U.S. FOOD & DRUG ADMINISTRATION

Human Foods Program

BoNTax

Genome-based classification and characterization of *Clostridium botulinum* Group I and *Clostridium sporogenes*

Jayanthi Gangiredla

Office of Laboratory Operations and Applied Science (OLOAS)
Human Foods Program, FDA
Annual GenomeTrakr-2026

Background & The Classification Challenge

- **Public Health Significance:** Botulism is a rare but life-threatening neuroparalytic illness — BoNT is one of the most potent biological toxins known, with food safety and public health implications.
- **The Organism:** *C. botulinum* Group I and *C. sporogenes* are closely related, anaerobic, spore-forming bacteria frequently encountered in food safety and clinical investigations.
- **The Classification Problem:** These two species share substantial genomic similarity yet have historically been classified based on toxin production alone — unreliable and inconsistent across laboratories.
- **Why Genomics?** BoNT genes reside on mobile genetic elements (plasmids, phages) and can be horizontally transferred — toxin genotype does not reliably reflect true species identity.
- **The Need:** A standardized, reproducible, genome-based framework is needed to accurately classify these organisms and support outbreak investigations and food safety surveillance.

BoNTax- ANI-Based Taxonomic Framework

A curated, taxonomy-first genomic decision framework prioritizing ANI species assignment over toxin gene content.

Integrates assembly QC, MLST, BoNT detection, and AMR profiling into one automated workflow.

Deployed directly on the FDA HFP GalaxyTrakr platform (galaxytrakr.org).

Highly accessible to non-bioinformaticians—no command-line execution required. (galaxytrakr.org).

Species assignment and toxin characterization in one workflow.

