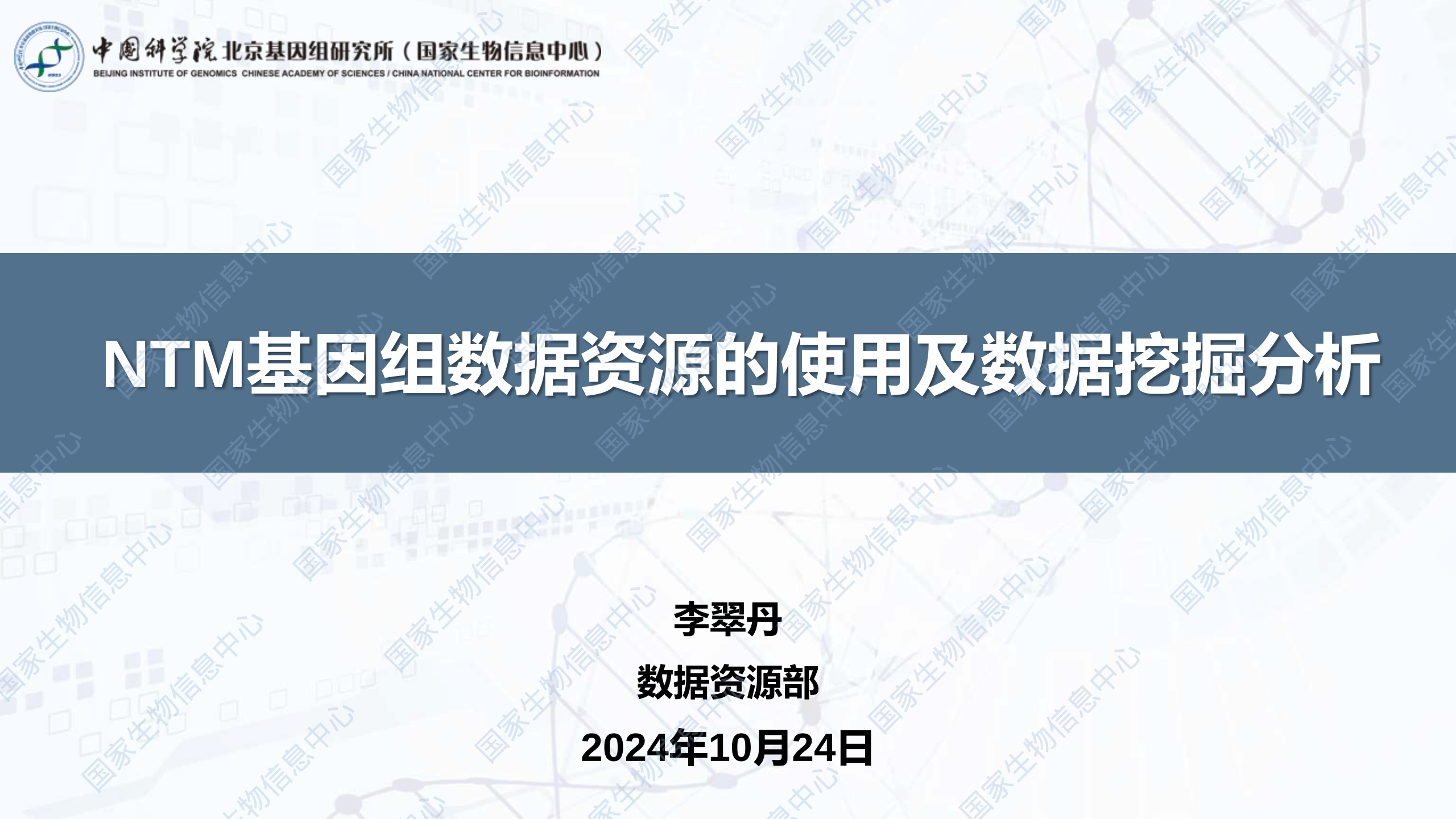




中国科学院北京基因组研究所（国家生物信息中心）
BEIJING INSTITUTE OF GENOMICS CHINESE ACADEMY OF SCIENCES / CHINA NATIONAL CENTER FOR BIOINFORMATION

NTM基因组数据资源的使用及数据挖掘分析

李翠丹

数据资源部

2024年10月24日

研究方向

致病微生物的基因组学研究：结核分枝杆菌、非结核分枝杆菌、肺炎克雷伯菌、铜绿假单胞菌、支原体

□ 全国首次结核病耐药性基线调查耐多药结核杆菌的组学研究;
Clin Infect Dis (2019)

□ 基于外泌体的潜伏结核病人液体活检技术研究;
Front Microbiol (2019);
Front Microbiol (2017)
□ 基于代谢组学的菌阴结核精准诊断研究;
BMC Infect Dis (2022)

□ 结核病危险因素与高风险人群早期筛查
(文章撰写)
□ 不同毒力Mtb感染巨噬细胞后转录组的动态变化研究(文章撰写)
□ 结核肉芽肿单细胞与空间转录组学研究
(文章撰写)

□ 肺结核患者痰液、肺泡盥洗液宏基因组研究 (正在开展)

□ TB泛基因组学研究;
Front Microbiol (2018)
□ MTBC比较基因组学研究;
Front Cell Infect Microbiol (2017)
□ NTM基因组学研究;
Microb Biotechnol (2021)



内容提纲

1

细菌基因组分析的流程和经验

2

NTM基因组分析、数据资源介绍及应用案例

细菌基因组解析的基础模块

细菌基因组组装

基因预测与注释

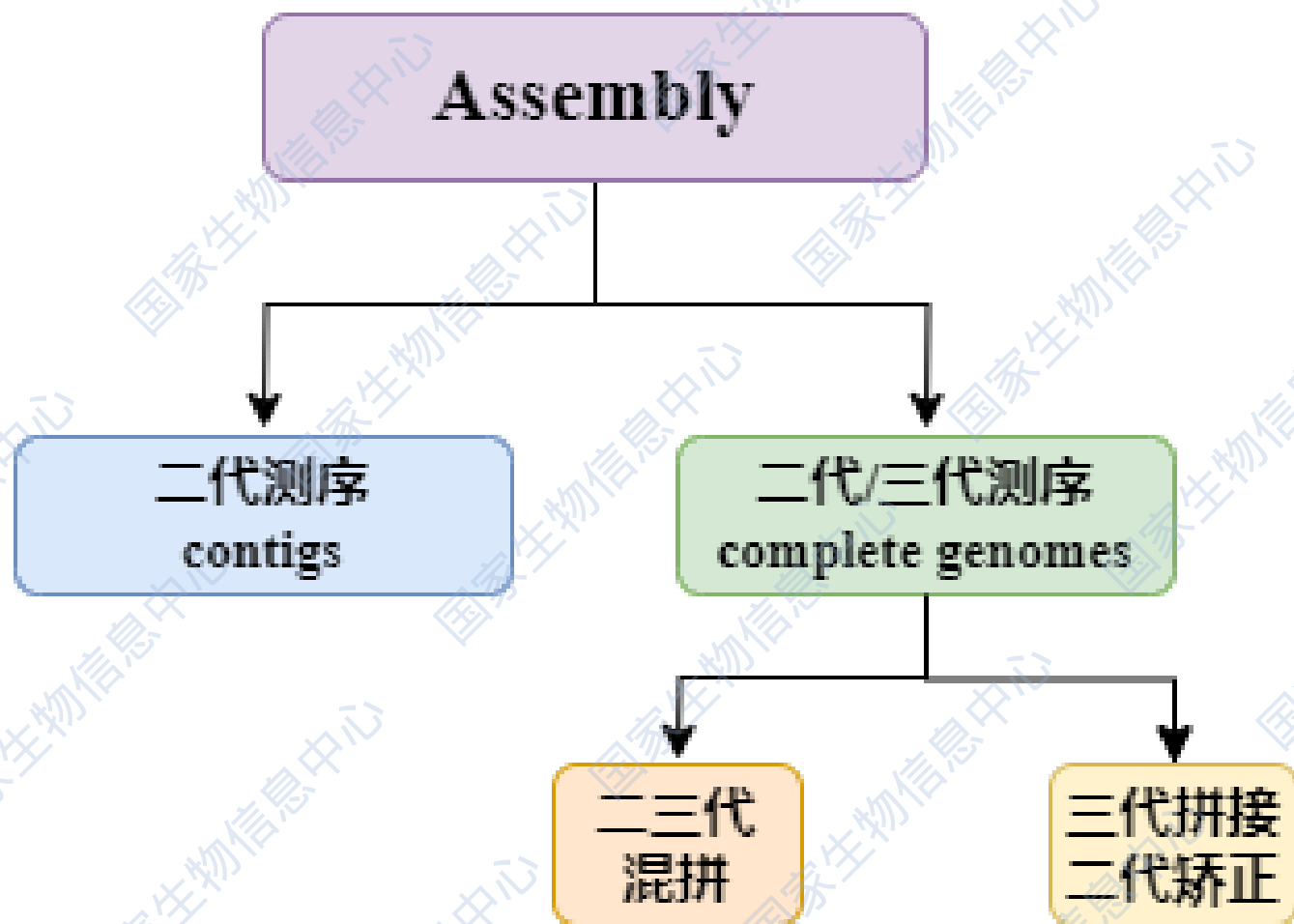
分子分型

进化

泛基因组

耐药和毒力

模块一：细菌基因组组装



□ 二代短读长测序 (Illumina)

✓ Contigs水平

□ 二代短读长+三代长读长测序 (Nanopore/Pacbio)

✓ Complete genome

✓ 二代短reads+三代长reads混拼

✓ 三代长reads拼接, 二代短reads进行矫正

模块一：细菌基因组组装-二代数据

二代短读长测序 (Illumina)

SPAdes (<https://github.com/ablab/spades>)

- 功能强大，其既可用于细菌基因组组装，也可用于宏基因组组装，既可以用于二代测序组装，也可以用于二代+三代测序数据的联合组装。
- 对计算硬件的要求较低

spades.py -1 left.fastq.gz -2 right.fastq.gz -o output_folder

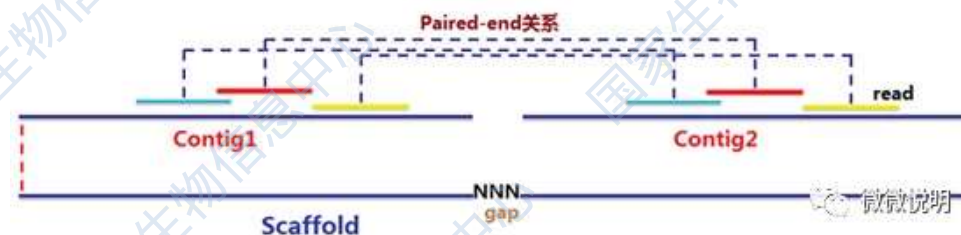
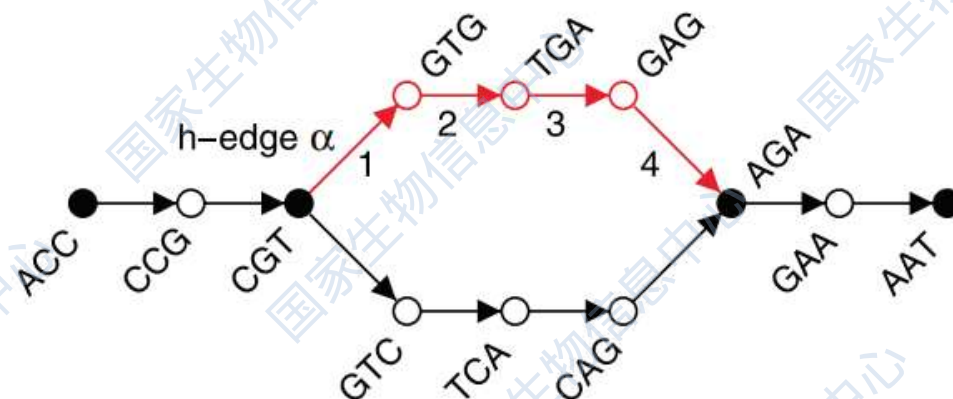
-k 21, 33, 55 (肺炎克雷伯菌)

35,45,55,65,75,85,95,105,115,125,127 (支原体)

--careful

--only-assembler

--cov-cutoff 'auto'



模块一：细菌基因组组装-二代数据**拼接质量评估**

- 拼接结果中contigs (> 1000 bp) 的数量应该小于100
- 组装后的基因大小total length, 如果与参考基因组的总长度有20%以上的差异就有问题
- N 's per 100 k bp该指标是指每100k bp碱基中所含有的N碱基数量, 值越高, 拼接结果越差
- 与参考基因组比对, coverage $\geq 95\%$
- 与参考基因组的平均核苷酸相似性 $\geq 94\%$
- 基因组的完整性和污染度: 完整性 $\geq 95\%$ 污染度 $\leq 5\%$

模块一：细菌基因组组装-二代数据拼接质量评估

- 拼接结果中contigs (> 1000 bp) 的数量应该小于100
- 组装后的基因大小total length, 如果与参考基因组的总长度有20%以上的差异就有问题
- N 's per 100 k bp该指标是指每100k bp碱基中所含有的N碱基数量, 值越高, 拼接结果越差

□ quast (<https://github.com/ablab/quast>)

```
quast.py contigs.fasta -r test_data/reference.fasta -1 reads1.fastq.gz -2 reads2.fastq.gz  
-o output_dir
```

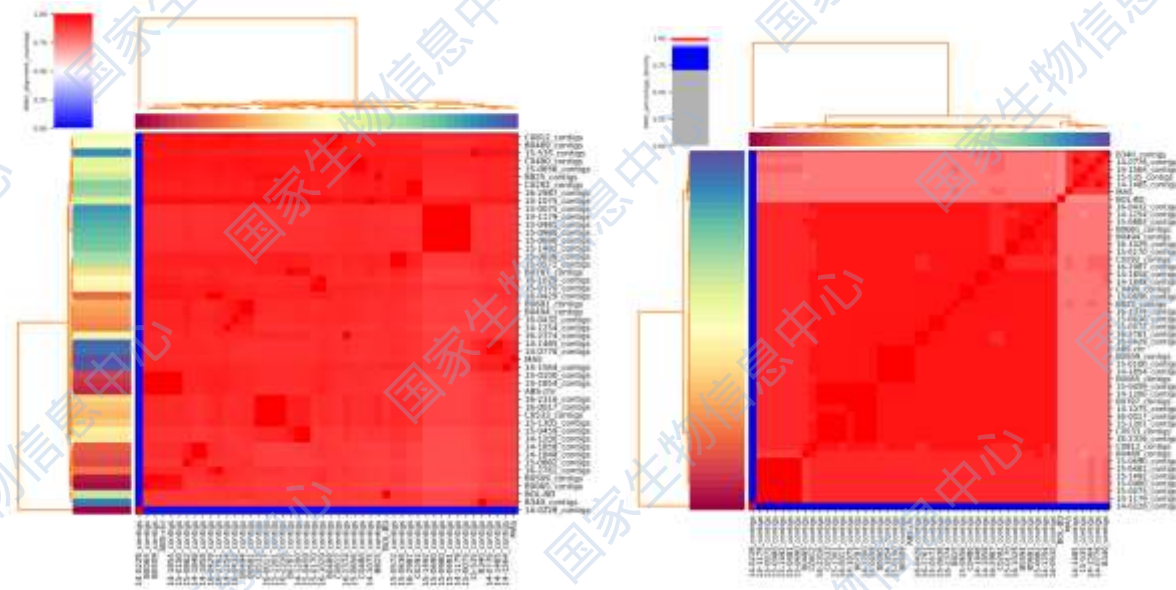
Assembly	contigs
# contigs (>= 0 bp)	312
# contigs (>= 1000 bp)	6
# contigs (>= 5000 bp)	5
# contigs (>= 10000 bp)	5
# contigs (>= 25000 bp)	4
# contigs (>= 50000 bp)	4
Total length (>= 0 bp)	909475
Total length (>= 1000 bp)	812243
Total length (>= 5000 bp)	808227
Total length (>= 10000 bp)	808227
Total length (>= 25000 bp)	785480
Total length (>= 50000 bp)	785480
# contigs	12
Largest contig	466237
Total length	815629
Reference length	816394
GC (%)	40.00
Reference GC (%)	40.01
N50	466237
NG50	466237
N90	67789
NG90	67789
auN	313822.5
auNG	313528.5
L50	1
LG50	1
L90	4
LG90	4
# N's per 100 kbp	0.00

模块一：细菌基因组组装-二代数据拼接质量评估

- 与参考基因组比对, coverage >= 95%
- 与参考基因组的平均核苷酸相似性 >= 94%
- 基因组平均遗传相似度(Average nucleotide identity, ANI)是指两细菌个基因组之间同源基因的相似性
- ANI 值可以通过两种运算方法得出：一种是以MUMmer 运算法则为基础(ANIm); 另一种是以BLASTn 方法为基础(ANId),
- ANI 具有方便、耗费工作量少、错误率低、分辨率高的优点
- Pyani (<https://github.com/widdowquinn/pyani>)

```
./average_nucleotide_identity.py -i tests/test_ani_data/ -o tests/test_ANIm_output -m ANIm -g
```

```
./average_nucleotide_identity.py -i tests/test_ani_data/ -o tests/test_ANId_output -m ANId -g
```



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
14-0228	1	0.842689	0.840449	0.840052	0.840239	0.840934	0.842559	0.842121	0.840091	0.839435	0.840091	0.840052	0.839435	0.840052
14-0776	0.842689	1	0.975518	0.976336	0.975452	0.975185	0.99727	0.989304	0.975846	0.976296	0.97584	0.976355	0.97628	0.97628
14-1075	0.840449	0.975518	1	0.990618	0.996201	0.993317	0.973631	0.975954	0.994664	0.994345	0.994668	0.990626	0.994349	0.994349
14-1179	0.840052	0.976336	0.990618	1	0.990639	0.989016	0.97526	0.977657	0.990156	0.990048	0.990153	0.999977	0.990045	0.990045
14-1200	0.840239	0.975452	0.996201	0.990639	1	0.993457	0.973698	0.976142	0.994522	0.994333	0.994521	0.990633	0.994329	0.994329
14-1254	0.840934	0.975185	0.993317	0.989016	0.993457	1	0.973247	0.975647	0.993306	0.992996	0.993303	0.98902	0.992997	0.992997
14-1485	0.842559	0.99727	0.973631	0.97526	0.973698	0.973247	1	0.990029	0.973457	0.973368	0.973457	0.975256	0.97335	0.97335
14-1564	0.842121	0.989304	0.975954	0.977657	0.976142	0.975647	0.990029	1	0.976304	0.976042	0.976297	0.977631	0.976025	0.976025
14-1848	0.840091	0.975846	0.994664	0.990156	0.994522	0.993308	0.973457	0.976304	1	0.994602	0.99999	0.990161	0.994586	0.994586
14-1854	0.839435	0.976296	0.994345	0.990048	0.994333	0.992996	0.973368	0.976042	0.994602	1	0.994583	0.990078	0.999997	0.999997
14-1858	0.840091	0.97584	0.994668	0.990153	0.994521	0.993303	0.973457	0.976297	0.99999	0.994583	1	0.990139	0.994581	0.994581
15-0075	0.840052	0.976355	0.990626	0.999977	0.990633	0.98902	0.975256	0.977631	0.990161	0.990078	0.990139	1	0.990053	0.990053
15-0100	0.839435	0.97628	0.994349	0.990045	0.994329	0.992997	0.97335	0.976025	0.994586	0.999997	0.994581	0.990053	1	0.990053
15-0170	0.839955	0.975356	0.993731	0.989656	0.993973	0.99195	0.973245	0.976009	0.99332	0.993128	0.993297	0.989665	0.993163	0.993163
15-0459	0.83997	0.975452	0.996198	0.990635	0.999994	0.993394	0.973584	0.976095	0.994497	0.994295	0.99449	0.990655	0.994284	0.994284

