

## 基于 16S rRNA 基因和基因组序列对细菌物种的初步鉴定

### Preliminary Identification of Bacterial Species Based on 16S rRNA Gene and Genomic Sequence

刘庆<sup>1,2</sup>, 辛玉华<sup>1,2</sup>, 东秀珠<sup>1,\*</sup>

<sup>1</sup>微生物资源前期开发国家重点实验室, 中国科学院微生物研究所, 北京; <sup>2</sup>中国普通微生物菌种保藏管理中心, 中国科学院微生物研究所, 北京

\*通讯作者邮箱: [dongxz@im.ac.cn](mailto:dongxz@im.ac.cn)

引用格式: 刘庆, 辛玉华, 东秀珠. (2021). 基于 16S rRNA 基因和基因组序列对细菌物种的初步鉴定. Bio-101 e2003932. Doi: 10.21769/BioProtoc.2003932.

How to cite: Liu, Q., Xin, Y. H. and Dong, X. Z. (2021). Preliminary Identification of Bacterial Species Based on 16S rRNA Gene and Genomic Sequence. Bio-101 e2003932. Doi: 10.21769/BioProtoc.2003932. (in Chinese)

**摘要:** 随着微生物培养组学技术的发展, 基于细菌纯培养物在物种和菌株水平的多样性研究日益受到重视。伴随着大量纯培养物的获取, 对其分类地位进行鉴定显得尤为重要。16S rRNA 基因依然是原核生物系统进化和分类学研究比较好的分子标识, 根据 16S rRNA 基因的序列同源性, 鉴定物种的分类学地位是目前最快捷的方法之一。目前 GenBank 和 EzBiocloud 数据库中存储了大量的原核生物模式菌株的 16S rRNA 基因序列, 通过 BLAST 在线比对可快速准确的鉴定未知细菌的分类学地位。本文的目的是基于本地化的数据库, 以 65 株 *Cryobacterium* 属菌株为例, 根据 16S rRNA 基因序列快速比对, 结合全基因组核苷酸和氨基酸序列一致性分析, 以及基于 16S rRNA 基因和全基因组序列的系统发育树构建, 明确分离菌株的分类学地位。对于培养组学研究过程中获取的大量菌种, 实现批量化鉴定十分重要。

**关键词:** 16S rRNA 基因, 菌种鉴定, ANI 分析, AAI 分析, 基因组系统树

## 材料与试剂

1. Chelex 树脂 (BIO-RAD, catalog number: 143-2832)
2. PCR 反应 Mix Taq: Premix Taq™ (Ex Taq™ Version 2.0) (宝生物工程 (大连) 有限公司, Takara, catalog number: RR003A)

## 仪器设备

1. PCR 仪 (德国 SENSO, Sensoquest LabCycler PCR 仪)
2. 离心机 (日本 TOMY, MicroOne 离心机)

## 软件和数据库

1. blast 2.8.1+: <https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.8.1/>; 操作系统: windows 或 linux。
2. FastANI v1.1: <https://github.com/ParBLiSS/FastANI>; 操作系统: linux。
3. R 语言 bactaxR 包: <https://github.com/lmc297/bactaxR>; 操作系统: windows 或 linux。
4. CompareM: <https://github.com/dparks1134/CompareM>; 操作系统: linux。
5. MAFFT V7: <https://mafft.cbrc.jp/alignment/software/>; 操作系统: windows 或 linux。
6. MEGA V5.2: <https://www.megasoftware.net/downloadform>; 操作系统: windows。
7. UBCG: <https://www.ezbiocloud.net/tools/ubcg>; 操作系统: linux。
8. 原核生物全基因组数据库: <https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/>
9. EzBioCloud 数据库: <https://www.ezbiocloud.net/>

## 实验步骤

1. 染色体 DNA 提取和纯化
  - 1.1 用无菌接种环或竹签挑取 1 个菌落, 加入 50  $\mu$ l 5% 的 Chelex 溶液, 充分振荡混匀。
  - 1.2 沸水浴 10-15 min。

1.3 2,000 × *g* 离心 1 min。上清液即为提取的基因组 DNA 溶液，-20 °C 保存备用。

## 2. 16S rRNA 基因的 PCR 扩增

50 μl 反应总体系包含：

2 × Mix Taq	25 μl
Primers (10 μM)	
27F	2.0 μl
1492R	2.0 μl
DNA 模板 (100 ng/μl)	2.0 μl
ddH <sub>2</sub> O	19 μl

其中扩增引物：27F (5'- AGAGTTTGATCMTGGCTCAG-3') 和 1492R (5'- GGTTACCTTGTTACGACTT-3')，为生工生物工程 (上海) 股份有限公司合成。

PCR 扩增参数：94 °C 预变性，4 min；94 °C 变性，1 min；55 °C 复性，1 min；72 °C 延伸，1.5 min，共 30 个循环；72 °C 延伸 10 min。

