Characterization of Prokaryote Strains for Taxonomic Purposes

Man Cai
China General Microbiological Culture Collection Center,
Institute of Microbiology,
Chinese Academy of Sciences
Nov. 2018
The tenet for novel taxa characterization: as comprehensively as possible.
Strain

Genetic-based characterization
- 16S rRNA gene
- Whole genome
- Phylogenetic tree
- DNA G+C content
- DNA-DNA hybridization
- Multilocus sequence analysis

Phenotypic characterization
- Morphology, physiology, biochemistry
  - Antibiotics resistance
  - Substrate utilization
  - Growth tolerance
  - Enzyme activity
  - Cell shape
  - Staining

Chemical characterization
- Fatty acids
- Polar lipids
- Polyamines
- Techoic acids
- Mycolic acids
- Peptidoglycan
- Lipopolysaccharides
- Respiratory lipoquinones
- Hydrophobic side chains of lipids
- Isoprenoid-based ether-linked lipids
I. Isolation information and importance of types

What should be remember before identification?

- Isolation method: Spread Plate Method

![Diagram showing soil samples and different media and temperatures]

- Location (e.g., GPS, latitude/longitude)
- Environment (e.g., pH, salinity, temperature, chemical composition)
- Designations (accession numbers and certifications from two culture collections e.g., CGMCC 1.9159\textsuperscript{T} = DSM 22955\textsuperscript{T})
I. Isolation information and importance of types

What should be remember before identification?

*Bacillus* Cohn 1872, *genus*. (Type genus of the order *Bacillales* Prévot 1953 [Approved Lists 1980]; type genus of the family *Bacillaceae* Fischer 1895 [Approved Lists 1980]).

**Type species:** *Bacillus subtilis* ( Ehrenberg 1835) Cohn 1872 (Approved Lists 1980).

**Etymology:** L. masc. n. *bacillus*, a small staff, a wand, a rod.


---

*Bacillus subtilis* (Ehrenberg 1835) Cohn 1872, *species*. (Type species of the genus.)

**Type strain:** (see also [StrainInfo.net](http://www.straininfo.net)) strain ATCC 6051 = ATCC 6051-U = CCM 2216 = BCRC (formerly CCRC) 10255 = CCUG 163 B = CFBP 4228 = CIP 52.65 = DSM 10 = IAM 12118 = IFO (now NBRC) 13719 = IFO (now NBRC) 16412 = IMET 10758 = JCM 1465 = LMG 7135 = NCAIM B.01095 = NCCB 70064 = NCCB 32009 = NCCB 53016 = NCIMB 3610 (formerly NCDO 1769) = NCTC 3610 = NRRL B-4219 = NRRL NRS-1315 = NRRL NRS-744 = VKM B-501.

**Sequence accession no.** (16S rRNA gene) for the type strain: [AJ276351](http://www.straininfo.net).

**Synonym:** "*Vibrio subtilis*" Ehrenberg 1835.

**Etymology:** L. masc. adj. *subtilis*, slender.
II. Genetic-based characterization

How to assign defined taxa by genetic data?

• Phylogenetic definition of species

  • DNA–DNA relatedness by DNA–DNA hybridization

  “The phylogenetic definition of a species generally would include strains with
  approximately 70% or greater DNA-DNA relatedness and with 5 °C or less
  ΔTm.”

  Wayne LG, Brenner DJ, Colwell RR. Report of the ad hoc committee on reconciliation

DNA-DNA relatedness
> 70 %

16S rRNA gene sequences
pairwise similarity
98.7 %

II. Genetic-based characterization

How to assign defined taxa by genetic data?

- 16S rRNA gene sequence
  - High quality (almost complete, no ambiguities)
II. Genetic-based characterization

How to assign defined taxa by genetic data?

• 16S rRNA gene sequences similarity
  • Multiple alignment (use expert-maintained seed alignments)
    ARB: www.arb-home.de
    RDP: http://rdp.cme.msu.edu
    SILVA: www.arb-silva.de

    Alternative:
    Robust multiple alignment programs (CLUSTAL_X, MEGA, T-COFFEE, MUSCLE) + by manual editing.