模块一：细菌基因组组装-二代数据拼接质量评估

□ 基因组的完整性和污染度：完整性 $\geq 95\%$ 污染度 $\leq 5\%$

□ checkM (<https://github.com/chklovski/CheckM2>)

checkm lineage_wf -t 16 --pplacer_threads 16 --tab_table -f checkm.txt -x fa contig_dir output

Bin_Id	Marker_lineage	#genomes	#markers	#marker_sets	0	1	2	3	4	5+	Completeness	Contamination	Strain heterogeneity
NY10810	o_Pseudomonadales (UID448	185	813	308	1	810	2	0	0	0	99.68	0.14	0
NY10816	o_Pseudomonadales (UID448	185	813	308	1	811	1	0	0	0	99.68	0.11	0
NY10820	o_Pseudomonadales (UID448	185	813	308	1	811	1	0	0	0	99.68	0.11	0
NY10823	o_Pseudomonadales (UID448	185	813	308	1	803	9	0	0	0	99.68	1.57	0
NY10825	o_Pseudomonadales (UID448	185	813	308	1	811	1	0	0	0	99.68	0.11	0
NY10829	o_Pseudomonadales (UID448	185	813	308	1	809	3	0	0	0	99.68	0.46	0
NY10839	o_Pseudomonadales (UID448	185	813	308	1	810	2	0	0	0	99.68	0.14	0
NY10840	o_Pseudomonadales (UID448	185	813	308	1	810	2	0	0	0	99.68	0.22	50
NY10847	o_Pseudomonadales (UID448	185	813	308	1	811	1	0	0	0	99.68	0.11	0
NY10876	o_Pseudomonadales (UID448	185	813	308	1	802	10	0	0	0	99.68	1.59	0
NY10878	o_Pseudomonadales (UID448	185	813	308	1	802	10	0	0	0	99.68	1.59	0
NY10894	o_Pseudomonadales (UID448	185	813	308	1	810	2	0	0	0	99.68	0.16	0
NY10898	o_Pseudomonadales (UID448	185	813	308	1	802	10	0	0	0	99.68	1.59	0
NY10975	o_Pseudomonadales (UID448	185	813	308	1	811	1	0	0	0	99.68	0.11	0
NY10992	o_Pseudomonadales (UID448	185	813	308	1	809	3	0	0	0	99.68	0.25	0
NY10998	o_Pseudomonadales (UID448	185	813	308	1	810	2	0	0	0	99.68	0.43	0
NY10999	o_Pseudomonadales (UID448	185	813	308	1	810	2	0	0	0	99.68	0.43	0
NY11000	o_Pseudomonadales (UID448	185	813	308	1	811	1	0	0	0	99.68	0.11	0
NY11002	o_Pseudomonadales (UID448	185	813	308	1	810	2	0	0	0	99.68	0.43	0
NY11040	o_Pseudomonadales (UID448	185	813	308	1	805	7	0	0	0	99.68	0.67	14.29
NY11052	o_Pseudomonadales (UID448	185	813	308	1	808	4	0	0	0	99.68	0.51	0

模块一：细菌基因组组装-二代+三代数据

二代短reads+三代长reads混拼

❑ Unicycler (<https://github.com/rrwick/Unicycler>)

```
unicycler -1 reads_1.fastq.gz -2 reads_2.fastq.gz -l nano.fastq.gz --mode bold --min_fasta_length 1000 -o outputdir
```

>1 length=5477879 depth=1.00x circular=true

>2 length=279711 depth=1.58x circular=true

>3 length=139521 depth=1.42x circular=true

>4 length=115305 depth=2.54x circular=true

>5 length=43474 depth=0.45x circular=true

>6 length=10060 depth=0.64x circular=true

三代长reads拼接，二代短reads进行矫正

1、三代数据拼接

❑ Flye (<https://github.com/fenderglass/Flye>)

```
flye --nano-raw nano.fq.gz -g 5.3m -t 12 -o dir
```

❑ canu (<https://canu.readthedocs.io/en/latest/>)

```
canu -d canu -p canu genomeSize=5.3m  
maxThreads=12 -nanopore-raw nano.fq.gz
```

2、二代数据矫正

❑ Pilon

```
java -Xmx16G -jar pilon-1.13.jar --genome  
assembly.fasta --frags reads_sorted.bam --output dir --  
changes --vcf --vcfqe --tracks --threads 4 --minmq 20 --  
minqual 25
```

3、人工环化

拼接的序列首尾截取不同长度的序列进行比对，找到相同的区域，写脚本，进行环化

基因预测: Prokka (<https://github.com/tseemann/prokka>)

Output Files

-gcode 默认是11，针对细菌

-gcode 4

[illegible]

Differences from the Standard Code:

Code	Standard
USA Try #	Far =

Alternative Initiation Codons:

Trypanosoma: UUA, UUG, CUG
Leishmania: AUU, AUA

<https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi#SG11>

模块二：细菌基因组基因预测和注释

□ tRNA预测：tRNAscan-SE (参数：-B -l -m lsu, ssu, tsu)

□ rRNA预测：RNAmmer (参数：-S bac)

□ 基因功能注释：

✓ **EggNOG-mapper** (<https://github.com/eggnogdb/eggno-mapper>) COG, GO和KEGG的注释

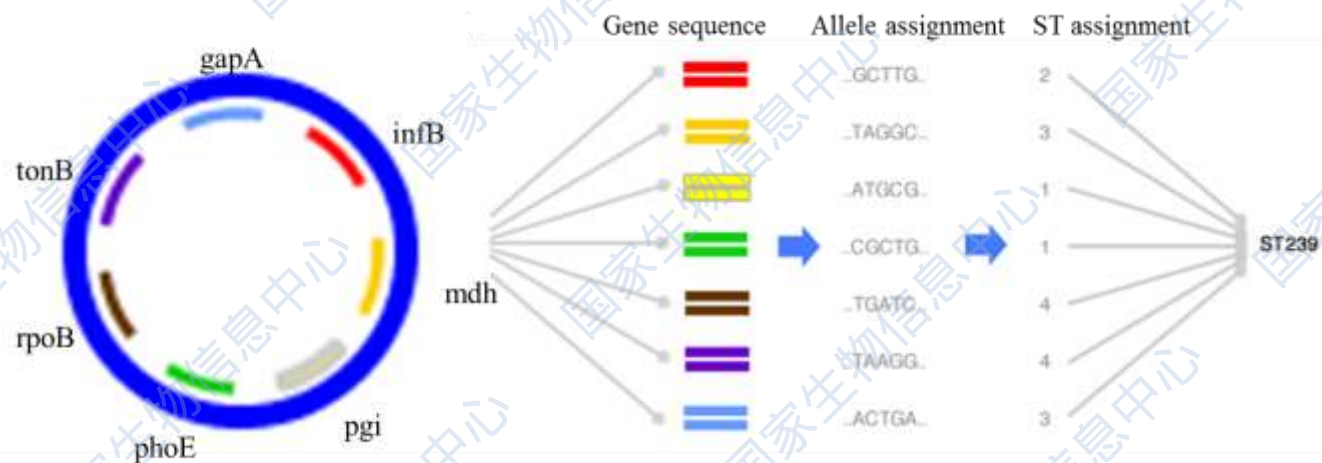
query	seed	ortvalue	score	eggNOG	Ortho	name	COG	category	Descript	Preferre	GOs	EC	KEGG	hs	KEGG	Path	KEGG	Mod	KEGG	Reac	KEGG	rel	BR	TE	KEGG	TC	CAZy	BiGG	Reas	PFAMs
30264_001722438	M10.0	1927.0		2A2P001	544448	Ti	-	-	crnA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	MspC
158_00271722438	M12.06e-30	842.0		COG22688	2	Bacter	T	-	Band 7 ps2012_042	-	-	-	ko:K0719	ko:K04910	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Band 7,
323_0073722438	M17.87e-36	201.0		2B50801	544448	Ti	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ICF3713
7461_001343273	M19.62e-78	252.0		2E23T01	544448	TiS	-	-	Peptidas	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ICF31
10319_001722438	M11.8e-65	204.0		COG4970	2	Bacter	L	-	DNA recod	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ICF16
10319_001722438	M11.36e-60	60.8		2B66901	544448	TiS	-	-	hms p001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Lipopro
366_0030722438	M10.0	906.0		COG49420	2	Bacter	D	-	peptidas	rnz	-	-	ko:K1208	ko:K03018	1W00364	-	-	-	-	-	-	-	-	-	-	-	-	-	-	AAA, DUF
332_00271722438	M12.72e-35	123.0		COG33310	544448	TiL	-	-	Endonuc	recU	-	-	ko:K03704	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	RecU
30107_001722438	M11.38e-20	593.0		COG49420	2	Bacter	D	-	peptidas	rnz	-	-	ko:K1208	ko:K03018	1W00364	-	-	-	-	-	-	-	-	-	-	-	-	-	-	AAA, DUF
10319_001722438	M11.82e-10	313.0		COG4970	2	Bacter	L	-	hms p001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	AAA, DUF
10319_001722438	M18.7e-38	111.0		COG4970	2	Bacter	L	-	DNA recod	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ICF16
3262_005722438	M18.51e-31	885.0		COG07470	544448	TiE	-	-	Family or	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ICF3396
10319_001722438	M12.73e-59	188.0		COG4970	2	Bacter	L	-	hms p001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ICF16
10319_001722438	M13.13e-20	598.0		COG07470	544448	TiE	-	-	Family or	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ICF3396
10319_001722438	M14.69e-50	166.0		COG4970	2	Bacter	L	-	hms p001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ICF16
30298_001722438	M10.0	1247.0		COG10530	544448	TiG	-	-	hms p001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Lipopro
30278_001722438	M11.01e-19	593.0		2B2VM01	544448	Ti	-	-	gapA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Adhesin
361_00371722438	M17.87e-16	300.0		2B2VM01	544448	Ti	-	-	gapA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Adhesin
736_0073722438	M17.37e-69	221.0		COG10530	2	Bacter	G	-	carbohyd	-	-	-	5.2.1.3	ko:K03701	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Flg_cow
158_00271722438	M12.74e-12	369.0		COG22688	2	Bacter	T	-	Band 7 ps2012_042	-	-	-	ko:K0719	ko:K04910	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Band 7,
7536_001722438	M14.11e-19	554.0		COG22688	544448	TiS	-	-	hms p001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ICF237
332_00401722438	M11.5e-27	629.0		COG06100	544448	TiL	-	-	Subunit I hsdX	-	-	-	3.1.21.3	ko:K0115	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	EcuR124
30267_001722438	M12.08e-30	884.0		2B50801	544448	Ti	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ICF3713

✓ **blastp** (参数：-evalue 1e-5 -outfmt 6 -max_target_seqs 5) 比对编码蛋白到GO、COG和KEGG数据库，保留比对coverage大于30%的最好结果作为注释结果

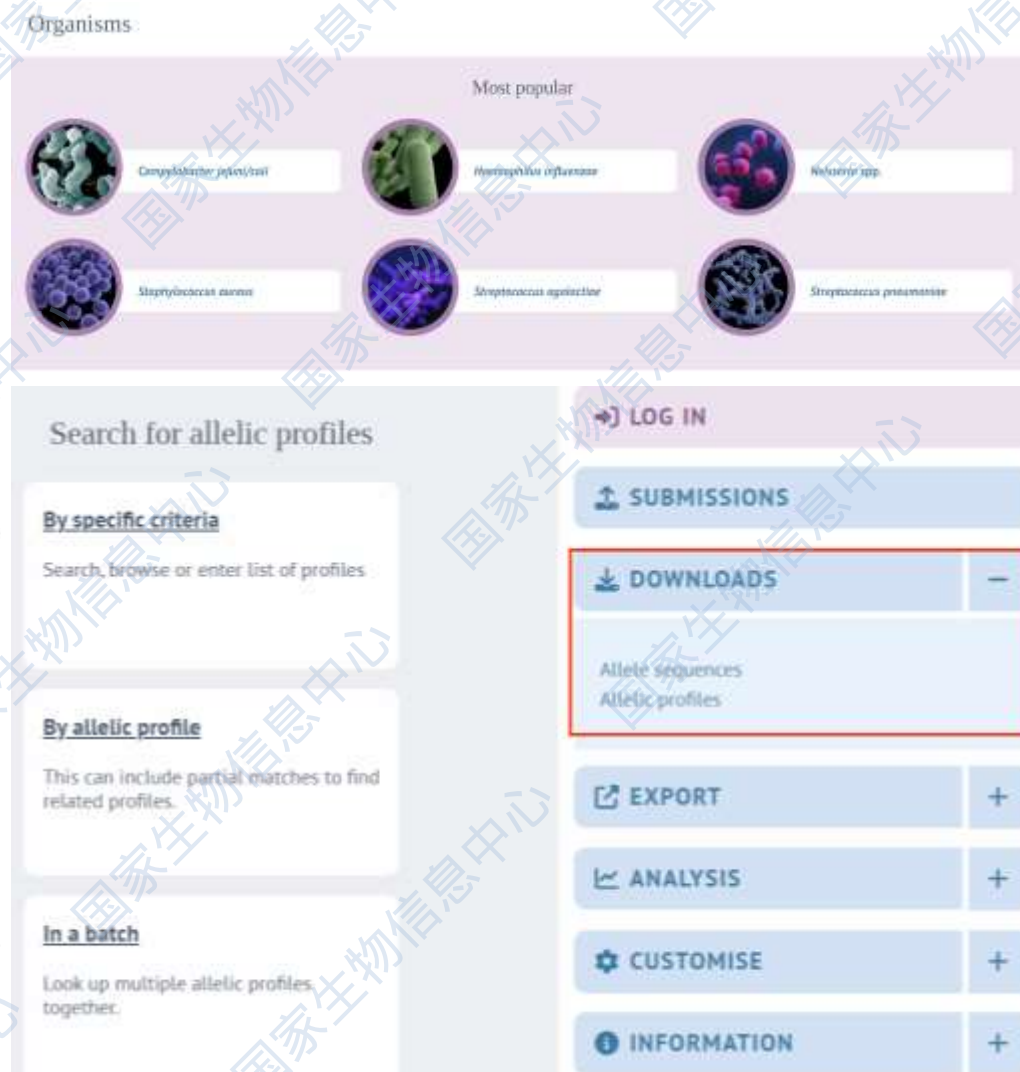
模块三：细菌基因组分子分型-ST分型

多位点序列分型 (multilocus sequence typing, MLST)

- 是一种基于核酸序列测定的细菌分型方法
- 这种方法通过多个管家基因的序列, 分析菌株的变异
- MLST操作简单, 逐渐成为细菌的常规分型方法



序列以及已知分型可以在PubMLST获得



模块三：细菌基因组分子分型-ST分型

在线分析方法：https://pubmlst.org/bigodb?db=pubmlst_spneumoniae_seqdef

Query a sequence

Single sequence

Query a single sequence or whole genome assembly to identify allelic matches.

Batch sequences

Query multiple independent sequences in FASTA format to identify allelic matches.

Sequence query

Please paste in your sequence to query against the database. Query sequences will be checked first for an exact match against the chosen (or all) loci - they do not need to be trimmed. The nearest partial matches will be identified if an exact match is not found. You can query using either DNA or peptide sequences. ①

Please select locus/scheme

Order results by

All loci

locus

Enter query sequence (single or multiple contigs up to whole genome in size)

Alternatively upload FASTA file

Action

Select FASTA file: ①

Click to select or drag and drop...

RESET

SUBMIT

Batch sequence query

Please paste in your sequences to query against the database. Query sequences will be checked first for an exact match against the chosen (or all) loci - they do not need to be trimmed. The nearest partial matches will be identified if an exact match is not found. You can query using either DNA or peptide sequences. ①

Please select locus/scheme

Order results by

All loci

locus

Enter query sequences (FASTA format)

Alternatively upload FASTA file

Action

Select FASTA file: ①

Click to select or drag and drop...