A curated reference database

966

NCBI genomes screened

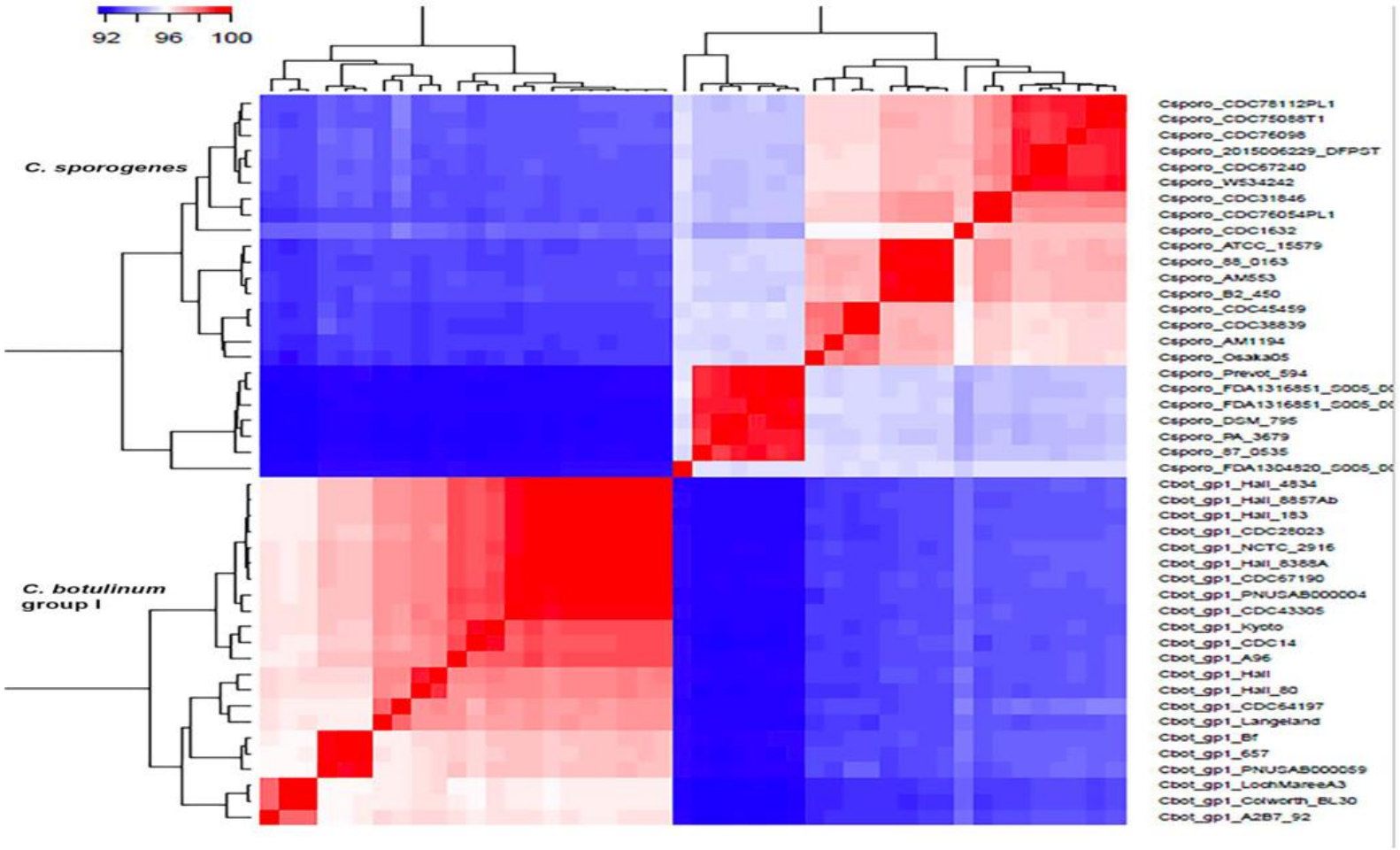
46

curated representative genomes

spanning intra-species ANI diversity of 95–100%.

Reference composition	22 *C. botulinum* Group I 24 *C. sporogenes*
Assembly-size QC	3.6–4.5 Mb screening range
ANI Criteria	Within-species ANI ≥95%; Between-species ANI <95%

ANI separation of the reference species



Reference ANI pattern

Two genomic clusters correspond to the reference species.

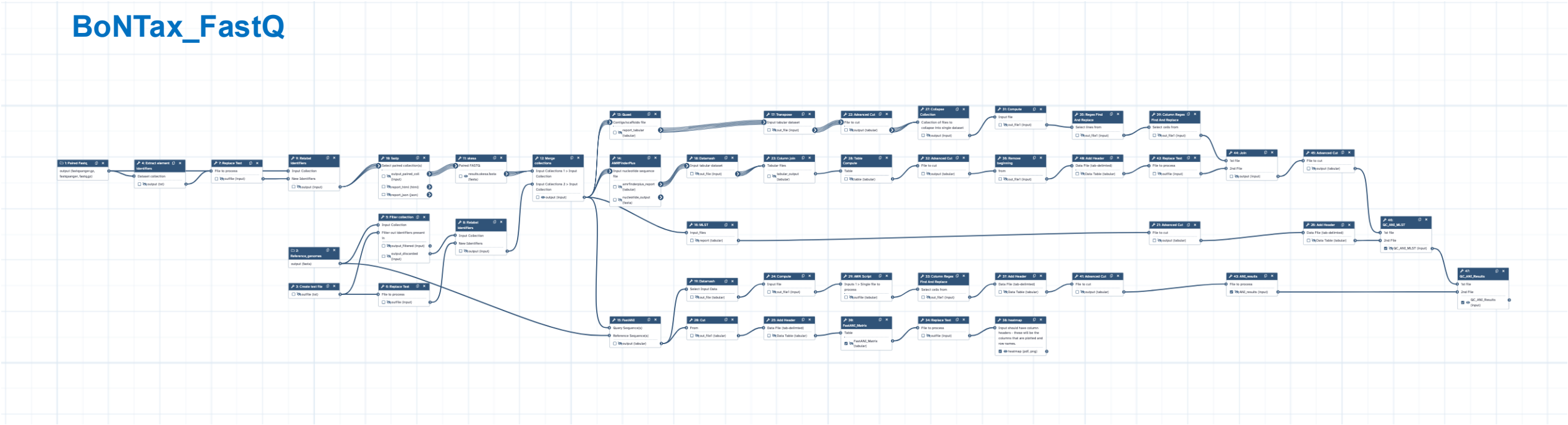
Framework threshold ANI ≥95%

Review borderline or ambiguous assignments.