• Pairwise similarity
  Calculation programs: eg. ARB, PHYDIT, jPHYDIT,
  EZBioCloud (www.ezbiocloud.net/)

  Don’t use local alignment programs (eg. BLAST and FASTA).
II. Genetic-based characterization

How to assign defined taxa by genetic data?

- Whole genome sequence
  - High quality
    - The sequencing instrument, library reagents and method for genome assembly should be described in detail.
  - At least the following statistics should be given for the final genome assembly: (i) the obtained genome size, (ii) DNA G+C ratio, (iii) the number of contigs, (iv) N50 and (v) the sequencing depth of coverage.
- Accession number from one of the public databases must be included, and all sequence data must be made publicly available prior to submission.

II. Genetic-based characterization

How to assign defined taxa by genetic data?

• **Overall genome related index (OGRI)**
  
  • OGRI can be used to check if a strain belongs to a known species by calculating the relatedness between genome sequences of the strains and type strain of a species.

• **Average nucleotide identity (ANI) and digital DDH (dDDH)** have been most widely used. EZBioCloud (www.ezbiocloud.net/); DDGC 2.1 (http://ggdc.dsmz.de/distcalc2.php)

• The proposed and generally accepted **species boundary** for ANI and dDDH values are **95~96 and 70 %**, respectively.

II. Genetic-based characterization

*How to assign defined taxa by genetic data?*

- **Assignment to defined taxa**

  1. **Genome assembly**
  2. Obtain full length 16S rRNA gene sequence
  3. Search
  4. Find the phylogenetic neighbors
  5. If 16S sequence similarity is
     - < 98.7%
     - ≥ 98.7%
  6. New species
  7. If 95~96% ANI or 70% dDDH
     - ≥ 95~96% ANI or 70% dDDH
  8. Identified as a known species

II. Genetic-based characterization

*How to assign defined taxa by genetic data?*

- Phylogenetic tree

  *eg.* Based on 16S rRNA gene sequences

  by MEGA: upload sequences, alignment, construction
Genetic-based characterization

How to assign defined taxa by genetic data?

- Phylogenetic tree based on 16S rRNA gene sequences

- Use high quality sequences. Do not mix full and partial sequences.

- Use high quality alignments.

- Apply alternative treeing methods (e.g., Neighbor-joining, maximum-parsimony, maximum-likelihood methods).

- Never use sequences from single distantly related organisms as outgroup.

- Only bootstrap proportions of 70 or higher presented in the dendrogram.

Fig. 1, Neighbor-joining Method

Fig. 2, Maximum-parsimony Method

Fig. 3, Maximum-likelihood Method
III. Phenotypic characterization
--morphology, physiology, biochemistry

• Morphology

Scanning/transmission electron micrograph

- Cell shape and size
- Characteristic features (e.g., stalks, prosthecae, budding or branching, cell aggregates)
- Spore formation
- Location of flagella
- Motility (form, speed)
- Intracellular structures
- Colony shape and size
- Cellular pigments
III. Phenotypic characterization
--morphology, physiology, biochemistry

- Morphology

- Staining
  - Gram stain (the reaction may alter as the cells age)
  - Acid-fast staining (strains containing mycolic acids)
  - Sudan Black staining (stains containing lipophilic cellular inclusions, eg. polyhydroxybutyric acid)
III. Phenotypic characterization
--morphology, physiology, biochemistry

• Physiology and biochemistry
  • The growth tolerance (e.g. pH, temperature, NaCl concentration)
  • Enzyme activity, substrate utilization, antibiotics resistance, etc. Fast methods: API and Biology test plates.

Note:
• To test with identical media and conditions or at least comparable.
• To compare with type strain of type species of appropriate genera.
• To analyze including strains of the most closely related taxa rather than using the previously published data.
III. Phenotypic characterization -- morphology, physiology, biochemistry

The API identification system is numerical taxonomy according to the microbial physiological and biochemical characteristics.