RESET

SUBMIT

模块三：细菌基因组分子分型-ST分型

本地分析方法

□ 基于拼接之后的序列：

mlst (<https://github.com/tseemann/mlst>) 简单快速

```
> mlst contigs.fa
```

```
contigs.fa neisseria 11149 abcZ(672) adk(3) aroE(4) fumC(3) gdh(8) pdhC(4) pgm(6)
```

□ 基于二代测序的reads：

SRST2 (<https://github.com/katholt/srst2>) 准确，可以提供变异的情况，以便鉴定提交新的ST型别序列

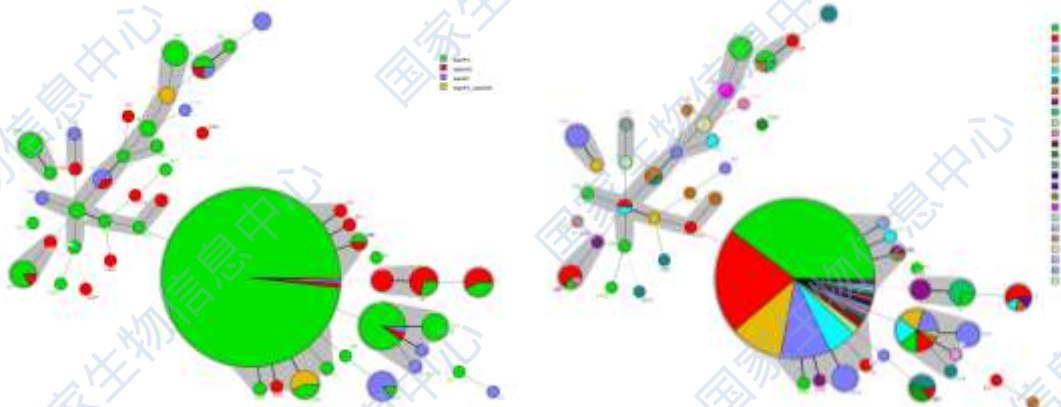
```
srst2 --input_pe strainA_1.fastq.gz strainA_2.fastq.gz --output strainA_test --log --mlst_db  
Escherichia_coli#1.fasta --mlst_definitions profiles_csv --mlst_delimiter _
```

Sample	ST	adk	fumC	gyrB	icd	mdh	purA	recA	Mis matches	uncertainty	depth
strainA	152	11	63	7	1	14	7	7	0	-	25.8319 955826

模块三：细菌基因组分子分型-ST分型可视化

BioNum

Strain	gapA	infB	mdh	pgi	phoE	rpoB	tonB	Province	Hospital	Year	Res
G001	2	1	2	1	3	4	25	Hebei	H27	2014	blaKPC
G006	3	3	1	1	1	1	4	Jiangsu	H42	2015	blaKPC
G007	3	3	1	1	1	1	4	Zhejiang	H50	2012	blaKPC
G008	3	3	1	1	1	1	4	Zhejiang	H50	2009	blaKPC
G009	3	3	1	1	1	1	4	Zhejiang	H50	2009	blaKPC
G010	1	1	1	1	1	1	1	Zhejiang	H50	2013	blaKPC
G011	1	1	1	1	1	1	1	Zhejiang	H50	2013	blaKPC
G012	3	3	1	1	1	1	4	Hebei	H28	2013	blaKPC
G013	3	3	1	1	1	1	4	Hebei	H28	2013	blaKPC
G014	3	3	1	1	1	1	4	Henan	H74	2015	blaKPC
G015	3	3	1	1	1	1	4	Chongqin	H76	2016	blaKPC
G016	3	3	1	1	1	1	4	Chongqin	H76	2014	blaKPC
G017	3	3	1	1	1	1	4	Fujian	H13	2015	blaKPC
G018	3	3	1	1	1	1	4	Fujian	H13	2015	blaKPC
G019	3	3	1	1	1	1	4	Fujian	H13	2015	blaKPC



PHYLOViZ

<https://online.phyloviz.net/index>

Upload your own data

NOTE: Without registration your data will not be saved and erased 24 hours after uploading.

Select one of the possible input formats

Possible Input Formats

Profile Data

Upload Profile Data

Browse

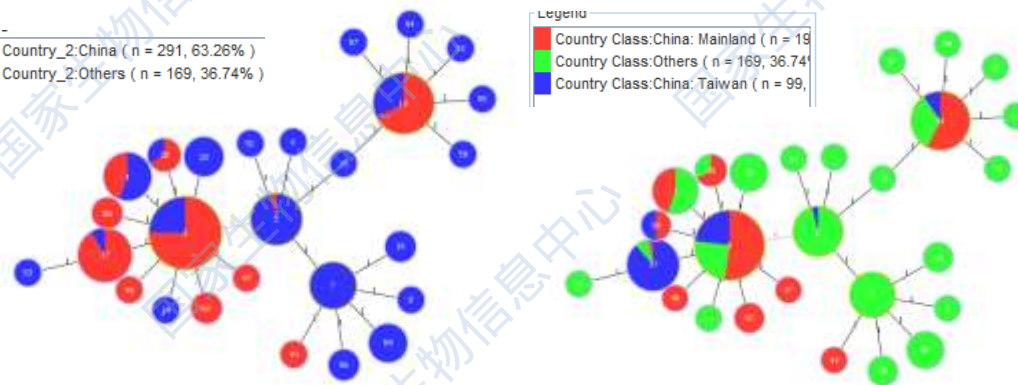
Upload Auxiliary Data

Browse

Country_2:China (n = 291, 63.26%)
Country_2:Others (n = 169, 36.74%)

Legend

Country Class:China: Mainland (n = 19)
Country Class:Others (n = 169, 36.74%)
Country Class:China: Taiwan (n = 99)



模块四：细菌基因组进化分析

SNP calling



- Snippy
- Mummer

SNP filtering

- 去除InDels
- 去除重组区域
- 去除Repeat区域



- Gubbins预测重组区域
- RepeatMasker预测重复区域

建树



- RaxML
- iqtree

模块四：细菌基因组进化分析

SNP calling

Snippy

- ❑ `snippy --cpus 4 --outdir dir --ref ref_genomic.fna --ctgs contigs.fasta`
- ❑ `snippy-core --ref ref_genomic.fna --prefix core_total snippy*`
- ❑ `snippy-clean_full_aln core.full.aln > clean.full.aln`

Mummer

- ❑ `nucmer -p ref_to_NJST258_1 NJST258_1.fasta ref.fasta`
- ❑ `delta-filter -r -q ref_to_NJST258_1.delta > ref_to_NJST258_1.filter`
- ❑ `show-snps -ClrT ref_to_NJST258_1.filter > ref_to_NJST258_1.snps`

模块四：细菌基因组进化分析

SNP filtering

- ❑ `run_gubbins.py --prefix gubbins_output clean_total.full.aln`
- ❑ `RepeatMasker -pa 4 -species bacteria -poly -html -dir ./3.repeatmasker reference.fasta`

建树

❑ RaxML

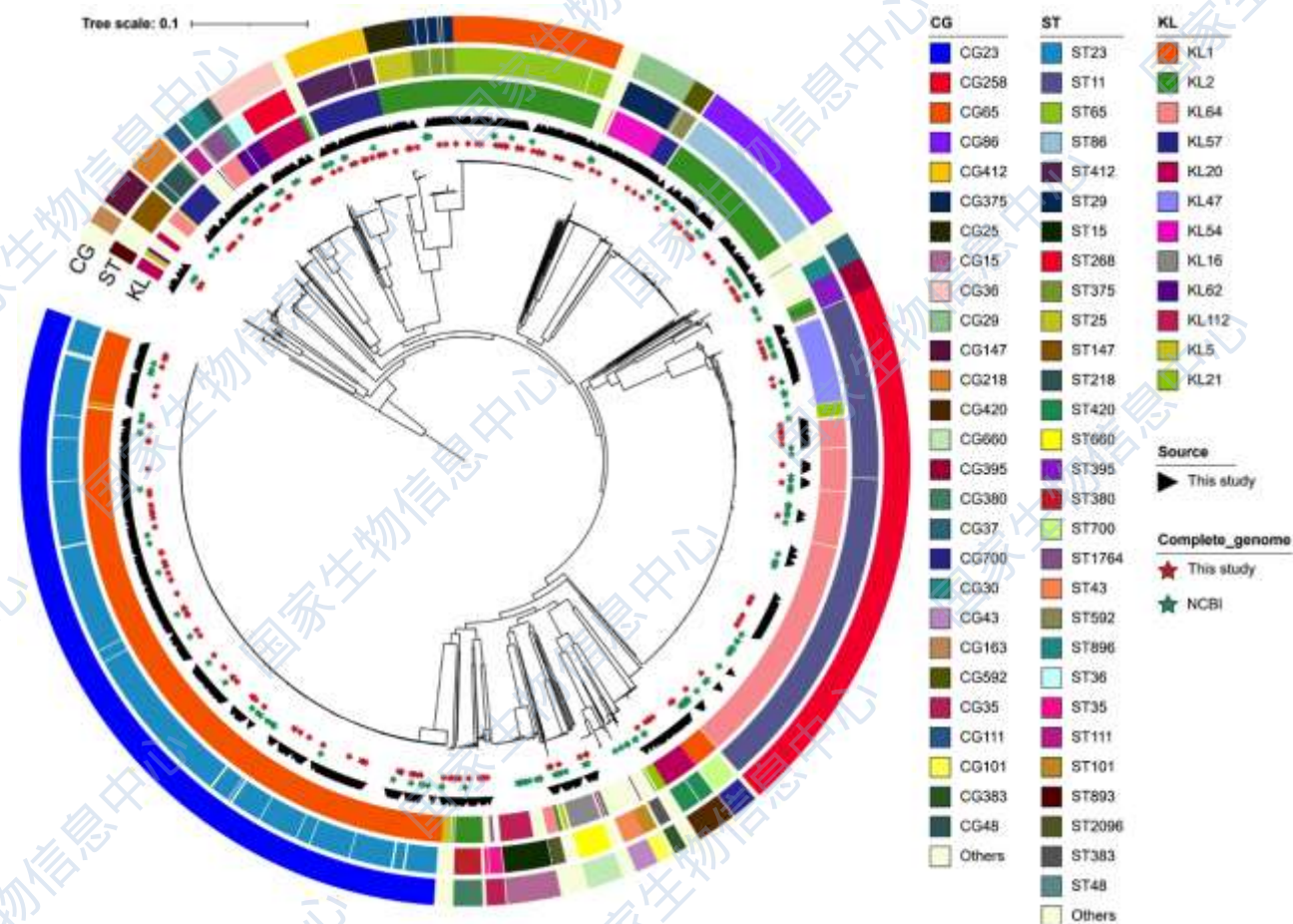
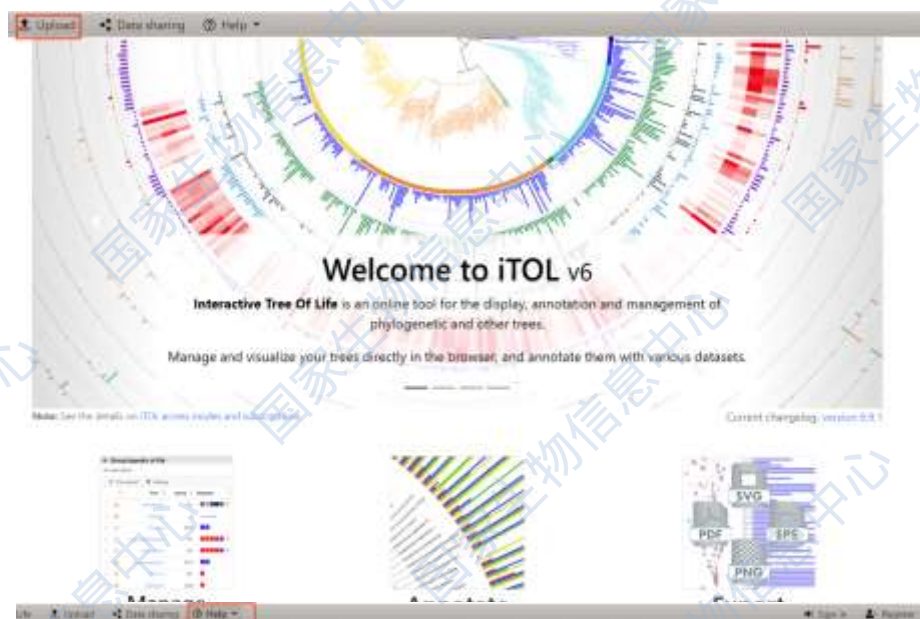
1. `raxmlHPC -m GTRGAMMA -p 12345 -# 20 -s dna.phy -n T1`
2. `raxmlHPC -x 12345 -p 12345 -# 1000 -m GTRGAMMA -k -s dna.phy -n T2`
3. `raxmlHPC -p 12345 -m GTRGAMMA -f b -t RAXML_bestTree.T1 -z RAXML_bootstrap.T2 -n T3`

❑ Iqtree

`iqtree -s aln.fasta -B 1000 -m GTR+I+G -T AUTO`

模块四：细菌基因组进化分析-进化树美化

itol:<https://itol.embl.de/>



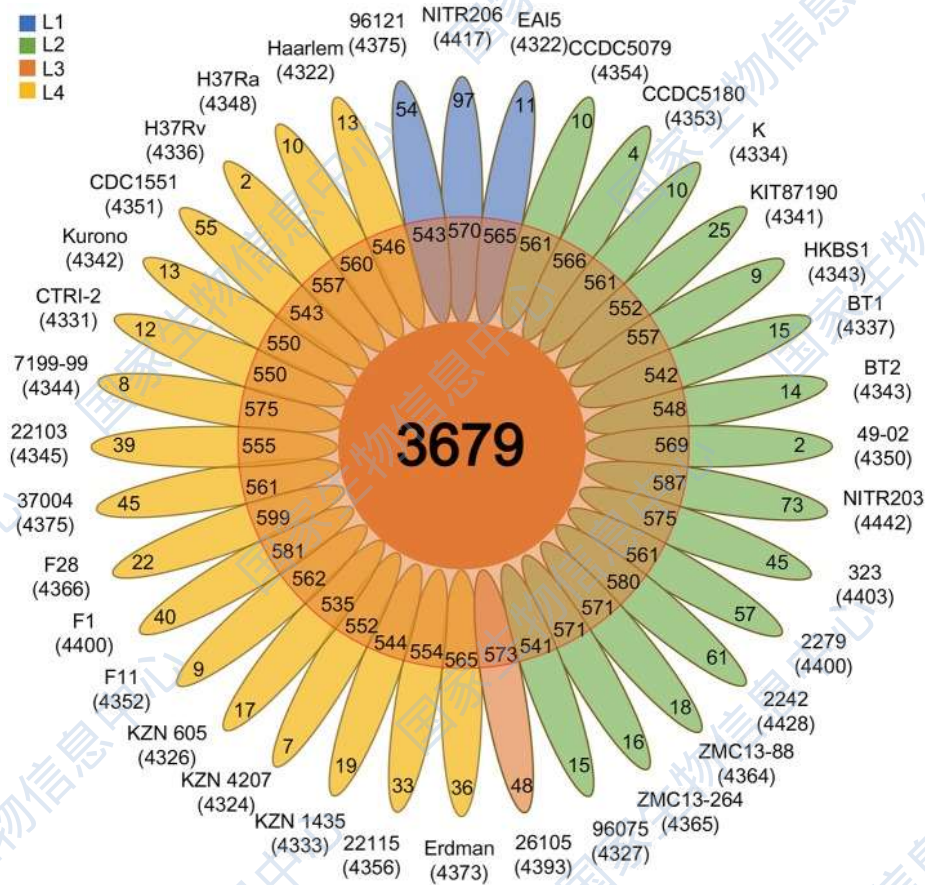
Viewer user interface

An overview of the most parts of iTOOL's tree viewer user interface is shown below.

- Main tree display:** Change the zoom level with your mouse wheel. To move pan the tree, simply click and drag with your mouse. You can also use the mouse wheel to pan the tree. Keep the [MM] key pressed while scrolling the wheel to pan horizontally. At the [MM] key to pan vertically. If you simultaneously hold the [MM] and [Ctrl] keys, the panning speed will increase.
- Zoom controls:** In addition to the mouse wheel, zoom keys can be used to change the tree zoom level. Holding the [Q] key will fit the tree to the parent window.
- Other functions:** Hover over the [Q] key to find out more information about the current tree. [Q] can be used to quickly find a specific node. Click the [Q] key to open the [Node information] panel.
- Main control panel:** The majority of iTOOL's controls and functions are available here, split over a tab: Basic, Advanced, Datasets and Export.
- Node functions menu:** Click on any tree branch or leaf node to display the available functions and interactive properties (during runtime).
- Datasets legend box:** If any datasets are present in the tree, they will be listed here. You can toggle the display of individual datasets across their options by clicking the [On] icon or dataset from its corresponding color-coded box and clicking the [On] icon.
- Colorful ranges legend box:** If any colorful ranges are present in the tree, they will be listed here. You can toggle the range display modes, across their options by clicking the [On] icon or adding them by hovering over the colored box and clicking the [On] icon.

Click the [H] icon in the control panel header to set the default position of dataset and interface panels. Toggle the display of extra user interface help icons and tree help panels.

模块五：细菌泛基因组分析



□ Core genes

□ Dispensable genes

□ Unique genes

□ Roary (<https://sanger-pathogens.github.io/Roary/>)

获得整体基因的情况

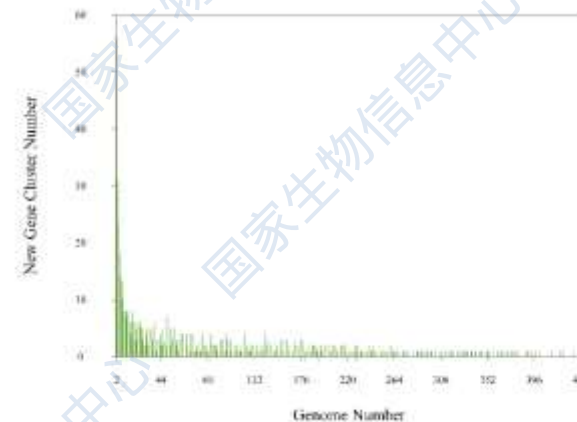
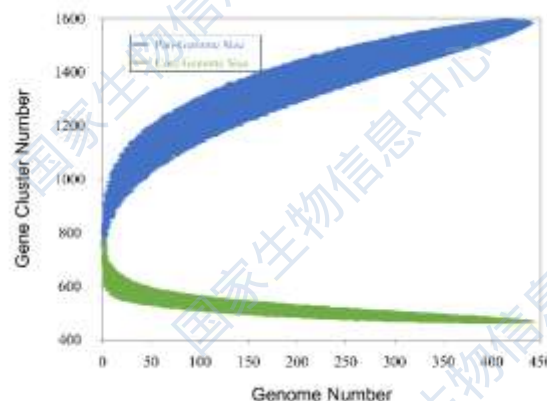
gene_presence_absence.csv

gene_presence_absence.pdf

gene_presence_absence.Rtab

Strain	Gene	Present	Absent	Core	Accessory	Strain	Gene	Present	Absent	Core	Accessory
H37Ra	Gene1	1	0	1	0	H37Rv	Gene1	1	0	1	0
H37Ra	Gene2	1	0	1	0	H37Rv	Gene2	1	0	1	0
H37Ra	Gene3	1	0	1	0	H37Rv	Gene3	1	0	1	0
H37Ra	Gene4	1	0	1	0	H37Rv	Gene4	1	0	1	0
H37Ra	Gene5	1	0	1	0	H37Rv	Gene5	1	0	1	0

□ PanGP (有Windows版本) 可视化泛基因组曲线



□ R包 “micropan” 计算是否闭合

$\alpha > 1$ 闭合

模块六：细菌基因组耐药和毒力分析

耐药和毒力数据库

- ❑ 耐药基因数据库：CARD, Resfinder, NCBI AMRFinderPlus, ARG-ANNOT等
- ❑ 毒力基因数据库：VFDB

分析方法

- ❑ **Blast**: blastp or blastn
- ❑ **ABRicate**
(<https://github.com/tseemann/abricate?tab=readme-ov-file>), 包含多种耐药和毒力基因的数据库, 并且可以指定使用哪一个数据库
- ❑ **AMRFinderPlus** (<https://github.com/ncbi/amr>), 同时可以解析突变导致的耐药