## 3. 16S rRNA 基因序列测定

将 16S rRNA 基因扩增产物送至测序公司测序。

## 4. 16S rRNA 基因序列本地化比对分析

### 4.1 本地化数据库构建

用于 BLAST 分析的原核生物模式菌株的 16S rRNA 基因序列本地化数据库，可直接在美国国家生物技术信息中心 (National Center for Biotechnology Information, NCBI) 的 ftp 站点 ([ftp://ftp.ncbi.nih.gov/blast/db/16S\\_ribosomal\\_RNA.tar.gz](ftp://ftp.ncbi.nih.gov/blast/db/16S_ribosomal_RNA.tar.gz)) 下载。本文以构建 *Cryobacterium* 属模式菌株 16S rRNA 基因序列本地化数据库为例，在 <https://lpsn.dsmz.de/genus/cryobacterium> 网页下载 *Cryobacterium* 属模式菌株的 16S rRNA 基因序列，并通过检索 EzBioCloud 数据库查看该属是否有更新，最终确定 *Cryobacterium* 属共 15 个模式菌株。序列命名为 Cryo16S.fasta 后，利用 BLAST+ 工具构建本地化数据库。在 windows 系统 CMD 命令行窗口中，输入以下命令：`makeblastdb.exe -in Cryo16S.fasta -parse_seqids -`

hash\_index -dbtype nucl -out D:\blast\db\Cryo16S, 如图 1, 生成以 Cryo16S 命名的本地化数据库。

```
D:\blast>makeblastdb.exe -in Cryo16S.fasta -parse_seqids -hash_index -dbtype nucl -out D:\blast\db\Cryo16S

Building a new DB, current time: 01/22/2021 10:54:16
New DB name: D:\blast\db\Cryo16S
New DB title: Cryo16S.fasta
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 15 sequences in 0.0132706 seconds.
```

图 1. 本地数据库构建

#### 4.2 16S rRNA 基因序列比对

本文以 Liu *et al.* (2020) 论文中报道的 *Cryobacterium* 属分离菌株数据为例, 可在文章所列的菌株信息附表中查询到 16S rRNA 基因序列的 GenBank 号。根据分离菌株的 GenBank 号, 利用 NCBI 的 Entrez 在线工具 (<https://www.ncbi.nlm.nih.gov/sites/batchentrez>) 下载原始数据。准备好其中 65 株菌株的 16S rRNA 基因序列 fasta 文件 (命名为 Cryo\_query.fasta) 后, 利用 BLAST+ 工具比对。在 windows 系统 CMD 命令行窗口中, 输入以下命令: blastn.exe -query Cryo\_query.fasta -out Cryo\_16S\_blast.xls -db D:\blast\db\Cryo16S -outfmt 6 -evaluate 1e-5 -max\_target\_seqs 1 -num\_threads 8, 如图 2。

```
D:\blast>blastn.exe -query Cryo_query.fasta -out Cryo_16S_blast.xls -db D:\blast\db\Cryo16S -outfmt 6 -evaluate 1e-5 -max_target_seqs 1 -num_threads 8
Warning: [blastn] Examining 5 or more matches is recommended
```

图 2. 本地 blast 命令

生成 Cryo\_16S\_blast.xls 命名的结果文件 (表 1), 每条序列输出一条与其相似性最高的物种信息。

Kim *et al.* (2014) 通过对数千个基因组和 16S rRNA 基因序列的统计分析发现, 当两菌株 16S rRNA 基因序列之间相似性低于  $\approx 98.65\%$  时, 可判断它们归属于不同的种; 而高于  $\approx 98.65\%$  时, 可能属于同一个种, 也可能属于不同的种, 需结合全基因组序列分析等其他方法判定。根据比对结果可知, 仅有两个菌株与已知物种间序列相似性低于 98.65%, 为潜在的新物种, 其余菌株与已知物

种最高相似在 98.76-100%之间, 因此, 需根据其基因组序列计算全基因组平均核苷酸序列一致性 (average nucleotide identity, ANI), 得到精确的物种鉴定结果。

**表 1. *Cryobacterium* 菌株的 16S rRNA 基因序列比对结果**

菌株	最相似的模式菌株	相似性 (%)
<b>MDB2-33-2</b>	<i>C. breve</i> TMT4-23 <sup>T</sup>	<b>98.42</b>
<b>MDB2-10</b>	<i>C. breve</i> TMT4-23 <sup>T</sup>	<b>98.48</b>
MDB2-B, MDB1-5, MDB1-18-1, MDB2-A-2, TMT1-19, MDB2-A-1, TMT1-23-1, TMT3-29-2, Hz7, TMT1-62, TMT2-48-2, Sr3, Hh4, MDB1-18-2, TMT2-16, TMT2-14, TMT2-17-1, TMT2-4, TMT4-10, TMT2-59, TMT1-21, TMT2-42-4, TMT1-22, TMT2-18-3, TMT2-18-2, TMT2-15-1, TMT2-10, TMT2-23, MDT1-3, HLT2-28, TMT1-51, HLT2-23, RHLT2-21, TMT4-31	<i>C. breve</i> TMT4-23 <sup>T</sup>	98.76-99.93
Hh14	<i>C. melibiosiphilum</i> Hh39 <sup>T</sup>	99.22
Sr8	<i>C. psychrotolerans</i> 0549 <sup>T</sup>	99.35
RHLS22-1, Sr59	<i>C. soli</i> GCJ02 <sup>T</sup>	99.57-99.93
Sr54, Hh38, Hz9, Hz16, Sr39, TMS1-13-1	<i>C. levicorallinum</i> Hh34 <sup>T</sup>	99.65-99.93
TMB1-8, TMB1-7, TMB3-15, TMB3-10, TMN-39-1, TMB3-12, TMB3-1-2	<i>C. zongtaii</i> TMN-42 <sup>T</sup>	99.658-99.93
TMT1-1, Hb1	<i>C. luteum</i> Hh15 <sup>T</sup>	99.86
TMT1-66-1	<i>C. flavum</i> Hh8 <sup>T</sup>	99.93
Hh7, TMS1-20-1, Hh11	<i>C. levicorallinum</i> Hh34 <sup>T</sup>	100
TMT1-3	<i>C. luteum</i> Hh15 <sup>T</sup>	100
TMT1-2-2	<i>C. flavum</i> Hh8 <sup>T</sup>	100
TMN-39-2	<i>C. zongtaii</i> TMN-42 <sup>T</sup>	100
TMT3-12	<i>C. breve</i> TMT4-23 <sup>T</sup>	100
TMT1-2-1, Sr47	<i>C. psychrotolerans</i> 0549 <sup>T</sup>	100

