Greuter et al., 1996). This format should be adopted for the valid publication of all names (whether as original articles, or by announcement in the 'Validation Lists' and reference to a previous effective publication) appearing in the IJSB starting on 1 January 2000 (or a suitable date thereafter). This would be included as part of Rule 27(2):

'(1) The new name or new combination should be clearly stated and indicated as such (i.e. fam.nov., gen. nov., sp. nov., comb. nov., etc.).

(2) The derivation of a new name must be given.

(3) The properties of the taxon being described must be given directly after (1) and (2). This may include reference to tables or figures in the same publication, or reference to previously effectively published work.

(4) All information contained in (3) must be accessible.

(5) The type of the taxon must be designated. In the case of species or subspecies the culture collection number(s) where a subculture of the type strain has been deposited must be indicated.'

## Gilroy et al. 2021 PeerJ

Table 1 Protologues for new Candidatus taxa identified from metagenomic analysis of chicken gut samples.

Description of Candidatus Acetatifactor stercoripullorum sp. nov.

✓ Define novelty ✓ Name taxon

Candidatus Acetatifactor stercoripullorum (ster.co.ri.pul.lo'rum. L. neut. n. stercus dung; L. masc. n. pullus a young chicken; N.L. gen. n. stercoripullorum of the faceces of young chickens)

#### **Description**?

A bacterial species identified by metagenomic analyses. This species includes all bacteria with genomes that show  $\geq$ 95% average nucleotide identity (ANI) to the type genome, which has been assigned the MAG ID CHK195-6426 and which is available via NCBI BioSample SAMN15816622. The GC content of the type genome is 48.46% and the genome length is 3.1 Mbp.

✓ State the type material

### A call to arms for systematists: revitalising the purpose and practises underpinning the description of novel microbial taxa

Iain C. Sutcliffe · Martha E. Trujillo · Michael Goodfellow

- HBC; human gut bacterial collection from Forster *et al* (2019)
- From the HBC collection 64 isolates were deposited at the DSMZ
- 40 were identified to represent
   34 novel taxa
- 17 novel species and 17 novel genera
- Includes; 3 high priority, 13 medium priority and 6 low priority species according to HMP





## **Remaining isolate collections**



Poyet et al. 2019 Nat. Biotec.; Forster et al. 2019 Nat. Biotec.; Seshadri et al. 2018 Nat. Biotec.

## The case for Candidatus taxa



Exclusively MAGs and/or SAGs
 Exclusively isolates
 Isolate and environmental genomes

# Extensive microbial diversity within the chicken gut microbiome revealed by metagenomics and culture

Rachel Gilroy<sup>1</sup>, Anuradha Ravi<sup>1</sup>, Maria Getino<sup>2</sup>, Isabella Pursley<sup>2</sup>, Daniel L. Horton<sup>2</sup>, Nabil-Fareed Alikhan<sup>1</sup>, Dave Baker<sup>1</sup>, Karim Gharbi<sup>3</sup>, Neil Hall<sup>3,4</sup>, Mick Watson<sup>5</sup>, Evelien M. Adriaenssens<sup>1</sup>, Ebenezer Foster-Nyarko<sup>1</sup>, Sheikh Jarju<sup>6</sup>, Arss Secka<sup>7</sup>, Martin Antonio<sup>6</sup>, Aharon Oren<sup>8</sup>, Roy R. Chaudhuri<sup>9</sup>, Roberto La Ragione<sup>2</sup>, Falk Hildebrand<sup>1,3</sup> and Mark J. Pallen<sup>1,2,4</sup>

## 150 Candidatus genus names.

650 distinctive binomials for new Candidatus species

## Lists of names of prokaryotic *Candidatus* taxa

Aharon Oren<sup>1,\*</sup>, George M. Garrity<sup>2,3</sup>, Charles T. Parker<sup>3</sup>, Maria Chuvochina<sup>4</sup> and Martha E. Trujillo<sup>5</sup>

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

The List of Prokaryotic names with Standing in Nomenclature (LPSN) currently contains 2,431 "Candidatus" taxa



9

Current issues:

- There are thousands of microbes still to be described
- There is a **lack of standardisation** regarding what is useful
- If this isnt addressed soon a high-volume of low-quality *Candidatus* protologues will be produced



Aims of ProtoBIOME:

- To provide high-quality descriptions of these novel taxa
- Help set standards in taxonomic description
- Spread these standards via community engagement



## 1. Select an environment to study

- Experts in the field are needed to ensure relavent descriptions
- Individuals who know these taxa to ensure accurate taxonomic placement

#### 2. Genomic database

- Identify a collection of isolates or MAGs for the selected environment
- Extract or sequence 16S rRNA gene sequences

#### 3. Genomic analysis via Protologger

 Each genome/16S pair will be run through Protologger to provide data

#### 4. Make the output available

- Put the raw Protologger output on the website for the community

- 5. Engage with the community
- Train members of the community so they can contribute to the analysis and describe taxa of interest to them



## Applying the ProtoBIOME framework to the human gut





MGYG-HGUT-00024
MGYG-HGUT-00025



# CandiBIOME framework and integration within NFDI4microbiota



## 1. Environment specific MAG collections2. Pro

# 2. Produce taxonomic, functional and ecological outputs

#### uts

**Ecology** Large-scale amplicon-based environmental survey MAG environmental survey

> Taxonomic placement ANI values POCP values mol% G+C difference Phylogeny 16S rRNA gene similarity (%)

Functionality CDS and CRISPR arrays Antibiotic resistance (CARD) Carbohydrate degradation (CAZy) Pathway analysis (KEGG)

4. Final curation of the crowd-sourced protologues

M2.1: Data and metadata standards M3.6: Long-term preservation of metadata 3. Teach the community how to understand the data and write a protologue



🔁 YouTube

M1.3: Community outreach M1.1: Training courses



****			
			ab
Thank your			
		🥣 🧖 tcabit	h i i i
		etterint	
	• • • • • • • • • • • •		
		NFDI4	
:::::::::::::::::::::::::::::::::::::::		NFDI4	:::::
		MICROBIOTA	:::::
		MICROBIOTA	
		NFDI4 MICROBIOTA	
		WWW.nfdi4microbiota.de	
		NFDI4 MICROBIOTA www.nfdi4microbiota.de	
		WWW.nfdi4microbiota.de	
		NFDI4 MICROBIOTA www.nfdi4microbiota.de @nfdi4microbiota	
		NFDI4 MICROBIOTA www.nfdi4microbiota.de @nfdi4microbiota	