Organism option	Point mutations	Blacklisted genes	Notes
Acinetobacter_baumannii	X		Use for the <i>A. baumannii</i> - <i>calcoaceticus</i> species complex
Burkholderia_cepacia	X		Use for the <i>Burkholderia cepacia</i> species complex
Burkholderia_mallei	X		
Burkholderia_pseudomallei	X		Use for the <i>Burkholderia pseudomallei</i> species complex
Campylobacter	X		Use for <i>C. coli</i> and <i>C. jejuni</i>
Citrobacter_freundii	X		
Corynebacterium_diphtheriae	X		
Clostridioides_difficile	X		
Enterobacter_cloacae	X		
Enterobacter_asburiae	X		
Enterococcus_faecalis	X		
Enterococcus_faecium	X		Use for <i>E. hirae</i>
Escherichia	X	X	Use for <i>Shigella</i> and <i>Escherichia</i>
Klebsiella_oxytoca	X	X	
Klebsiella_pneumoniae	X	X	Use for <i>K. pneumoniae</i> species complex and <i>K. aerogenes</i>

内容提纲

1

细菌基因组分析的流程和经验

2

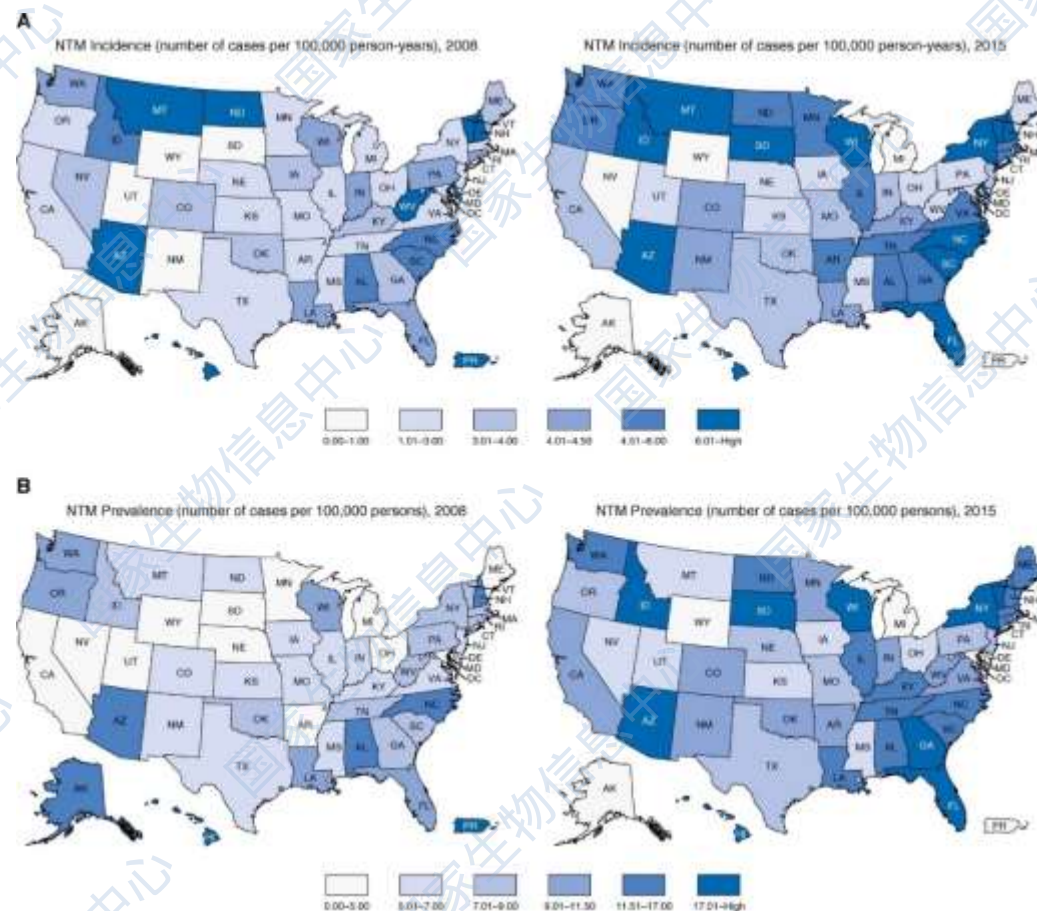
NTM基因组分析、数据资源介绍及应用案例

非结核分枝杆菌感染逐年上升且危害严重

非结核分枝杆菌 (nontuberculous mycobacteria, NTM) 是分枝杆菌属内除MTBC和麻风分枝杆菌以外的其他分枝杆菌。



- **可引起多组织感染：**多侵犯肺组织，此外还可侵犯肺外组织器官，引起角膜炎、心内膜炎、脑膜炎等，重症患者可引起全身播散性疾病、菌血症
- **可引发院内爆发感染**
- **感染率逐年上升：**2008年至2015年，NTM肺病年发病率由3.13 / 10万人上升至4.73 / 10万人，年患病率由6.78 / 10万人上升至11.70 / 10万人



NTM物种多样，临床表现不一

Rapid
growing

M. abscessus

M. chelonae

M. smegmatis

95 others

Slow
growing

M. avium complex

M. ulcerans

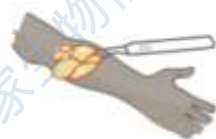
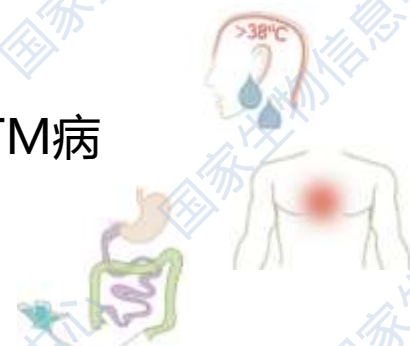
M. kansasii

84 others



肺部感染

NTM肺病——最常见的NTM病
常伴有胸痛、发烧、腹泻
以及体重减轻



皮肤，软组织感染

往往发生在**针刺伤口**或**开放性伤口**
或骨折处
局部皮肤发红、肿痛和硬结



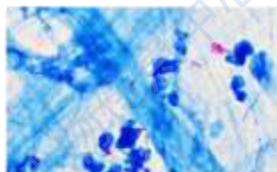
淋巴结感染和散播性感染

淋巴结感染常见于**儿童**
散播性感染主要见于HIV患者和**免疫抑制人群**

NTM感染诊断困难

现有诊断方法

痰涂片染色



影像学



病理学



Mantoux皮肤试验



难以区分

NTM or 结核感染?



◆ ~30%的 NTM感染患者在治疗开始时被误诊为耐多药结核病

无法鉴定

是哪种NTM感染?

五类常见的致病性NTMs (临床分离率 > 80%)



◆ 16S rRNA, rpoB, hsp65 和gyrA/gyrB等基因可以在一定程度上快速检测NTM的菌株

◆ 但由于不同NTM物种之间的高度同源性, 不同靶基因的基因检测结果往往不一致

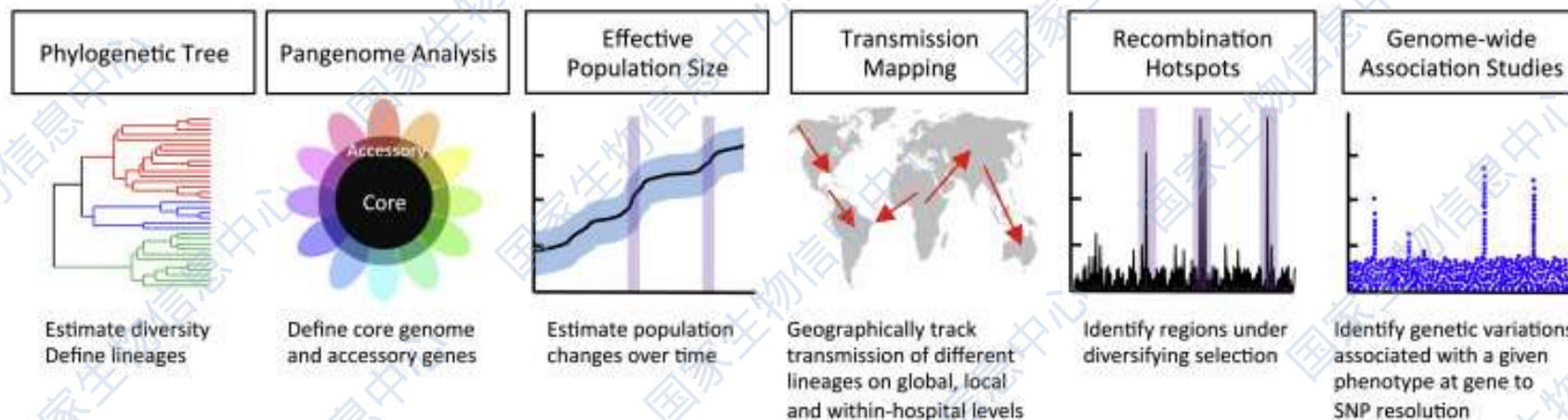
大规模的基因组数据为NTM研究提供重要资源

Whole genome sequencing of single isolates

Assembled reference genome

Gene content
Genome organization

Comparative genomic studies with multiple isolates



- 基因组流行病学
- 基因组特征
- 进化
- 传播
- 种群结构
- GWAS
- 耐药和毒力

当前，NTM基因组数据分散

Mycobrowser



- 仅包含3个结核基因组以及7个NTM基因组
- 仅进行了基因预测和基因注释

MicroScope



- 仅收集了154株NTM细菌基因组
- 提供了单个菌株之间的比较基因组分析在线分析工具

BV-BRC



- 收集了大量细菌基因组，但NTM物种数较少
- 没有进一步的整合分析

系统性NTM基因组数据库构建需求迫切

收集的物种不全



数据分析不深入



没有足够的可视化



现有数据库
的不足



缺少分析工具



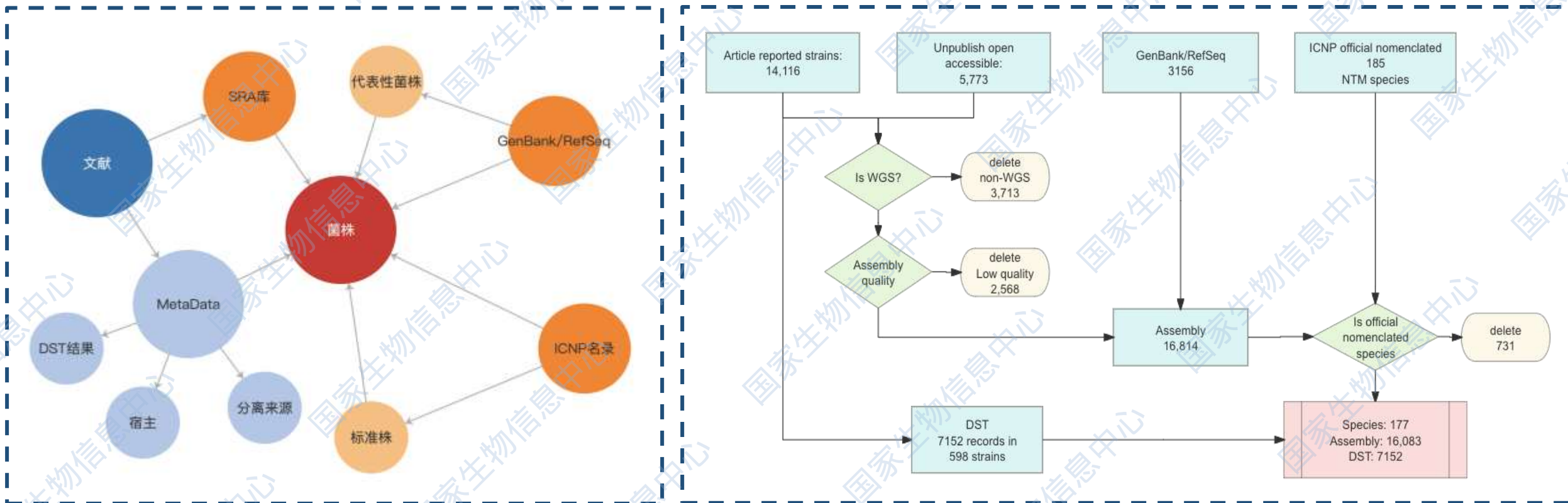
缺少物种层面的
整合分析



很多数据库已经
不维护

基因组学研究 + 数据库构建 → 整合零散信息、探究机制、提供资源

全面收集NTM基因组数据16000+, 177个物种



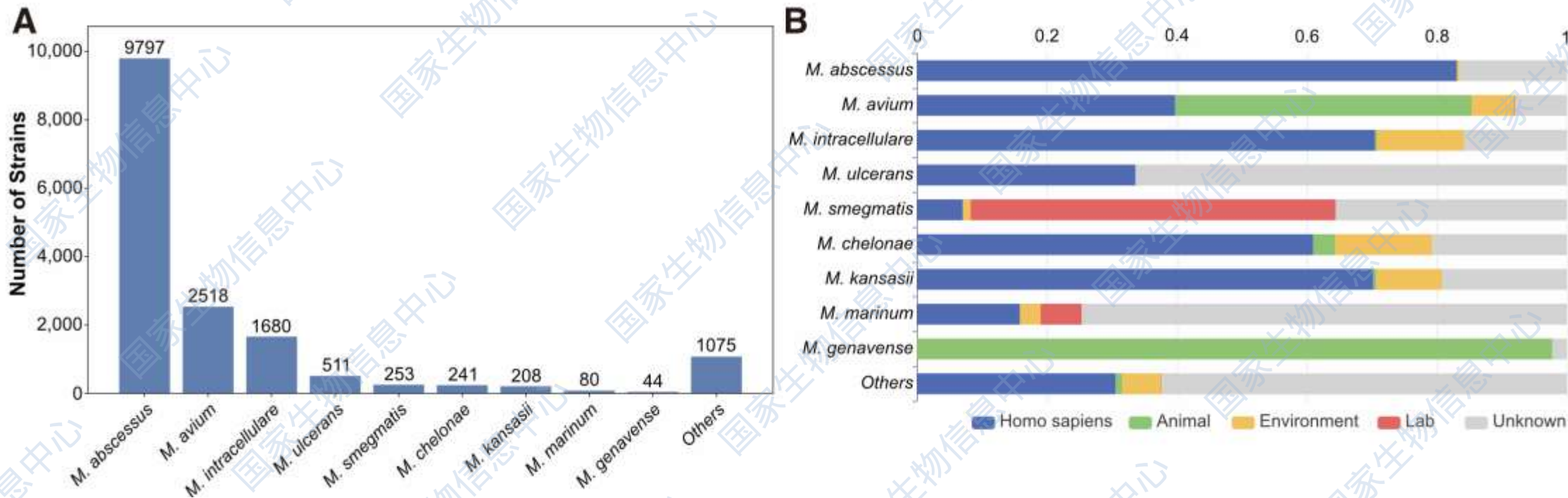
□ NCBI Genbank: 3156条

□ 文献审编: 14,116条

□ SRA原始二代测序数据: 5773条

□ ICNP: 185条

菌株的物种分布特征



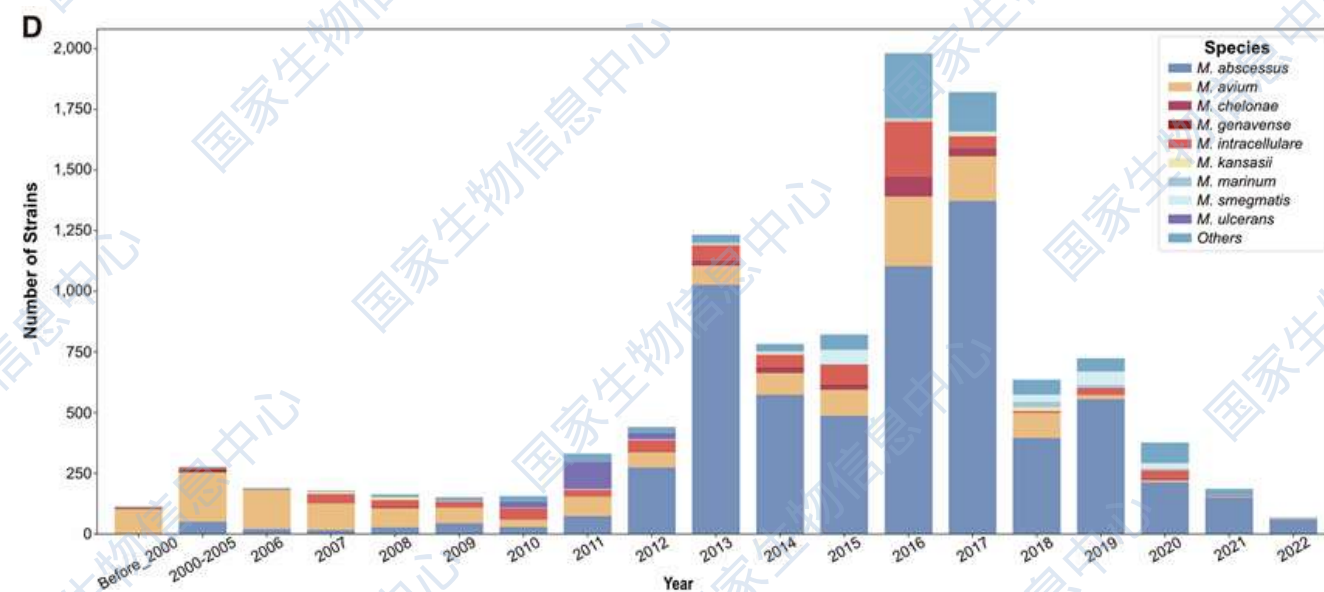
共纳入16,469个组装好的NTM全基因组

*M. abscessus*是最常见的物种，其次是*M. avium*和*M. intracellulare*

*M. abscessus*和*M. intracellulare*表现出对人类感染的偏好，分别有83%和70%的菌株是人类来源

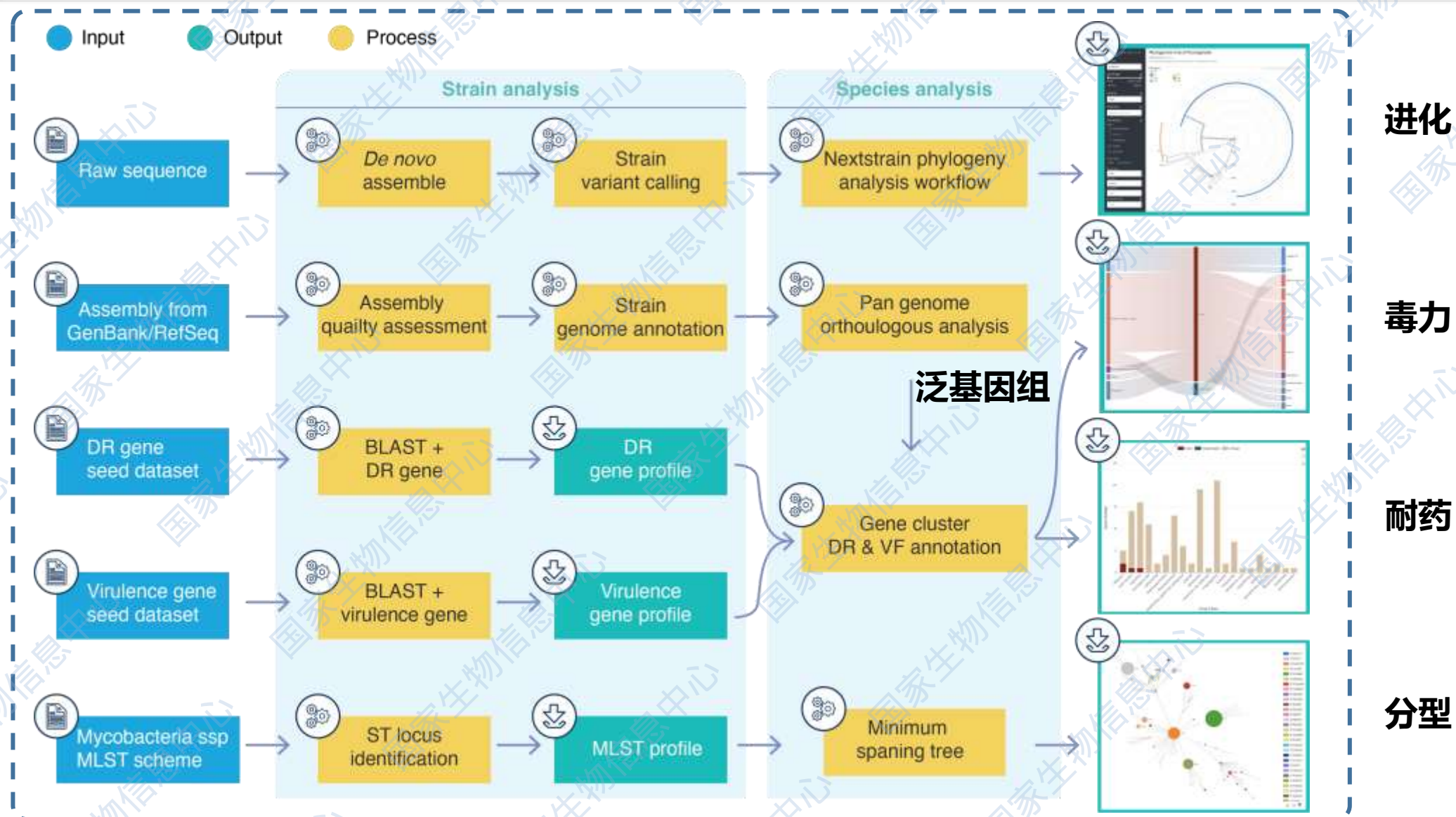
尽管*M. avium*和*M. intracellulare*具有密切的遗传关系，*M. avium*展示了不同的模式，其45.6%的菌株来自动物，39.8%来自人