## 5. 基因组平均核苷酸和氨基酸序列一致性分析

ANI 是基于两两基因组之间所有直系同源蛋白编码基因序列比较的平均值, 该值可反映基因组之间的进化距离。基因组之间的 ANI 值与 16S rRNA 基因序列相似性分析及 DNA-DNA 杂交结果相一致, 因此, ANI 分析方法已经代替繁琐的 DNA-DNA 杂交技术。两菌株基因组 ANI 值为 95-96%时, 相当于 DNA-DNA 杂交值 70%。因此, 当两菌株基因组 ANI 值大于 96%时, 鉴定为同一个物种; 小于 95%时, 鉴定为不同物种 (Richter *et al.*, 2009)。

## 5.1 基因组数据准备

本文所使用全基因组数据来源于 Liu *et al.* (2020) 文章或原核生物全基因组数据库。目前有多个工具可直接对全基因组进行 ANI 值计算，如，Jspeciesws 和 ANI Calculator 等在线计算工具，Jspecies 等图形界面工具，以及 FastANI、autoANI 等命令行工具 (表 2)。本文选择计算速度最快、操作最为方便的 FastANI 工具介绍，FastANI 需在 linux 系统中运行，可在普通计算机中安装 ubuntu 系统运行 FastANI 工具。首先建立 *Cryobacterium* 属模式菌株以及待鉴定菌株的全基因组序列文件，准备两个文本文件：将待比对序列文件名称保存至 query.list 文件 (图 3a)，参比序列文件名称保存至 ref.list (图 3b)，其中每一个基因组文件占用一行，用于 FastANI 软件计算的输入文件。

表 2. ANI 计算工具

名称	软件网页
FastANI	<a href="https://github.com/ParBLISS/FastANI">https://github.com/ParBLISS/FastANI</a>
Jspeciesws	<a href="http://jspecies.ribohost.com/jspeciesws/#Home">http://jspecies.ribohost.com/jspeciesws/#Home</a>
ANI Calculator	<a href="http://www.ezbiocloud.net/tools/ani">http://www.ezbiocloud.net/tools/ani</a>
Jspecies	<a href="http://imedea.uib-csic.es/jspecies/docu.html">http://imedea.uib-csic.es/jspecies/docu.html</a>
OAT	<a href="http://www.ezbiocloud.net/tools/orthoani">http://www.ezbiocloud.net/tools/orthoani</a>
OrthoANI	<a href="http://www.ezbiocloud.net/tools/">http://www.ezbiocloud.net/tools/</a>
autoANI	<a href="https://github.com/osuchanglab/autoANI">https://github.com/osuchanglab/autoANI</a>
GET_HOMOLOGUE S	<a href="https://github.com/eead-csic-compbio/get_homologues">https://github.com/eead-csic-compbio/get_homologues</a>

**a**

```

1 ~/data/my_genomes/Cryobacterium_sp._Hh1.fna
2 ~/data/my_genomes/Cryobacterium_sp._Hh11.fna
3 ~/data/my_genomes/Cryobacterium_sp._Hh14.fna
4 ~/data/my_genomes/Cryobacterium_sp._Hh38.fna
5 ~/data/my_genomes/Cryobacterium_sp._Hh4.fna
6 ~/data/my_genomes/Cryobacterium_sp._Hh7.fna
7 ~/data/my_genomes/Cryobacterium_sp._HLT2-23.fna
8 ~/data/my_genomes/Cryobacterium_sp._HLT2-28.fna
9 ~/data/my_genomes/Cryobacterium_sp._H216.fna
10 ~/data/my_genomes/Cryobacterium_sp._H27.fna
11 ~/data/my_genomes/Cryobacterium_sp._H29.fna
12 ~/data/my_genomes/Cryobacterium_sp._MDB1-18-1.fna
13 ~/data/my_genomes/Cryobacterium_sp._MDB1-18-2.fna
14 ~/data/my_genomes/Cryobacterium_sp._MDB1-5.fna
15 ~/data/my_genomes/Cryobacterium_sp._MDB2-10.fna