BoNTax- GalaxyTrakr workflow

An integrated, user-friendly tools and workflow system. Access via web browser with **no** installation and command line experience is required.

BoNTax_FastQ

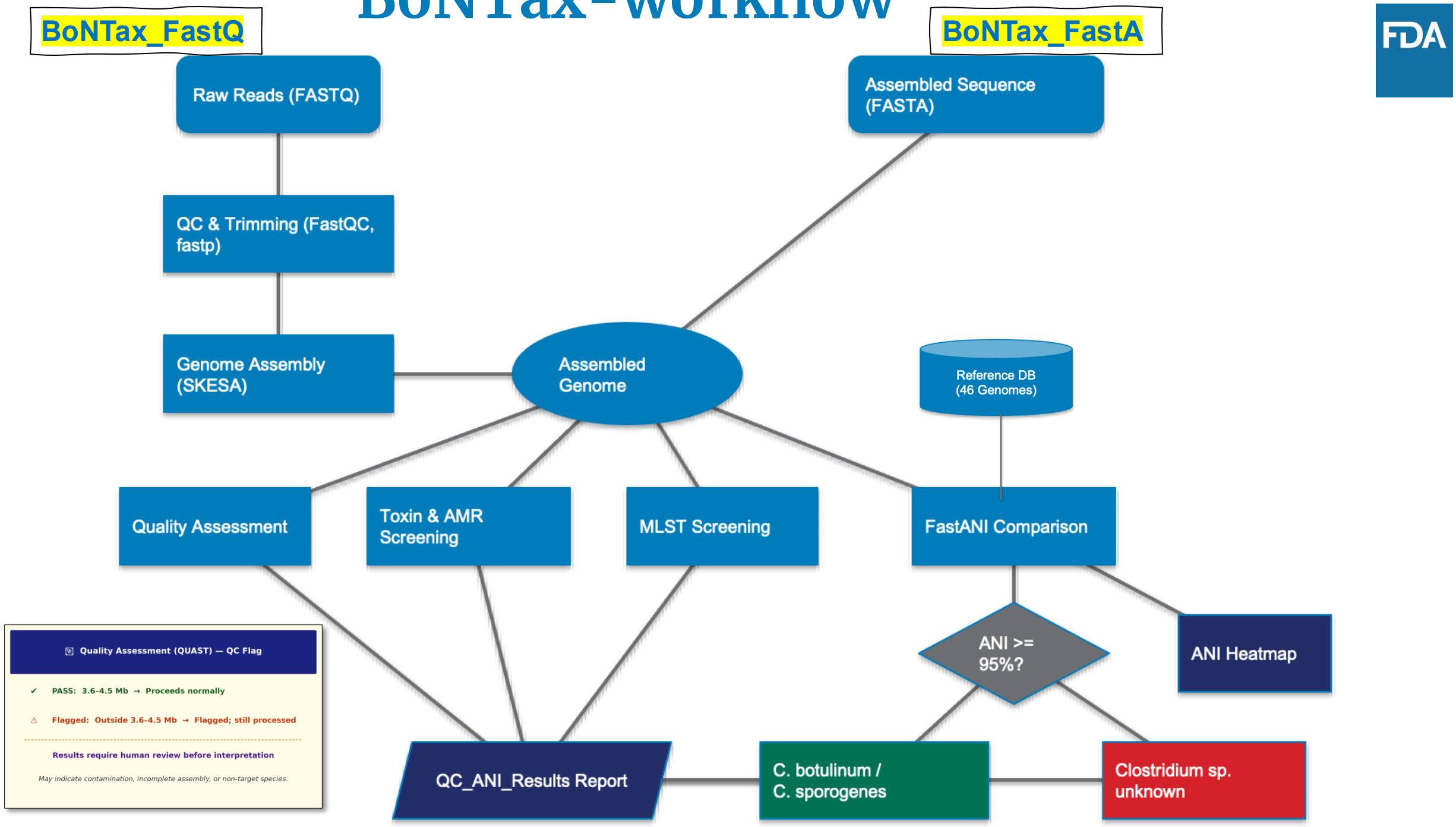


All tools available in the GalaxyTrakr panel.