NTM菌株的时空分布特征



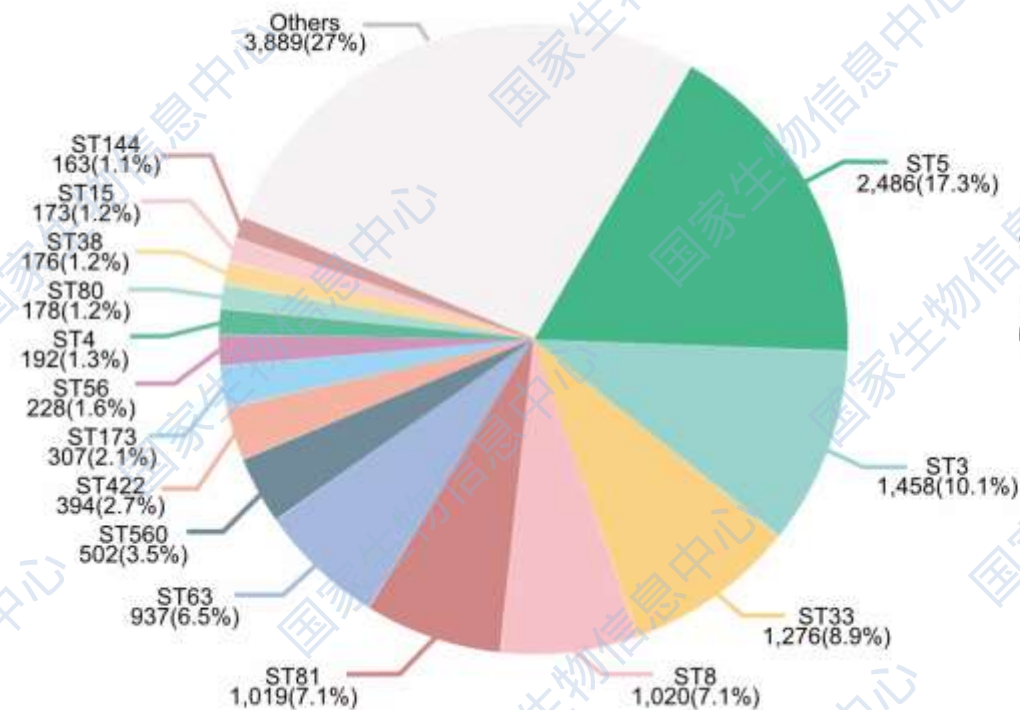
- 涵盖了五大洲的42个国家，其中21个国家报告的NTM菌株数量超过50株
- 英国、美国、澳大利亚、德国和日本成为菌株数量最多的前五个国家
- 时间跨度从1901年至2023年，2011年之后数量上升
- 在2010年前，*M. avium*是最常见的物种（60.8%），而*M. abscessus*仅占15.6%。
- 然而，*M. abscessus*在2011年后迅速增长，从2011年的22.35%上升到2022年的86.6%。

NTM基因组解析流程

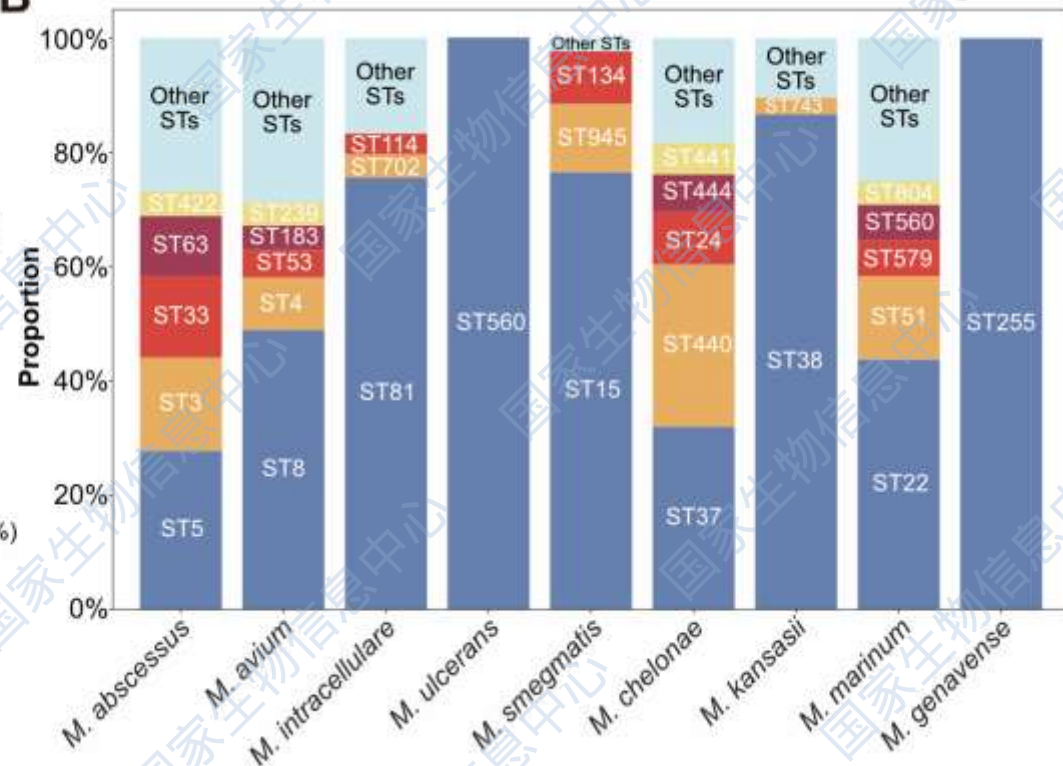


NTM分子分型特征

A



B



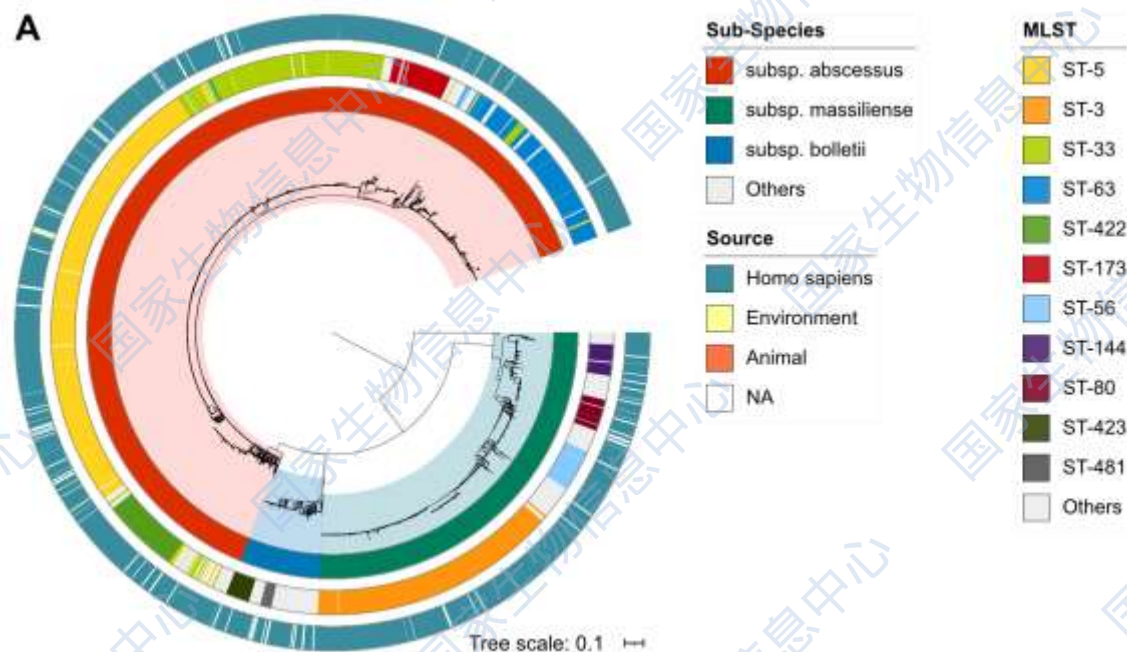
□ MLST分析共鉴定出**705种**，前三位分别是ST5、ST3和ST33

□ *M. abscessus*、*M. chelonae*和*M. avium*表现出较高的**ST类型多样性**，较高的遗传变异性

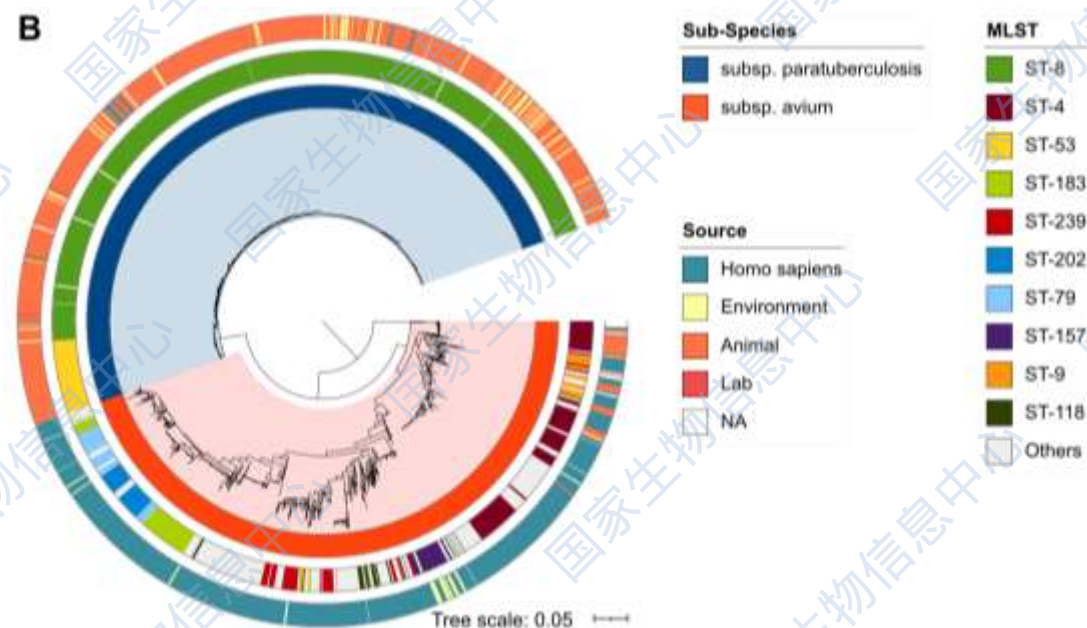
□ *M. intracellulare*、*M. ulcerans*、*M. kansasii*等物种由**单一优势ST类型**主导，较保守的遗传特征

NTM不同物种系统发育解析

M. abscessus

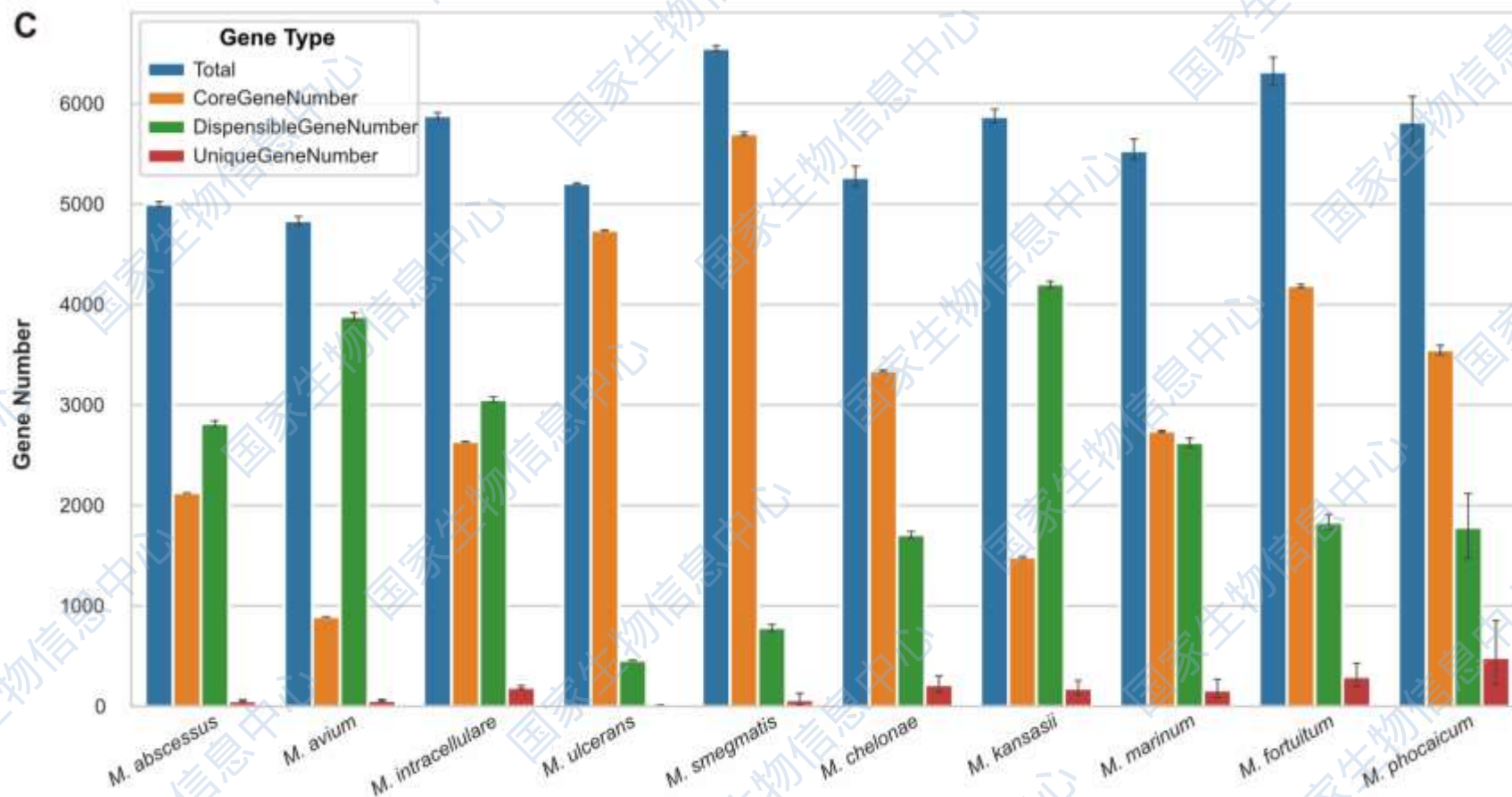


M. avium



- ❑ *M. abscessus*已进化为三个主要支系，分别对应其三个亚种：*M. abscessus subsp. abscessus*，*M. abscessus subsp. massiliense*和*M. abscessus subsp. bolletii*。
- ❑ 这三个亚种对人类都有一定的致病性
- ❑ *M. avium*主要由两个亚种组成，分别是*M. avium subsp. paratuberculosis*和*M. avium subsp. avium*
- ❑ *M. avium subsp. paratuberculosis*高度保守，而*M. avium subsp. avium*展示了显著的进化多样性