```

**b**

```

1 ~/data/my_genomes/Cryobacterium_arcticum_PAMC_27867.fna
2 ~/data/my_genomes/Cryobacterium_aureum_Hh31.fna
3 ~/data/my_genomes/Cryobacterium_breve_TMT4-23.fna
4 ~/data/my_genomes/Cryobacterium_flavum_Hh8.fna
5 ~/data/my_genomes/Cryobacterium_levicorallinum_Hh34.fna
6 ~/data/my_genomes/Cryobacterium_luteum_Hh15.fna
7 ~/data/my_genomes/Cryobacterium_messophilum_DSM_19267.fna
8 ~/data/my_genomes/Cryobacterium_psychrophilum_DSM_4854.fna
9 ~/data/my_genomes/Cryobacterium_psychrotolerans_CGMCC_1.5382.fna
10 ~/data/my_genomes/Cryobacterium_roopkundense_DSM_21065.fna
11 ~/data/my_genomes/Cryobacterium_tuani_i_3r36.fna
12 ~/data/my_genomes/Cryobacterium_solii_GCJ02.fna
13 ~/data/my_genomes/Cryobacterium_sp._Hh39.fna
14 ~/data/my_genomes/Cryobacterium_tepidiphilum_NEAU-85.fna
15 ~/data/my_genomes/Cryobacterium_zongtali_TMN-42.fna

```

图 3. query.list 和 ref.list 文件部分内容示例

## 5.2 ANI 计算

打开 shell 窗口后，进入基因组文件、query.list 和 ref.list 文件所在的目录，输入命令：fastANI --ql query.list --rl ref.list -o ANI.out。运行过程如图 4，计算结束后，输出结果到 ANI.out 文件。



MDB2-A-2	<i>C. psychrotolerans</i> 0549 <sup>T</sup>	81.20	新种
TMT1-19	<i>C. breve</i> TMT4-23 <sup>T</sup>	87.00	新种
MDB2-A-1	<i>C. psychrotolerans</i> 0549 <sup>T</sup>	81.19	新种
Hh14	<i>C. melibiosiphilum</i> Hh39 <sup>T</sup>	87.43	新种
TMT1-23-1	<i>C. breve</i> TMT4-23 <sup>T</sup>	87.35	新种
Tmt3-29-2	<i>C. breve</i> TMT4-23 <sup>T</sup>	87.47	新种
HZ7	<i>C. breve</i> TMT4-23 <sup>T</sup>	86.80	新种
Sr8	<i>C. psychrotolerans</i> 0549 <sup>T</sup>	98.15	<i>C. psychrotolerans</i>
TMT1-62	<i>C. breve</i> TMT4-23 <sup>T</sup>	86.76	新种
Tmt2-48-2	<i>C. breve</i> TMT4-23 <sup>T</sup>	87.25	新种
Sr3	<i>C. breve</i> TMT4-23 <sup>T</sup>	86.86	新种
Hh4	<i>C. breve</i> TMT4-23 <sup>T</sup>	82.71	新种
MDB1-18-2	<i>C. psychrotolerans</i> 0549 <sup>T</sup>	81.22	新种
TMT2-16	<i>C. breve</i> TMT4-23 <sup>T</sup>	86.86	新种
TMT2-14	<i>C. breve</i> TMT4-23 <sup>T</sup>	86.82	新种
TMT2-17-1	<i>C. breve</i> TMT4-23 <sup>T</sup>	86.87	新种
TMT2-4	<i>C. breve</i> TMT4-23 <sup>T</sup>	86.87	新种
TMT4-10	<i>C. breve</i> TMT4-23 <sup>T</sup>	84.45	新种
Tmt2-59	<i>C. breve</i> TMT4-23 <sup>T</sup>	84.46	新种
TMT1-21	<i>C. breve</i> TMT4-23 <sup>T</sup>	84.43	新种
TMT2-42-4	<i>C. breve</i> TMT4-23 <sup>T</sup>	86.88	新种
TMT1-22	<i>C. breve</i> TMT4-23 <sup>T</sup>	84.34	新种
TMT2-18-3	<i>C. breve</i> TMT4-23 <sup>T</sup>	86.85	新种
TMT2-18-2	<i>C. breve</i> TMT4-23 <sup>T</sup>	86.80	新种
TMT2-15-1	<i>C. breve</i> TMT4-23 <sup>T</sup>	87.00	新种
TMT2-10	<i>C. breve</i> TMT4-23 <sup>T</sup>	84.47	新种
TMT2-23	<i>C. breve</i> TMT4-23 <sup>T</sup>	84.21	新种
RHLS22-1	<i>C. zongtaii</i> TMN-42 <sup>T</sup>	90.73	新种
MDT1-3	<i>C. breve</i> TMT4-23 <sup>T</sup>	81.64	新种
Sr54	<i>C. aureum</i> Hh31 <sup>T</sup>	86.47	新种
TMB1-8	<i>C. zongtaii</i> TMN-42 <sup>T</sup>	96.31	<i>C. zongtaii</i>
TMB1-7	<i>C. zongtaii</i> TMN-42 <sup>T</sup>	97.87	<i>C. zongtaii</i>
HLT2-28	<i>C. breve</i> TMT4-23 <sup>T</sup>	82.57	新种
TMT1-51	<i>C. breve</i> TMT4-23 <sup>T</sup>	89.38	新种
HLT2-23	<i>C. luteum</i> Hh15 <sup>T</sup>	90.01	新种
Hh38	<i>C. levicorallinum</i> Hh34 <sup>T</sup>	99.11	<i>C. levicorallinum</i>
HZ9	<i>C. ruanii</i> Sr36 <sup>T</sup>	86.61	新种