- API 50 CH – Performance of carbohydrate metabolism tests
- API ZYM® – Semiquantification of enzymatic activities
- API 20E Reagent QC Test – Performance of QC testing for ferric chloride, alpha-naphthol, sulfaninic acid, NN-Dimethyl-Alpha-Naphthylamine and Kovac's Reagent
III. Phenotypic characterization
--morphology, physiology, biochemistry

eg. API 20E

Isolate | Prepare | Incubate | Read

Positive | Negative
III. Phenotypic characterization
-- chemical characterization

- Respiratory lipoquinones (cell membrane)
  - Three structural classes:
    - ubiquinones (classes α, γ, β-proteobacteria)
    - rhodoquinones (some of the classes α, γ-proteobacteria)
    - plastoquinones

Benzoquinones:

Naphthoquinones: * include menaquinones, demethylmenaquinones, mono-
methylmenaquin, dimethylmenaquinones and
menathioquinones.
  * the vast majority of *Bacteria* and *Archaea* known to synthesize
  naphthoquinone (menaquinone) derivatives.

Benzothiophene derivatives: members of the order *Sulfolobales*


• **Polar lipids (cell membrane)**

  • Vast diversity in prokaryotes and have yet to be fully elucidated.
  
  • Document the lipids by image with all visualized known and unidentified lipids.
  
  • Good image quality for publication:

    8 bit, grey scale, $7 \times 7$ cm, 300 d.p.i.
III. Phenotypic characterization
-- chemical characterization

• Peptidoglycan (outer cell layers)
  • Analysis is requirement for all novel Gram-positive species description.
  • Analysis includes characteristic diamino acid in the cross-linking peptide, peptidoglycan type (A or B), mode of cross-linkage, complete amino acid composition.

• Mycolic acids (outer cell layers)
  • occur in certain high G+C Gram-positive bacteria
  • additional taxonomic markers
III. Phenotypic characterization
--- chemical characterization

• Other constituents

• Lipopolysaccharides (LPS)
• Polyamines
• Hydrophobic side chains of lipids
• Non-/Isoprenoid-based ether-linked lipids

IV. Minimal standards

• To provide detailed information on the characterization of specific organisms, and complement these guidelines.

• Aerobic, endospore-forming bacteria (Logan et al., 2009)
• Anaerobic phototrophic bacteria (Imhoff & Caumette, 2004)
• Genus *Brucella* (Corbel & Brinley Morgan, 1975a, b)
• Family *Campylobacteraceae* (Ursing et al., 1994)
• Family *Flavobacteriaceae* (Bernardet et al., 2002)
• Order *Halobacteriales* (Oren et al., 1997)
• Family *Halomonadaceae* (Arahall et al., 2007; Arahall et al., 2008)
• Genus *Helicobacter* (Dewhirst et al., 2000)
• Methanogenic bacteria (*Archaea*) (Boone & Whitman, 1988)
• Suborder *Micrococcineae* (Schumann et al., 2009)
• Class *Mollicutes* (Division *Tenericutes*, Order *Mycoplasmatales*) (International Committee on Systematic Bacteriology Subcommittee on the Taxonomy of *Mollicutes*, 1979; International Committee on Systematic Bacteriology Subcommittee on the Taxonomy of *Mycoplasmatales*, 1972; Brown et al., 2007)
• Genera *Moraxella* and *Acinetobacter* (Bøvre & Henriksen, 1976)
• Genus *Mycobacterium* (Lévy-Frédault & Portaels, 1992)
• Family *Pasteurellaceae* (Christensen et al., 2007)
• Root and stem nodulating bacteria (Graham et al., 1991)
• Staphylococci (Freney et al., 1999)
• Genus *Streptomyces* (not a minimal standard, but a standard reference work, Shirling & Gottlieb, 1966)
Thanks for your attention!