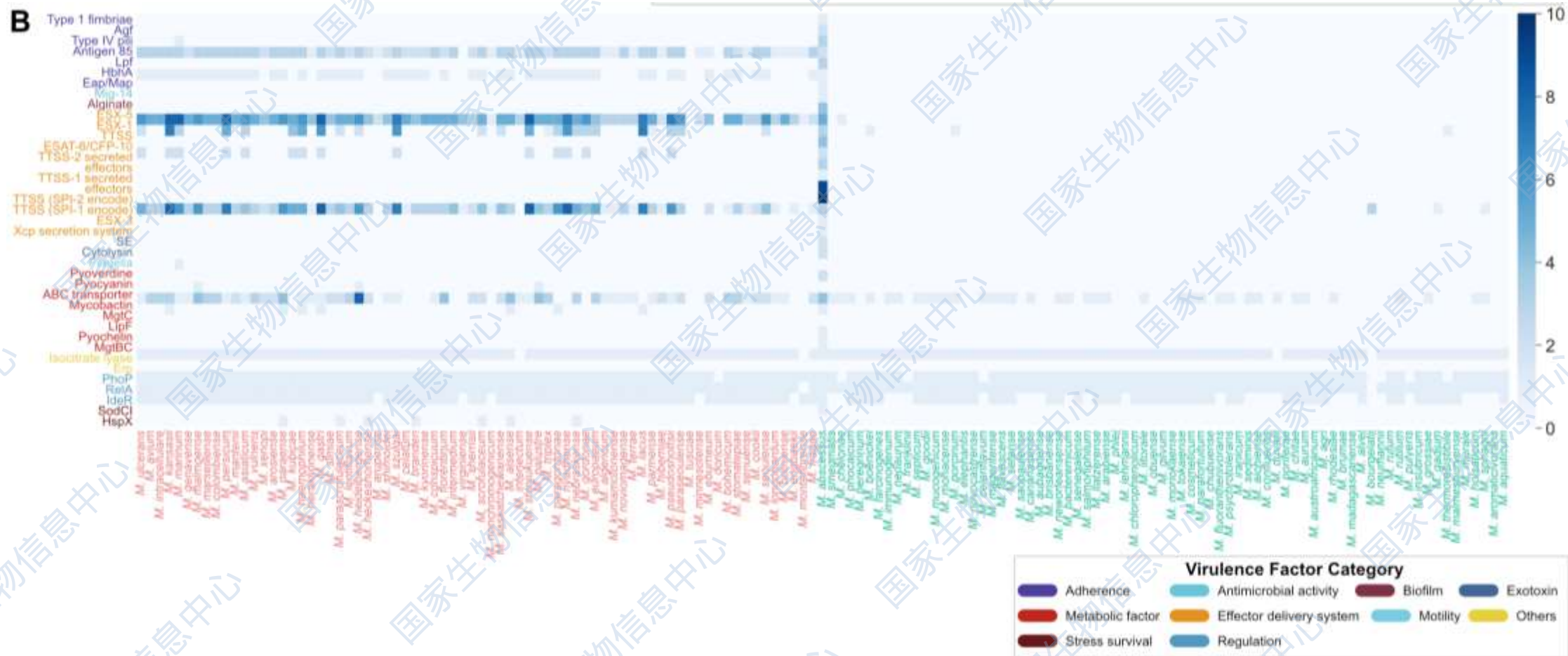
NTM泛基因组特征



- *M. abscessus*、*M. avium* 和 *M. kansasii* 的核心基因比例较低
- *M. marinum* 和 *M. intracellulare* 的核心基因在 ~50%
- *M. smegmatis*、*M. ulcerans*、*M. fortuitum* 等的核心基因比例较高



NTM毒力特征解析

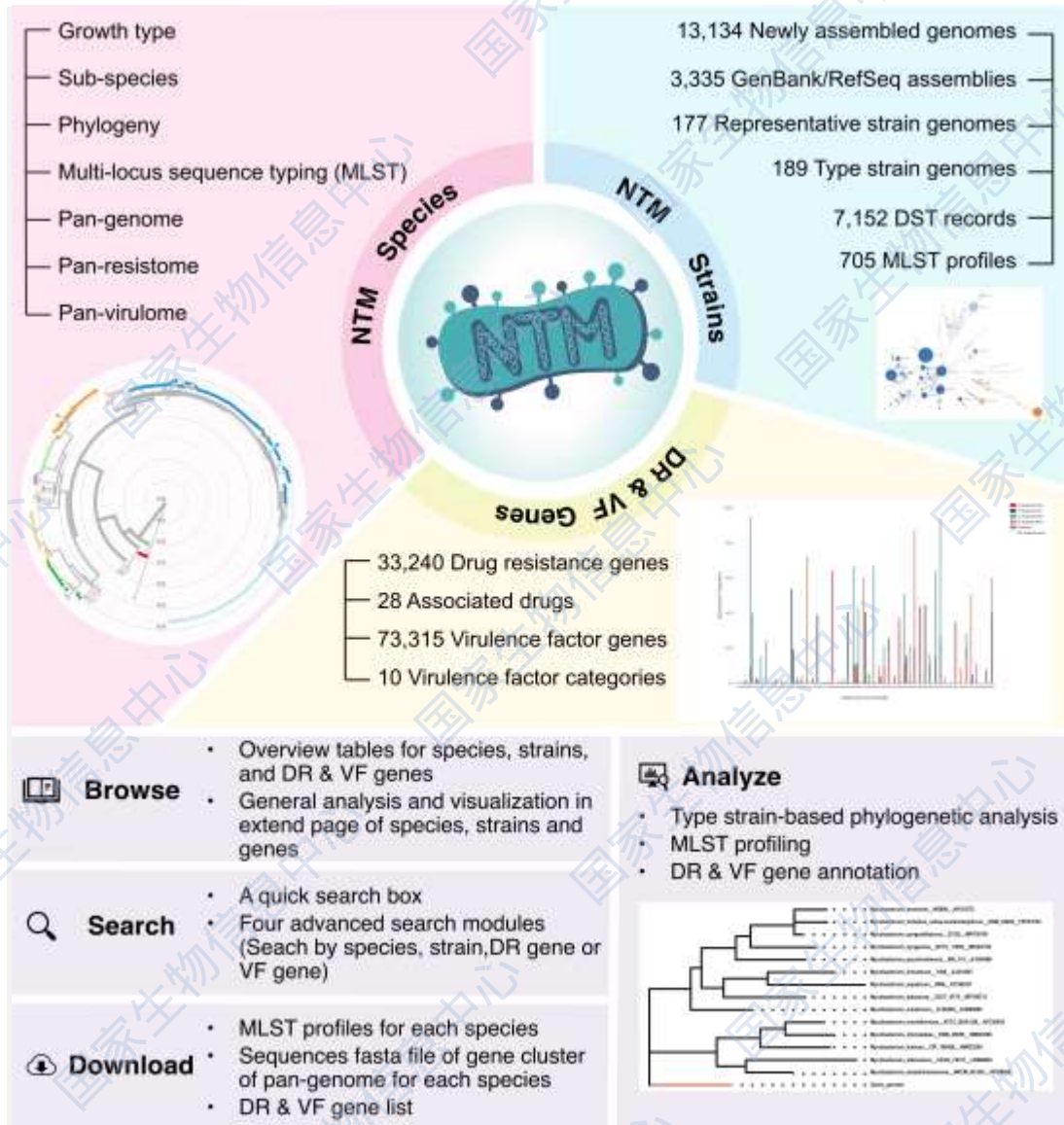


□ 38种不同的毒力因子（VF）基因

□ 慢生长的物种具有较多的毒力因子

□ 快生长中只有*M. abscessus*携带大量毒力基因，可见其在致病性以及耐药性都很强，促进其流行播散

NTM基因组数据资源平台构建



□ 大规模基因组数据

□ 涵盖了177个NTM物种

□ 整合了7,152条抗生素药敏试验结果记录

□ 多位点序列分型、跨物种系统发育分析、泛基因组分析、毒力和耐药基因注释

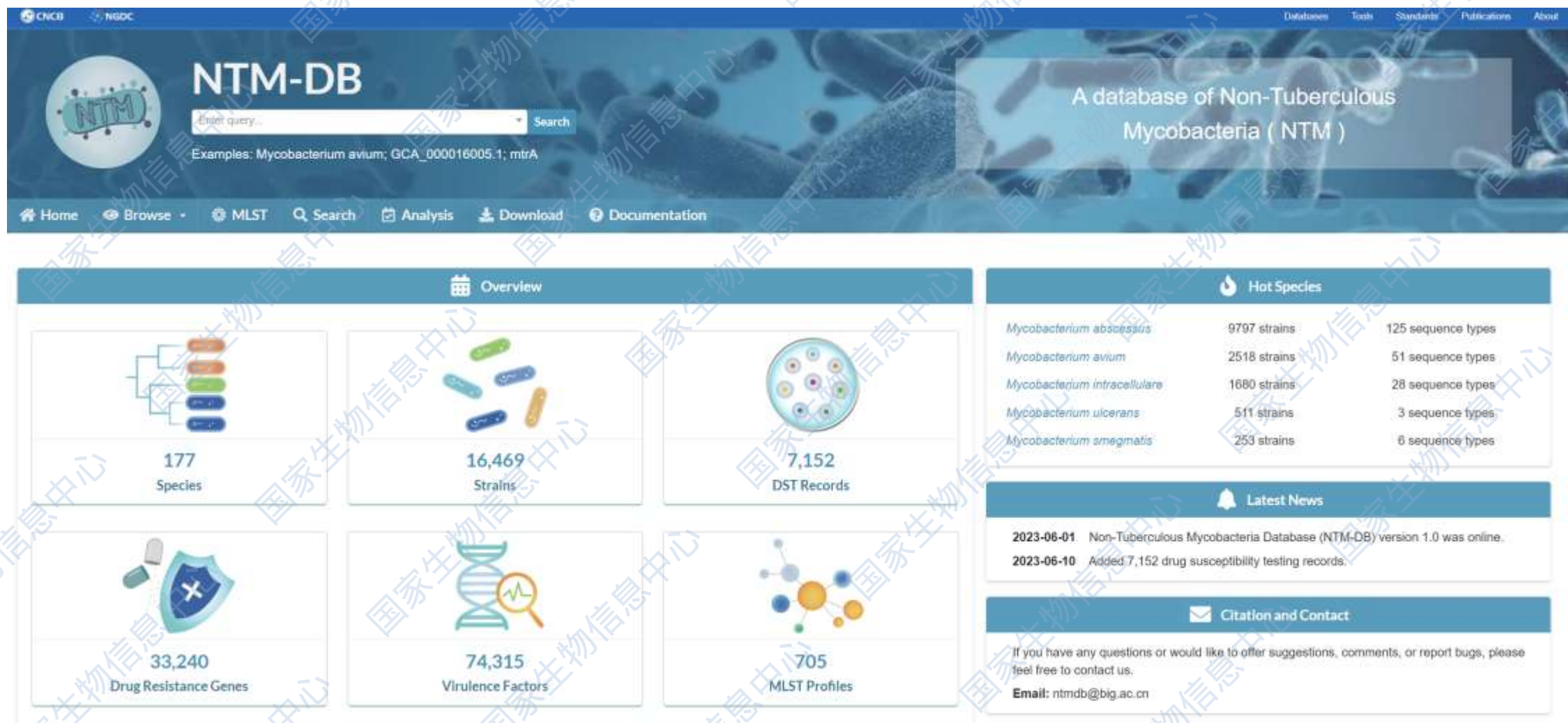
□ 浏览、搜索、下载

□ 在线分析工具

□ NTM-DB基于这些数据提供了丰富的可视化和交互功能

□ 为NTM临床诊断和治疗提供重要数据资源

NTM基因组数据资源库概览



<https://ngdc.cncb.ac.cn/ntmdb/>



NTM-DB浏览模块-物种

□ 一级页面

Home > Browse Species

Show 20 entries [Download Table](#)

Species	Taxonomy ID	Type Strain Genome	Representative Genome	Risk Group	Growth Type	Number of Strains
Mycobacterium abscessus	36809	GCF_000069185.1	GCF_004028015.1	2	rapidly	9333
Mycobacterium avium	1764	GCF_009741445.1	GCF_022175585.1	2	slowly	2271
Mycobacterium intracellulare	1767	GCF_000277125.1	GCF_020736005.1	2	slowly	1620
Mycobacterium ulcerans	1809	GCF_022374915.2	GCF_020616615.1	3	slowly	511
Mycobacterium smegmatis	1772	GCF_001457595.1	GCF_013349145.1	2	rapidly	227
Mycobacterium chelonae	1774	GCF_001632805.1	GCF_001632805.1	2	rapidly	214
Mycobacterium kansasii	1768	GCF_000157895.3	GCF_000157895.3	2	slowly	208
Mycobacterium mageritense	1781	GCF_003391395.1	GCF_016745295.1	2	slowly	63
Mycobacterium goodii	36812	GCF_007526915.1	GCF_000526915.1	2	slowly	44
Mycobacterium parafortuitum	1389713	GCF_003614435.1	GCF_003614435.1	1	slowly	38
Mycobacterium fortuitum	1766	GCF_000295855.1	GCF_001307545.1		rapidly	29

NTM-DB浏览模块-物种

□ 二级页面-overview

Overview of *Mycobacterium abscessus*

General Information

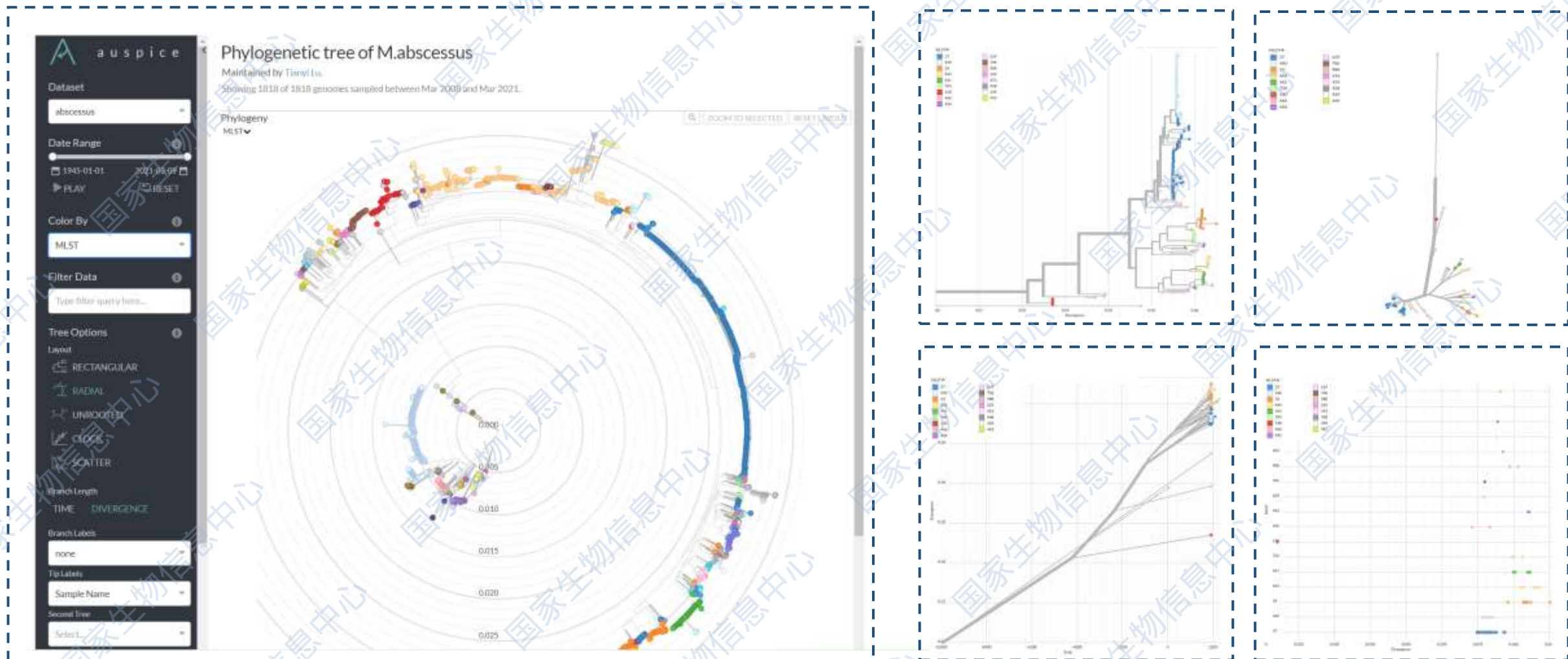
Taxonomy ID	36809
Type Strain	ATCC 19977; CCUG 20993; CIP 104536; DSM 43491; DSM 44196; Hauduroy L948; JCM 13569; L948; NCTC 13031; TMC 1543
Type Strain Genome	GCF_000069185.1
RefSeq Representative Genome	GCF_004028015.1
Risk Group	2
Growth Type	rapidly growing mycobacteria
Strain Number	9333

Subspecies

Subspecies Name	Taxonomy ID	Type Strain Genome	Representative Genome	Strain Number
<i>M.abscessus</i> subsp. <i>abscessus</i>	1185650	GCF_000069185.1	GCF_004028015.1	5037
<i>M.abscessus</i> subsp. <i>massiliense</i>	1962118	GCF_000239055.1		2057
<i>M.abscessus</i> subsp. <i>bolletii</i>	319705	GCF_003609715.1		421