RHLT2-21	<i>C. breve</i> TMT4-23 <sup>T</sup>	82.56	新种
TMT1-1	<i>C. aureum</i> Hh31 <sup>T</sup>	86.13	新种
Hb1	<i>C. aureum</i> Hh31 <sup>T</sup>	85.92	新种
Hz16	<i>C. ruanii</i> Sr36 <sup>T</sup>	86.83	新种
TMB3-15	<i>C. zongtaii</i> TMN-42 <sup>T</sup>	97.89	<i>C. zongtaii</i>
TMB3-10	<i>C. zongtaii</i> TMN-42 <sup>T</sup>	97.86	<i>C. zongtaii</i>
Sr39	<i>C. aureum</i> Hh31 <sup>T</sup>	87.46	新种
TMT1-66-1	<i>C. flavum</i> Hh8 <sup>T</sup>	97.33	<i>C. flavum</i>
TMT4-31	<i>C. zongtaii</i> TMN-42 <sup>T</sup>	97.86	<i>C. zongtaii</i>
TMN-39-1	<i>C. zongtaii</i> TMN-42 <sup>T</sup>	99.10	<i>C. zongtaii</i>
Sr59	<i>C. zongtaii</i> TMN-42 <sup>T</sup>	88.16	新种
TMS1-13-1	<i>C. levicorallinum</i> Hh34 <sup>T</sup>	96.49	<i>C. levicorallinum</i>
TMB3-12	<i>C. zongtaii</i> TMN-42 <sup>T</sup>	97.90	<i>C. zongtaii</i>
TMB3-1-2	<i>C. zongtaii</i> TMN-42 <sup>T</sup>	97.87	<i>C. zongtaii</i>
Hh7	<i>C. levicorallinum</i> Hh34 <sup>T</sup>	97.02	<i>C. levicorallinum</i>
TMS1-20-1	<i>C. levicorallinum</i> Hh34 <sup>T</sup>	97.09	<i>C. levicorallinum</i>
Hh11	<i>C. levicorallinum</i> Hh34 <sup>T</sup>	96.94	<i>C. levicorallinum</i>
TMT1-3	<i>C. luteum</i> Hh15 <sup>T</sup>	99.49	<i>C. luteum</i>
TMT1-2-2	<i>C. flavum</i> Hh8 <sup>T</sup>	97.46	<i>C. flavum</i>
TMN-39-2	<i>C. zongtaii</i> TMN-42 <sup>T</sup>	99.07	<i>C. zongtaii</i>
TmT3-12	<i>C. breve</i> TMT4-23 <sup>T</sup>	99.99	<i>C. breve</i>
TMT1-2-1	<i>C. psychrotolerans</i> 0549 <sup>T</sup>	99.58	<i>C. psychrotolerans</i>
Sr47	<i>C. psychrotolerans</i> 0549 <sup>T</sup>	90.92	新种

### 5.3 平均氨基酸一致性分析

ANI 值可作为种水平的鉴定方法之一,而在更高分类单元的鉴定可通过计算氨基酸一致性 (amino acid identity, AAI) 进行初步判定,在属水平的 AAI 阈值大约为 65-72% (Konstantinidis and Tiedje, 2007)。以本文中的基因组数据为例,AAI 值可通过 CompareM 软件计算,运行操作系统为 ubuntu。以 ~/data/my\_genomes 文件夹中的基因组序列为输入文件,输入命令: `comparem aai_wf ~/data/my_genomes aai_output`。

运行结束后,结果保存至 aai\_output 文件夹,AAI 值保存在 aai\_summary.tsv 文件,结果显示 *Cryobacterium mesophilum* DSM 19267<sup>T</sup> 与其他 79 株菌 (包括 14 个模式菌株和 65 个分离菌株) 为 64.3-65.86%, 其余菌株两两间 AAI 值为 70.86-100%。

### 5.4 ANI 聚类分析

以上过程初步得到了每株菌的鉴定结果，为了分析菌株间亲缘关系，可对两两菌株间 ANI 值聚类分析。将所有待聚类分析的序列文件路径保存至 `pairwise.list`，打开 shell 窗口后，进入工作目录，输入命令：

```
fastANI --ql pairwise.list --rl pairwise.list -o ANI_pairwise.out
```

得到 `ANI_pairwise.out` 结果文件，结果包含两两菌株间 ANI 值。以此文件为输入文件，在 R 语言 `bactaxR` 包中进行聚类分析并生成带有 95% 和 96% 阈值的聚类树状图，操作系统 windows 或 linux 均可。代码如下：

```
1 library(bactaxR)
2 ani <- read.ANI(file = "ANI_pairwise.out")
3 dend <- ANI.dendrogram(bactaxRObject = ani, ANI_threshold = 95,
4   xline = c(4,5), xlinecol = c("#ffc425", "#f37735"), Label_size = 0.5)
```

图 5. ANI 聚类分析代码

代码逻辑为：

- a) 加载 `bactaxR` 程序包；
- b) 读取 `ANI_pairwise.out` 文件，命名为 `ani`；
- c) 运行 `bactaxR` 程序包中的 `ANI.dendrogram` 函数，该函数将两两间 ANI 值转换为距离矩阵后，利用 `hclust` 函数中的非加权组平均法 (UPGMA) 聚类。

运行结束后，生成 ANI 聚类图，如图 6 所示。以 95% 为阈值，本文中 65 株待分析菌株和 15 个已知物种，可划分为 34 个种，其中 45 株菌分别归属于 19 个新种。