Validated, Open-Source Tools in BoNTax

- **Read Processing: Trimmomatic**
  Trims low-quality reads (FASTQ workflow only)
- **Genome Assembly: SKESA**
  Assembles trimmed reads into genome (FASTQ workflow only)
- **Quality Control: QUAST**
  Evaluates assembly metrics (size, contigs, N50, completeness)
- **ANI Calculation: FastANI**
  Compares genomes for species-level assignment
- **MLST Typing: MLST**
  Determines sequence types for strain differentiation
- **Gene Detection: AMRFinderPlus**
  Identifies AMR genes, toxins, and virulence factors

BoNTax-workflow



User Interface & Workflow Walkthrough

1. Upload genome(s) (FastQ or Fasta)
2. Select BoNTax_FastQ or BoNTax_FastA workflow
3. Select reference database
4. Run the workflow
5. **Output Results** QC/ANI/Genotype table and ANI heatmap

BoNTax validation and results

42 genomically diverse samples including *C. botulinum*, *C. sporogenes*, other *Clostridium* species (**3 samples**), unknown taxa (**10 samples**), and mixed cultures (**8 samples**).

Evaluation	Reported finding
Agreement with existing labels	29 of 32 known-label genomes
Unknown-label genomes	10 received ANI-based assignments
Discordant classifications	3 genomes support taxonomic re-evaluation
Known mixed-culture datasets	QC warnings flagged all 8 tested

ANI assignments and QC flags support targeted review.

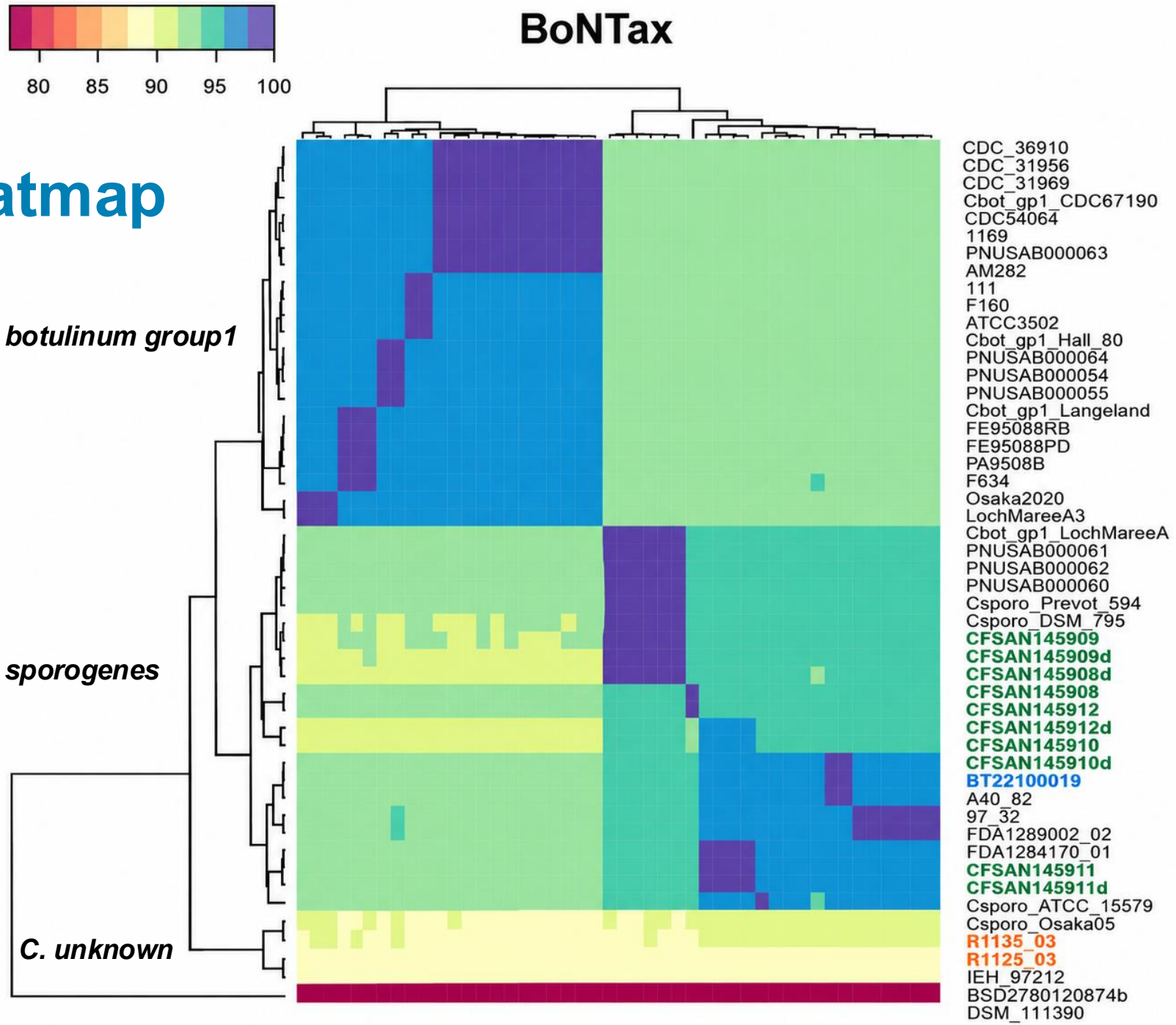
Validation: three discordant classifications