NTM-DB浏览模块-物种

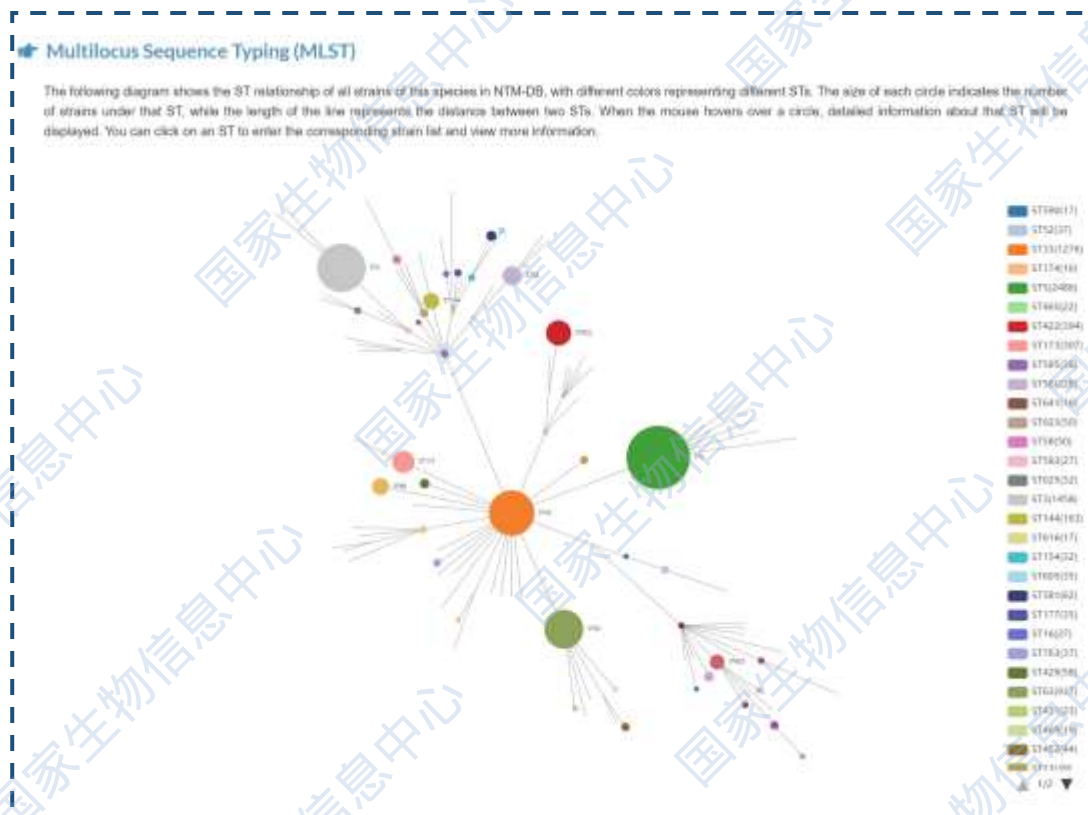
□ 二级页面-进化





NTM-DB浏览模块-物种

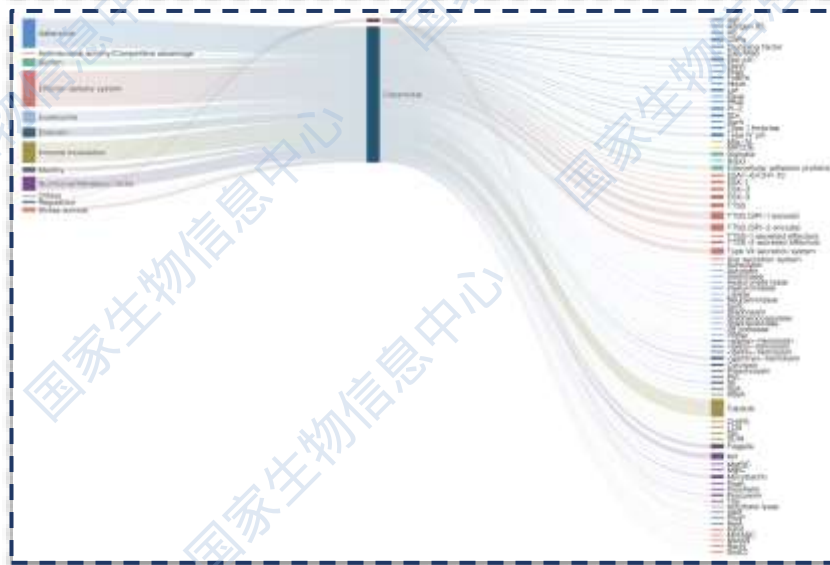
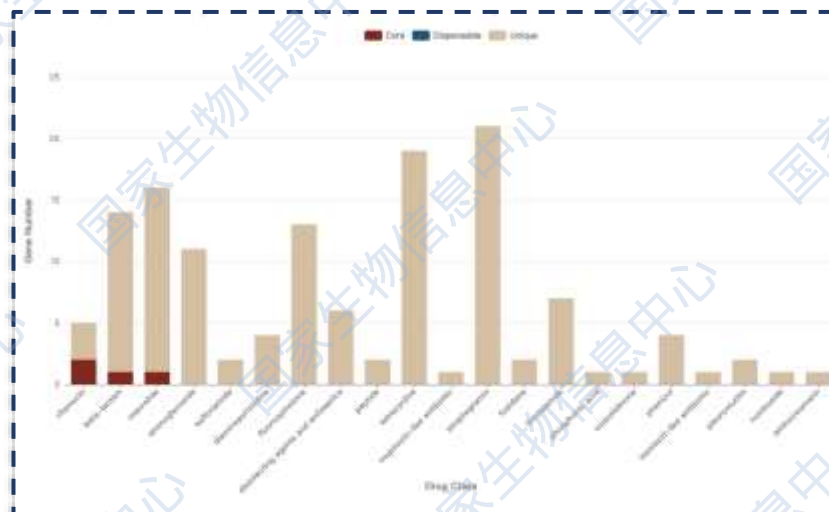
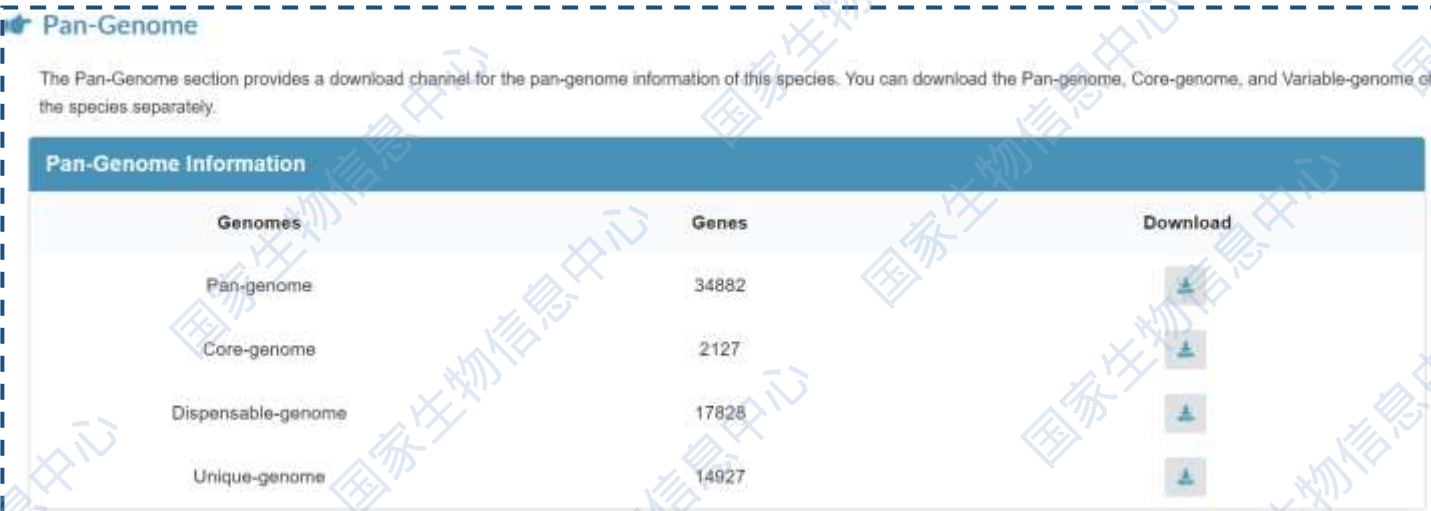
□ 二级页面-ST分型



Accession	Taxonomy	Representati...	MLST	Isolation Source	VF Genes	Resistance Genes
GCF_004028015.1	<i>Mycobacterium abscessus</i>	Representative Gene me	ST5	-	Others: isochitrate lyase Regulation: PhoP; RelA Nutritional/Metabolic factor: Mycobactin	
GCF_800132815.1	<i>Mycobacterium abscessus</i>	-	ST5	-	Others: isochitrate lyase Regulation: PhoP; RelA Nutritional/Metabolic factor: Mycobactin	antibiotic target alteration: rpoB2 antibiotic efflux: mtrA antibiotic target protection: RbpA
GCF_800133215.1	<i>Mycobacterium abscessus</i>	-	ST5	-	Others: isochitrate lyase Regulation: PhoP; RelA Nutritional/Metabolic factor: Mycobactin	antibiotic efflux: mtrA antibiotic target alteration: rpoB2 antibiotic target protection: RbpA
GCF_800133175.1	<i>Mycobacterium abscessus</i>	-	ST5	Human sapiens: pulmonary	Others: isochitrate lyase Regulation: PhoP; RelA Nutritional/Metabolic factor: Mycobactin	antibiotic target alteration: rpoB2 antibiotic efflux: mtrA antibiotic target protection: RbpA
GCF_800133115.1	<i>Mycobacterium abscessus</i>	-	ST5	Human sapiens: pulmonary	Others: isochitrate lyase Regulation: PhoP; RelA Nutritional/Metabolic factor: Mycobactin	antibiotic target alteration: rpoB2 antibiotic efflux: mtrA antibiotic target protection: RbpA
GCF_800133075.1	<i>Mycobacterium abscessus</i>	-	ST5	-	Others: isochitrate lyase Regulation: PhoP; RelA Nutritional/Metabolic factor: Mycobactin	antibiotic efflux: mtrA antibiotic target alteration: rpoB2 antibiotic target protection: RbpA
GCF_800132995.1	<i>Mycobacterium abscessus</i>	-	ST5	-	Others: isochitrate lyase Regulation: PhoP; RelA	antibiotic target alteration: rpoB2 antibiotic efflux: mtrA

NTM-DB浏览模块-物种

□ 二级页面-泛基因组、耐药和毒力



- 泛基因组结果下载
- 按核心、非必需和菌株特异性基因分类
- 统计药物的耐药基因数量
- 根据VFDB整理的主要毒力机制类别和对应毒力因子的基因数量

NTM-DB浏览模块-菌株

□ 一级页面

Home > Browse Strains

More Information

Select All Clear All

Show Columns ☐ Year ☐ Country ☐ Assembly total length ☐ GC percentage ☐ Contig L50 ☐ Contig N50

View drug susceptibility testing (DST) results >

Show 20 entries Download Table

Accession	Taxonomy	Representative	MLST	Isolation Source	VF Genes	Resistance Genes	DST
GCF_900566085.1	Mycobacterium attenuatum	Type strain Genome	ST851	-	Regulation: PhoP, RelA, IdeR Others: Isocitrate lyase Effector delivery system: ESX-3; ESX-1; ESX-5, 6 SAT-6/CFP-10 Adherence: Antigen 85; Hbha Nutritional/Metabolic factor: Mycobactin	-	-
GCF_900454025.1	Mycobacterium gilvum	Type strain Genome	ST682	-	Regulation: PhoP, IdeR, RelA Others: Isocitrate lyase	antibiotic target alteration: rpsB2 antibiotic efflux: ntrA; tap antibiotic target protection: RbpA	-



NTM-DB浏览模块-菌株

□ 一级页面-人工审编的DST结果

Filter of drug class											
Select Default Select All											
Select Drug Class		☒ Aminoglycosides ☒ β -lactams ☐ Chloramphenicol ☐ Glycopeptides ☒ Macrolides ☐ Oxazolidinones ☐ Quinolones ☐ Tetracyclines ☐ Others									
Pin Columns		Default Pinned Clear Pinned									
Show 10 entries		Download Table									
Species	Strain	Aminoglycosides					β -lactams				
		Aminoglycosides	Amikacin	Capreomycin	Streptomycin	Tobramycin	Cefoxitin	Faropenem	Imipenem	Meropenem	
<i>Mycobacterium smegmatis</i>	DRR207720	-	[S] / [1.453]	[-] / [1.331]	[-] / [4.311]	-	-	-	-	[S] / [1.882]	
<i>Mycobacterium smegmatis</i>	DRR207721	-	[S] / [1.138]	[-] / [0.928]	[-] / [4.311]	-	-	-	-	[S] / [2.338]	
<i>Mycobacterium smegmatis</i>	DRR207722	-	[S] / [1.138]	[-] / [0.928]	[-] / [3.840]	-	-	-	-	[S] / [2.338]	
<i>Mycobacterium smegmatis</i>	DRR207723	-	[S] / [1.610]	[-] / [0.827]	[-] / [4.311]	-	-	-	-	[S] / [1.653]	
<i>Mycobacterium smegmatis</i>	DRR207724	-	[S] / [1.610]	[-] / [0.928]	[-] / [5.431]	-	-	-	-	[S] / [1.856]	



NTM-DB浏览模块-菌株

□ 二级页面

菌株基础信息

Overview of *M.intracellulare*: GCF_000277125.1

Organism name	<i>Mycobacterium intracellulare</i>
GenBank assembly accession	GCF_000277125.1
Representative strain	No
Type strain	Yes
Isolation source	Homo sapiens:clinical isolates from a Korean pulmonary patient; male; 64 age; Asan medical center
DST	Not tested
CheckM Completeness	99.94%
CheckM Contamination	0.00%
SSU Count	1
tRNA Count	46
Genome Size	5,402,402
Coding density	91.56%
GC content	68.10%
Contig N50	5402402
Contig L50	1

耐药基因注释信息

Resistome

4 resistance genes found in this strain.

Accession	Gene Name	Drug Class	Mechanism	Coverage	Identity
ARO:3003955	efpA	isoniazid,rifamycin	antibiotic efflux	99.94	83.70
ARO:3000816	mtrA	macrolide,penam	antibiotic efflux	100.00	90.25
ARO:3000245	RbpA	rifamycin	antibiotic target protection	93.04	88.47
ARO:3000501	rpoB2	rifamycin	antibiotic target alteration	97.31	81.17

毒力基因注释信息

Virulome

27 virulence genes found in this strain.

Accession	Gene Name	Virulence Factor	VF full name	VF Category	Coverage	Identity
VFG001811	fbpC	Antigen 85		Adherence	95.50	84.03
VFG001406	ideR	IdeR	Iron-dependent regulator	Regulation	99.57	86.72
VFG002394	esxM	ESX-5	ESAT-like secretion system	Effector delivery system	99.33	85.19



NTM-DB浏览模块-基因

耐药基因：一级页面

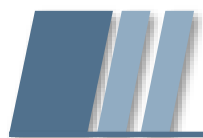
Home Browse MLST Search Analysis Download Documentation						
Home > Browse Resistance Genes						
Resistance Genes Virulence Genes						
Show 10 entries Download Table						
Gene Name	Accession	Number of Species	Number of Strains	Mechanism	Drug	
<i>mphA</i>	ARO3000816	121	9592	antibiotic efflux	macrolide:penam	
<i>RfpA</i>	ARO3000245	123	9583	antibiotic target protection	rifamycin	
<i>rpoB2</i>	ARO3000501	120	9571	antibiotic target alteration	rifamycin	
<i>efpA</i>	ARO3003995	46	909	antibiotic efflux	isoniazid:rifamycin	
<i>AAC(2)-Ic</i>	ARO3002525	24	738	antibiotic inactivation	aminoglycoside	
<i>APH(3)-Ia</i>	ARO3002641	4	604	antibiotic inactivation	aminoglycoside	
<i>lfrA</i>	ARO3003967	14	196	antibiotic efflux	fluoroquinolone	
<i>tet(V)</i>	ARO3000181	11	195	antibiotic efflux	tetracycline	
<i>arr-7</i>	ARO3002846	14	193	antibiotic inactivation	rifamycin	



NTM-DB浏览模块-基因

耐药基因：二级页面





NTM-DB浏览模块-基因

毒力基因：一级页面

Home > Browse Virulence Genes

Resistance Genes Virulence Genes

Show 10 entries Download Table

Gene Name	Accession	Number of Species	Number of Strains	Virulence Factor	Virulence Factor Category
icl	VFG001381	152	10288	Isocitrate lyase	Others
relA	VFG001826	154	10283	RelA	Regulation
phoP	VFG001386	148	10055	PhoP	Regulation
mbtH	VFG001818	87	9862	Mycobactin	Nutritional/Metabolic factor
ideR	VFG001406	124	1411	IdelR	Regulation
esxN	VFG002400	76	1319	ESX-5	Effector delivery system
esxM	VFG002394	72	1311	ESX-5	Effector delivery system
esxH	VFG022098	71	1263	ESX-3	Effector delivery system
fbpC	VFG001811	76	1261	Antigen 85	Adherence