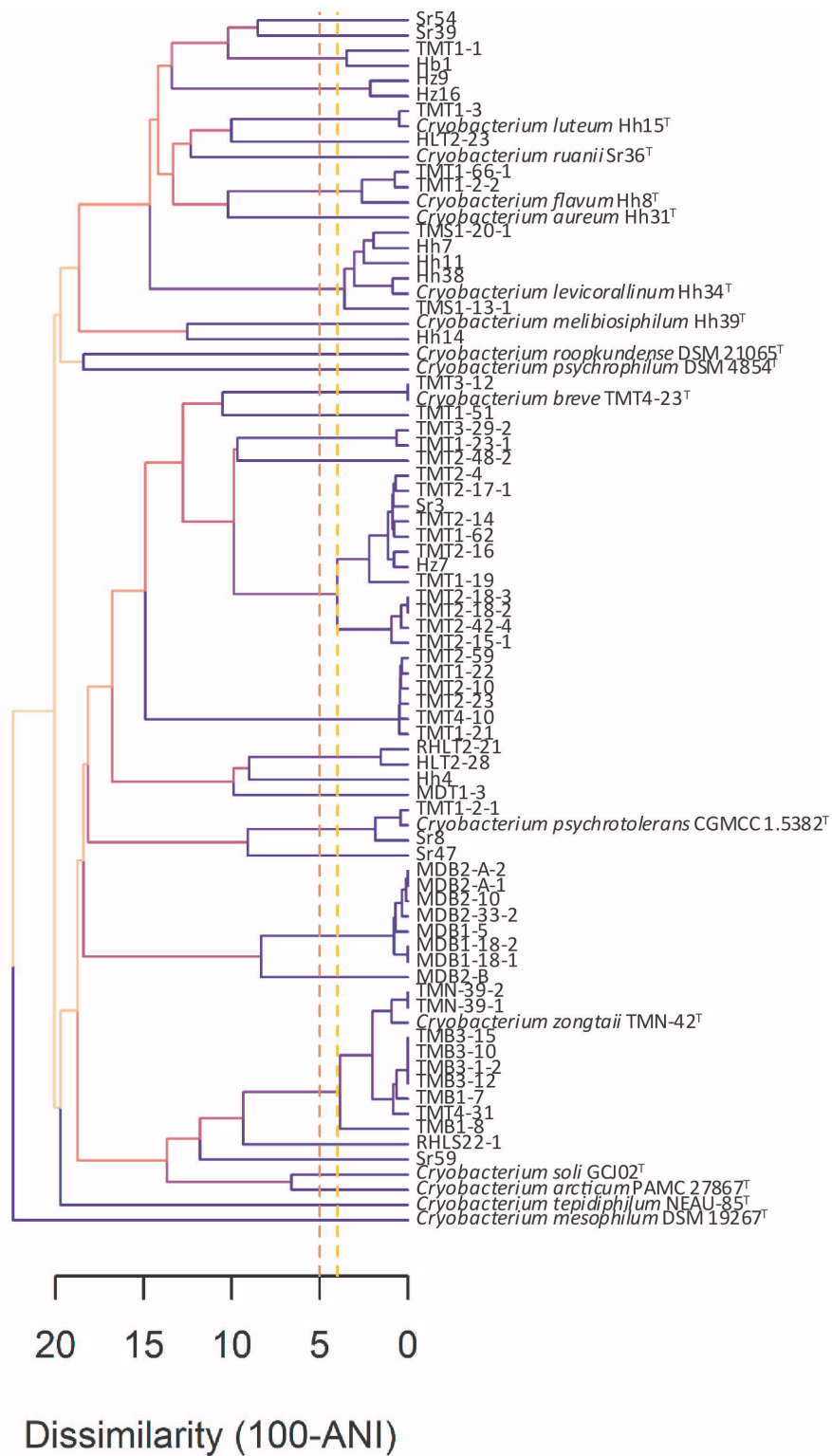


图 6. ANI 聚类分析结果

## 6. 系统发育分析

系统发育分析是细菌分类、鉴定和命名的基础，建立在物种间的进化关系上，而不是普遍相似性基础上。本文利用上述 *Cryobacterium* 数据为例，构建基于 16S rRNA 基因序列和全基因组序列的系统发育树。

### 6.1 16S rRNA 基因系统发育树构建

首先,将待分析的 16S rRNA 基因序列保存为一个 **fasta** 格式文件 (16S.fasta), 采用 **mafft** 软件对序列进行 **alignment**, 命令如下:

```
mafft 16S.fasta > 16S_aligned.fas
```

输出文件为 16S\_aligned.fas。打开 MEGA 软件, 依次点击 **Phylogeny—Construct/Test Neighbor-joining Tree**, 找到 16S\_aligned.fas 文件并选中后, 选择 **Nucleotide sequences**, 点击 **OK**, 进入参数选择界面, 调整 **Bootstrap method** 等参数后 (图 7), 点击 **Compute**, 生成系统发育树 (图 8)。