Sample	Reference metadata	ANI-supported match	ANI (%)
BT-22100019	*C. botulinum* Group I	*C. sporogenes* CDC31846	99.92
R1125/03	*C. botulinum*	*C. caseinilyticum* DSM 46279	99.96
R1135/03	*C. botulinum*	*C. caseinilyticum* DSM 46279	99.96

BoNTax assigned BT-22100019 to *C. sporogenes*.
The other two genomes fell outside both reference groups and flagged as ***C. unknown***.

Separate ANI analysis supported *C. caseinilyticum* assignments

ANI heatmap


BoNTax
80
85
90
95
100
C. botulinum group1
C. sporogenes
C. unknown
CDC_36910
CDC_31956
CDC_31969
Cbot_gp1_CDC67190
CDC54064
1169
PNUSAB000063
AM282
111
F160
ATCC3502
Cbot_gp1_Hall_80
PNUSAB000064
PNUSAB000054
PNUSAB000055
Cbot_gp1_Langeland
FE95088RB
FE95088PD
PA9508B
F634
Osaka2020
LochMareeA3
Cbot_gp1_LochMareeA
PNUSAB000061
PNUSAB000062
PNUSAB000060
Csporo_Prevot_594
Csporo_DSM_795
CFSAN145909
CFSAN145909d
CFSAN145908d
CFSAN145908
CFSAN145912
CFSAN145912d
CFSAN145910
CFSAN145910d
BT22100019
A40_82
97_32
FDA1289002_02
FDA1284170_01
CFSAN145911
CFSAN145911d
Csporo_ATCC_15579
Csporo_Osaka05
R1135_03
R1125_03
IEH_97212
BSD2780120874b
DSM_111390