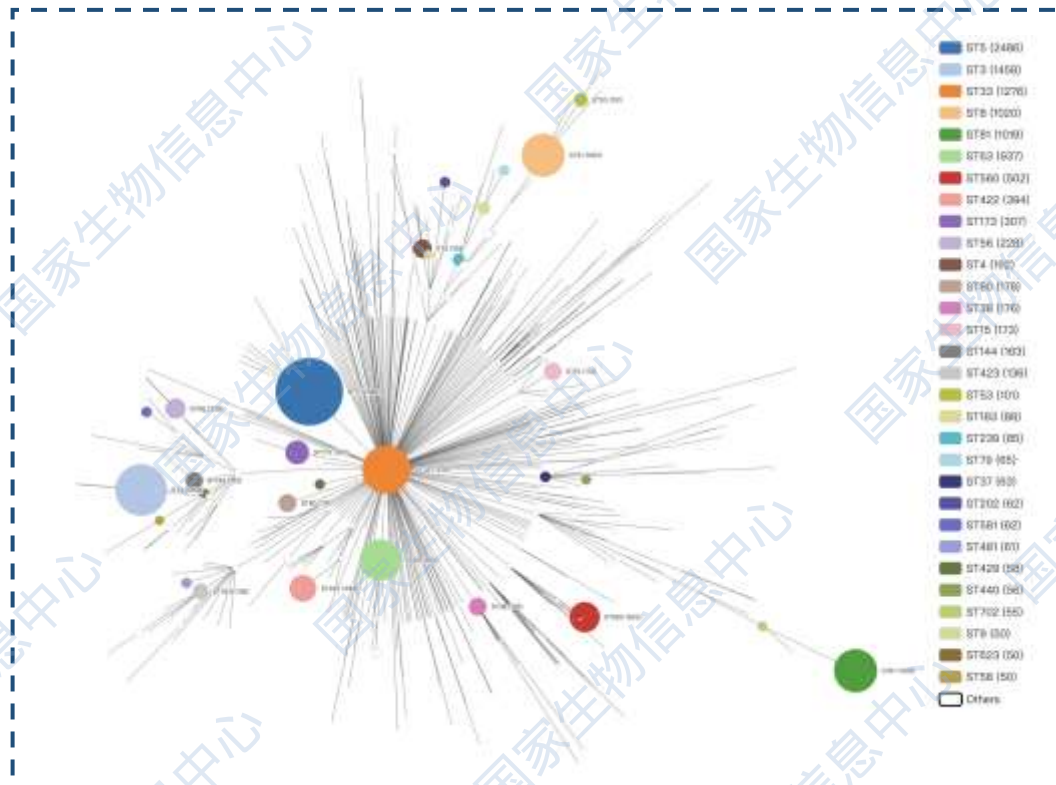
NTM-DB浏览模块-基因

毒力基因：二级页面

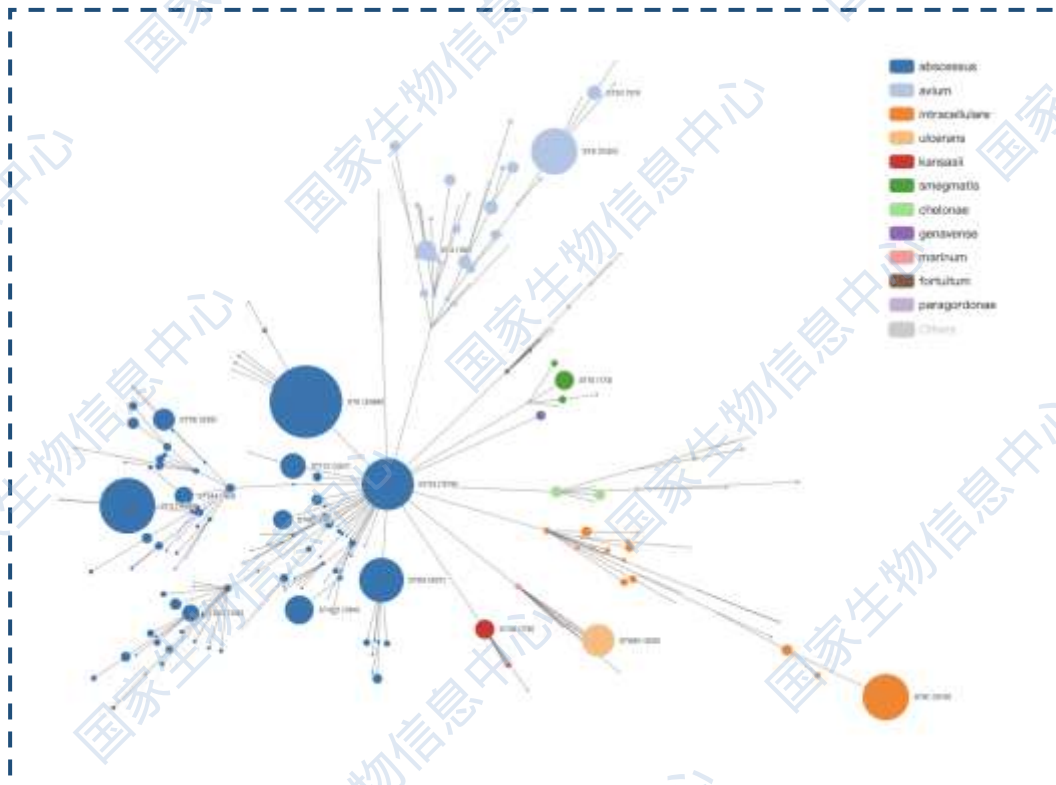


NTM-DB之MSLT模块

所有ST型的分型



主要致病NTM物种ST型的分型



- ✓ 每个ST的菌株都来自一个物种
- ✓ 每个物种的ST型之间呈现聚集成簇
- ✓ 说明ST型对NTM的物种有较好的分辨率

NTM-DB搜索模块

Search by Species

Species Name

e.g. *M. abscessus*

Search by Strains

Accession

e.g. GCA_000016005.1

Search by Resistance Genes

Gene Name

e.g. *mtrA*

Drug Class

Select a drug class

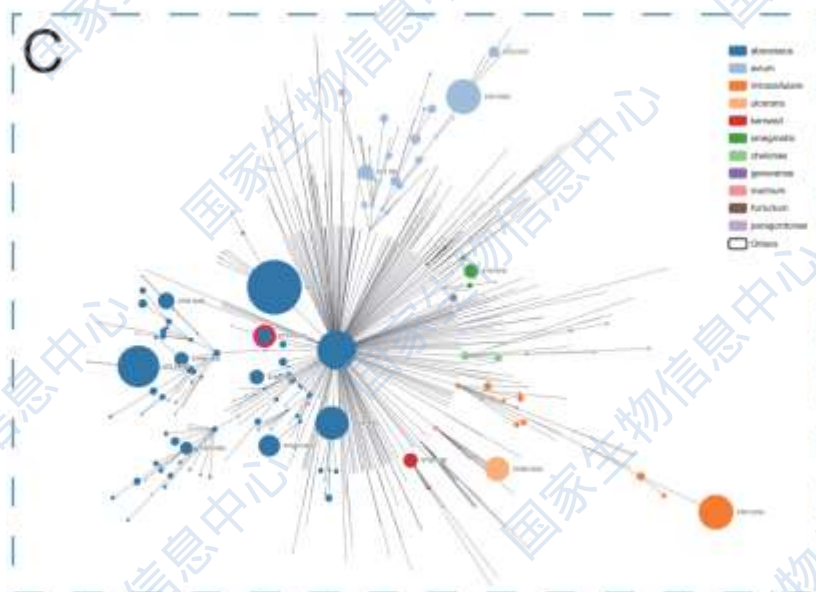
Search by Virulence Genes

Gene Name

e.g. *relA*

VF Category

Select a VF category



NTM基因组数据资源库应用案例（一）

针对NTM不同亚种鉴定困难的问题，运用基因组学+机器学习思路



比较基因组学

特异基因和SNPs



机器学习之随机森林

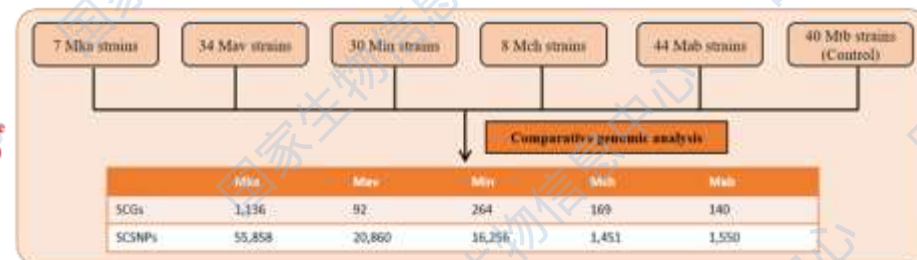
二分类模型构建



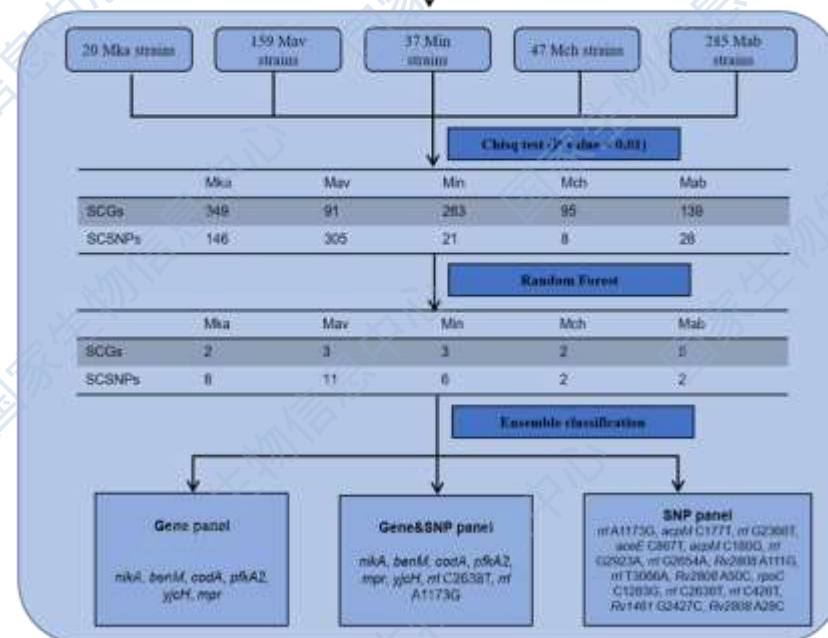
机器学习之集成学习

多分类模型构建

NTMs with complete genome (Discovery set)

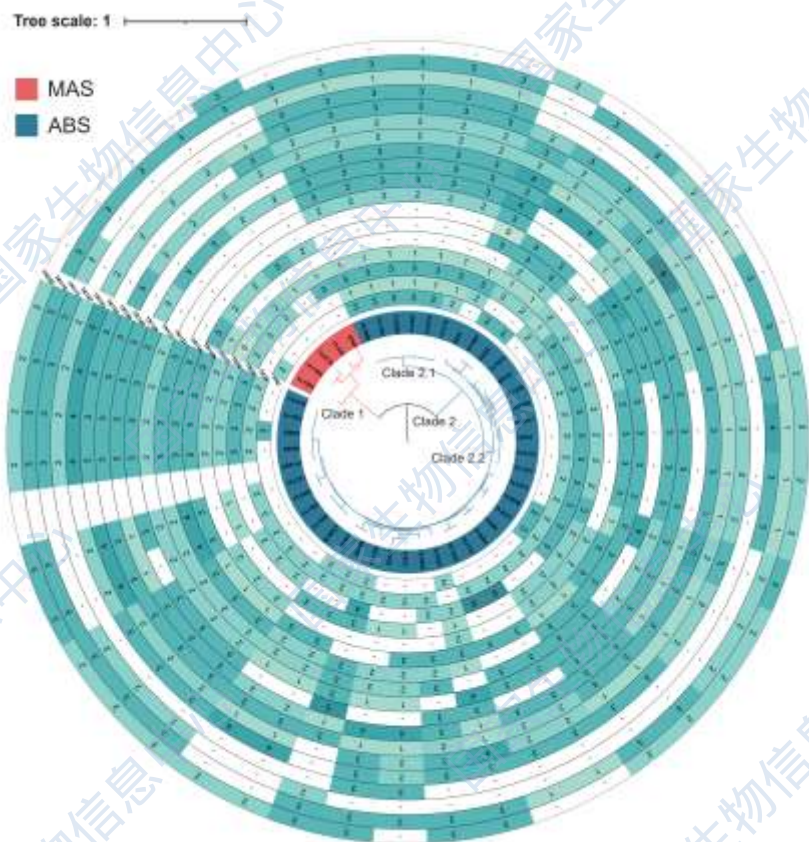


NTMs with raw sequencing reads (Validation set)

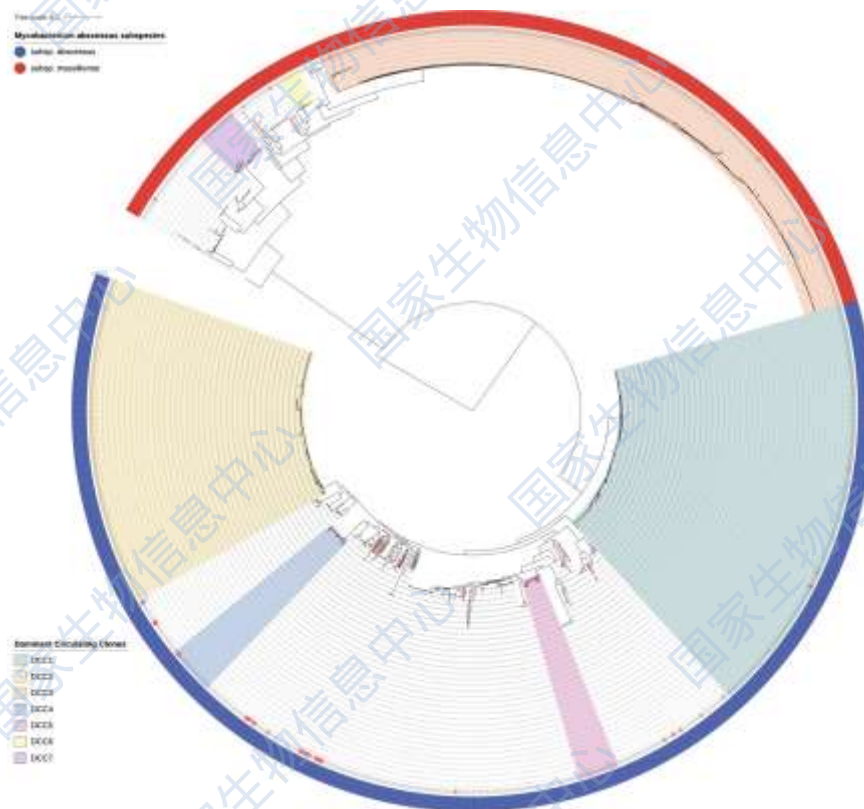


NTM基因组数据资源库应用案例（二）

基于自测临床样本，整合本数据库中的基因组数据，解析整个种群的基因组特征



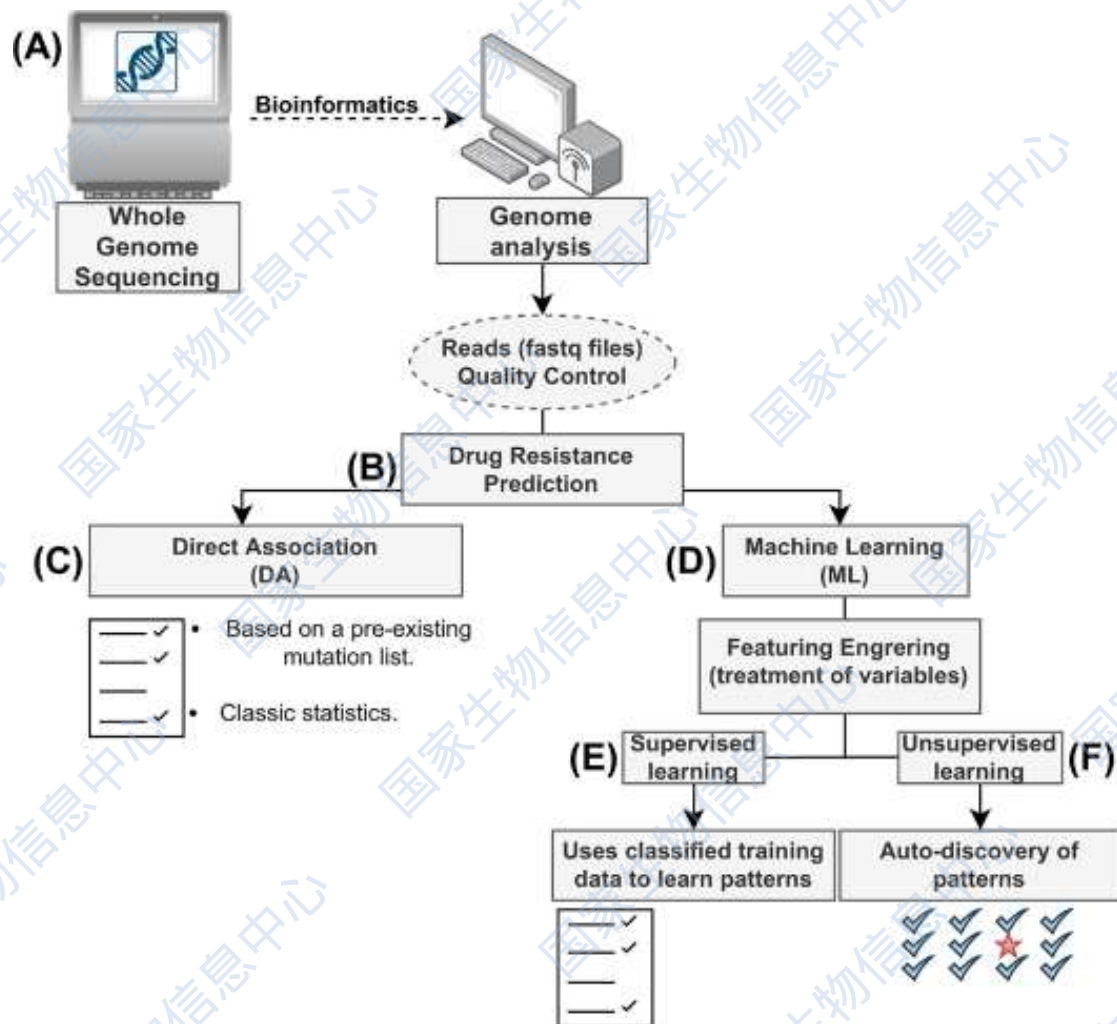
- 来自30个病人的45个样本
- (有22株来自8个病人连续采样)



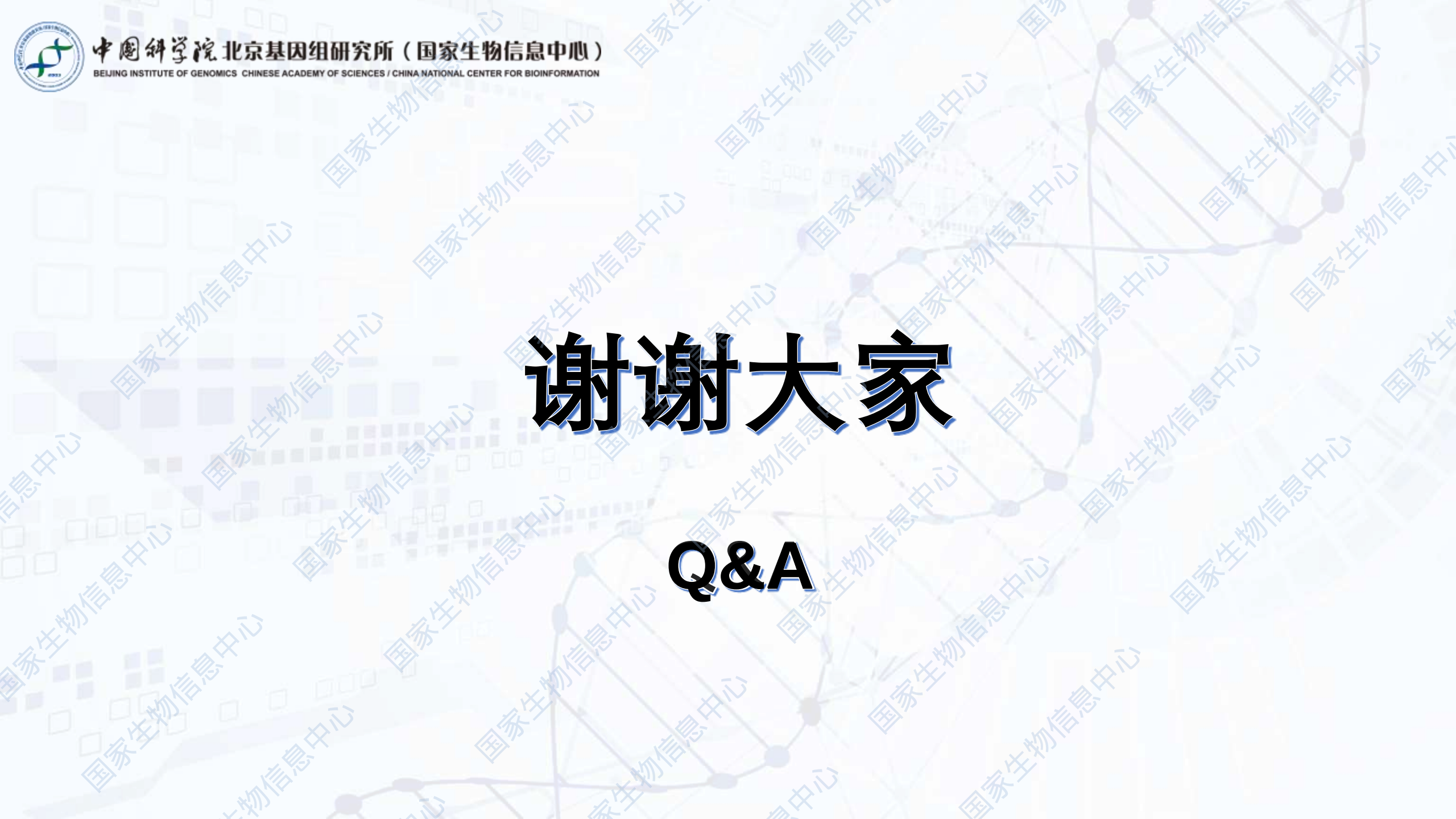
- 鉴定属于高危克隆群的菌株：有8株来来自DCC1、DCC3和DCC4的显性循环克隆群

NTM基因组数据资源库应用案例（三）

耐药表型+基因组+统计学/机器学习解析耐药机制



- 人工审编的耐药表型信息
- 基因组学信息
- 统计学、机器学习



中国科学院北京基因组研究所（国家生物信息中心）

BEIJING INSTITUTE OF GENOMICS CHINESE ACADEMY OF SCIENCES / CHINA NATIONAL CENTER FOR BIOINFORMATION

谢谢大家

Q&A