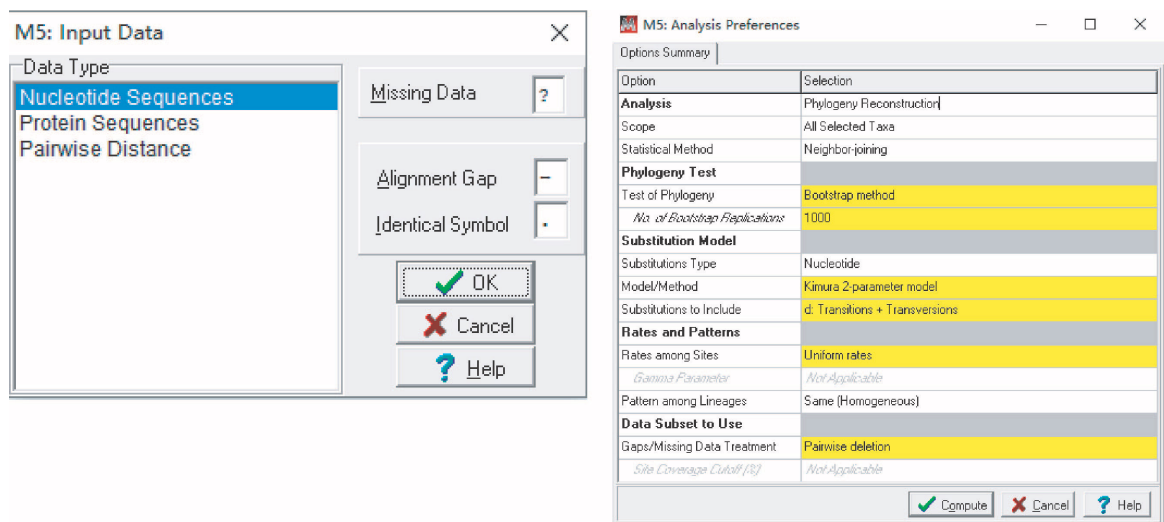


图 7. MEGA 软件 Neighbor-joining 系统发育分析参数示例

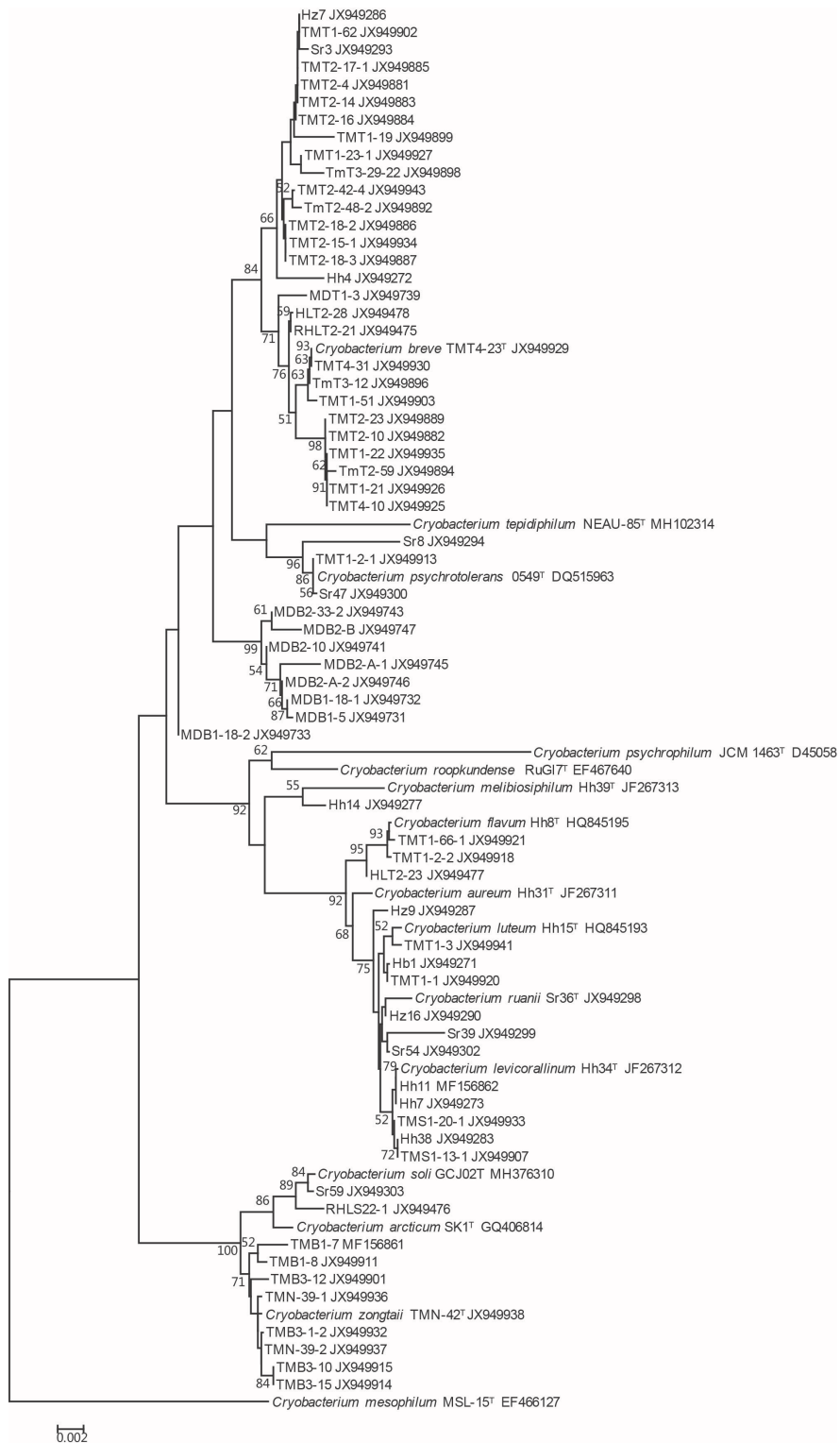


图 8. *Cryobacterium* spp. 基于 Neighbor-joining 方法的系统发育树 *C. mesophilum* MSL-15<sup>T</sup> 设置为外群。图中显示了大于 50% 的 Bootstrap 值，标尺代表 0.2% 核苷酸差异。

## 6.2 基于基因组的系统发育树构建

本文以 UBCG 工具为例，构建基于基因组序列的系统发育树。UBCG 可快速抽取细菌 92 个 (或以下) 共有的保守单拷贝基因序列，然后对每个基因序列和串联序列分别构建基于最大似然法的系统发育树。运行命令如图 9 所示：

```
1 cd ~/soft/UBCG/
2 for i in /home/liuq/data/my_genomes/*.fna
3 do
4 java -jar UBCG.jar extract -bcg_dir bcg -i $i -Label $i -acc "" -taxon "" -strain "" -type
5 done
6
7 java -jar UBCG.jar align -bcg_dir bcg -prefix Cryo_genome_tree
```

图 9. 利用 UBCG 构建基因组系统树命令

代码解释：首先进入 UBCG 软件所在目录，将~/data/my\_genomes 文件夹中所有 fna 后缀的基因组序列转换为 bcg 格式文件，此处应注意基因组序列文件名称不能含有空格。bcg 文件保存至 UBCG 软件目录下的 bcg 文件夹，全部转换完成后，执行系统树构建命令，结果保存至~/soft/UBCG/output/Cryo\_genome\_tree 文件夹。基于核心基因组的系统发育树如图 10 所示。

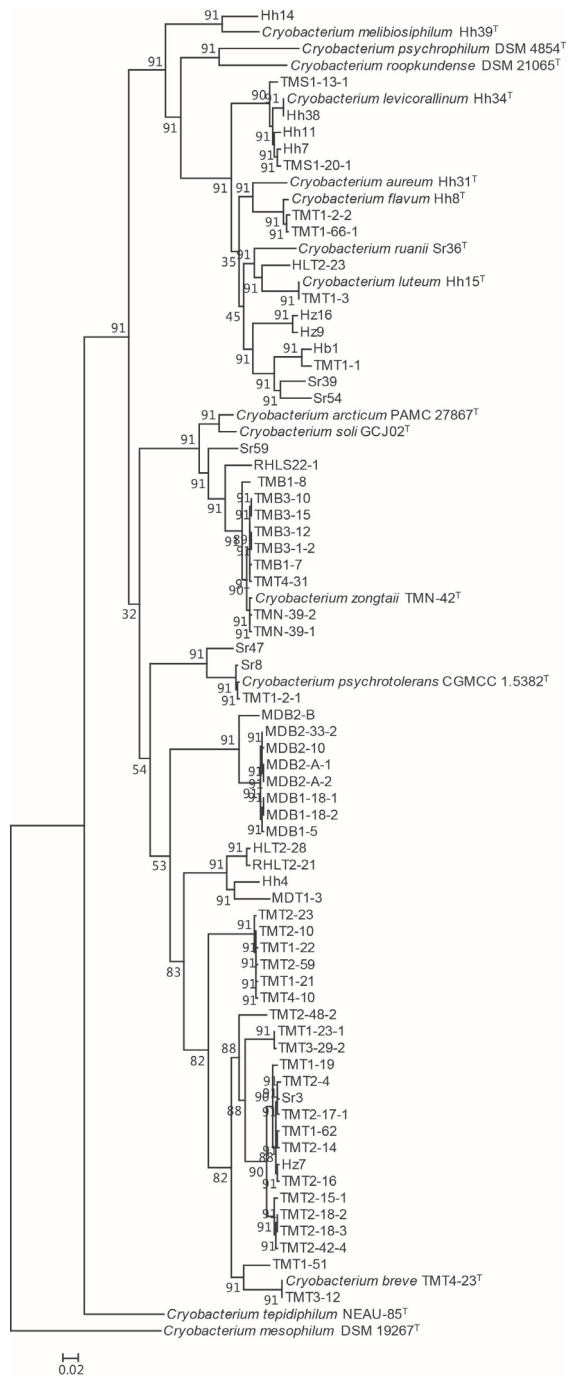


图 10. *Cryobacterium* spp. 基于 91 个单拷贝持家基因的系统发育树 *C. mesophilum* MSL-15<sup>T</sup> 设置为外群。节点处显示出基因支持指数 (GSI) 值，标尺代表 2%核苷酸差异。

## 致谢

本文所使用数据内容由国家自然科学基金青年基金项目 (31600007) 资助，数据来源于 Liu *et al.* (2020)。

## 参考文献

1. Kim, M., Oh, H. S., Park, S. C. and Chun, J. (2014). [Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes](#). *Int J Syst Evol Microbiol* 64: 346-351.
2. Liu, Q., Song, W. Z., Zhou, Y. G., Dong, X. Z. and Xin, Y. H. (2020). [Phenotypic divergence of thermotolerance: Molecular basis and cold adaptive evolution related to intrinsic DNA flexibility of glacier-inhabiting \*Cryobacterium\* strains](#). *Environ Microbiol* 22: 1409-1420.
3. Richter, M. and Rosselló-Móra, R. (2009). [Shifting the genomic gold standard for the prokaryotic species definition](#). *Proc Natl Acad Sci USA* 106: 19126-19131.
4. Konstantinidis, K. T. and Tiedje, J. M. (2007). [Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead](#). *Curr Opin Microbiol* 10: 504-9.