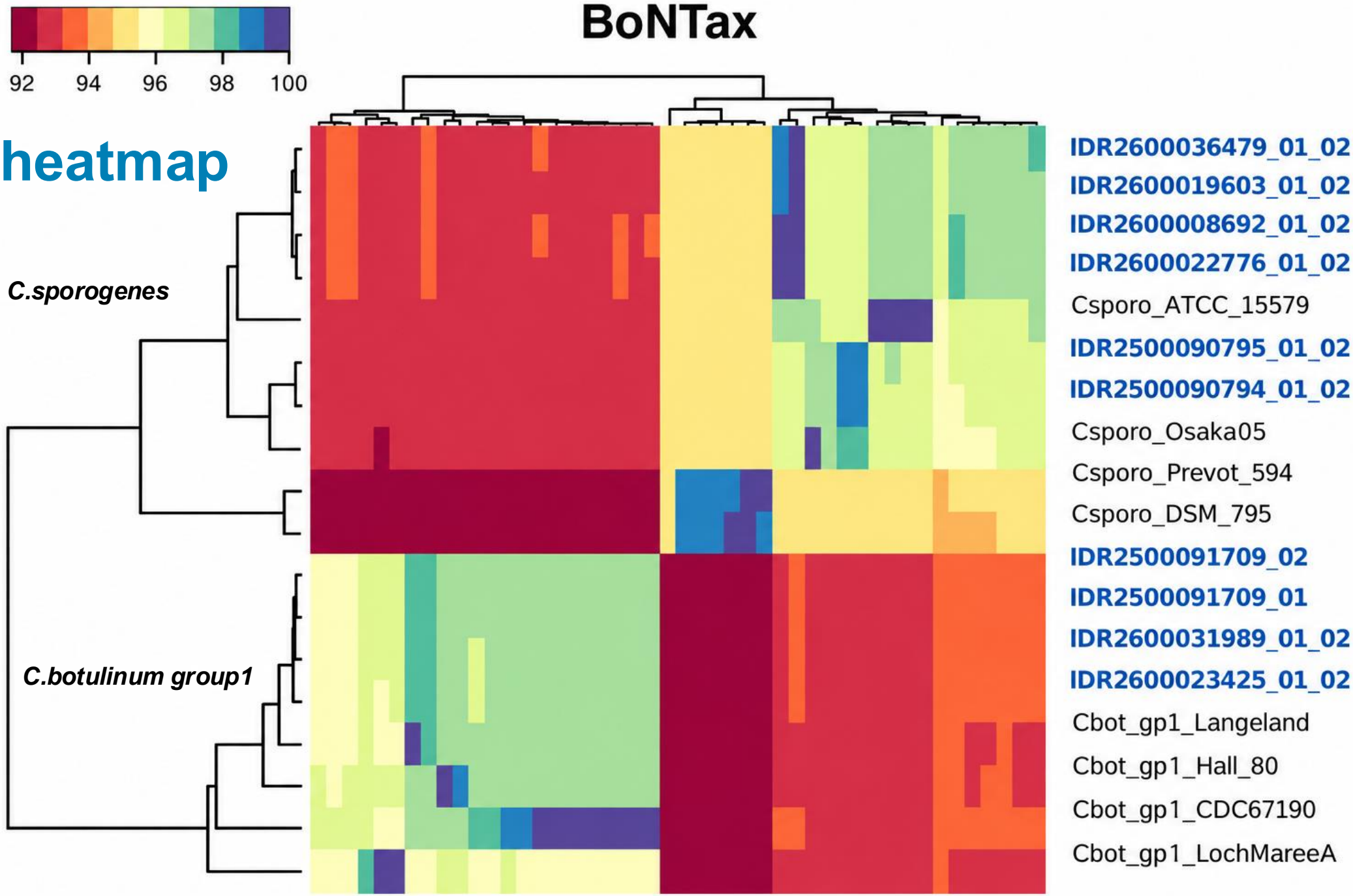
BoNTax: recent-sample metadata

SAMPLE NAME	BIOSAMPLE	COLLECTION_DATE	CONSENT	GEO_LOC_NAME	ISOLATION_SOURCE	ORGANISM	PLATFORM	SEROTYPE
IDR2600036479-01-02	SAMN63663785	2026	public	USA: NY\, New York County	stool	Clostridium botulinum	ILLUMINA	B1
IDR2600031989-01-02	SAMN63663784	2026	public	USA: CT	stool	Clostridium botulinum	ILLUMINA	B7
IDR2600022776-01-02	SAMN63663783	2026	public	USA: CT	stool	Clostridium botulinum	ILLUMINA	B1
IDR2600023425-01-02	SAMN63663782	2026	public	USA: NY\, Kings County	stool	Clostridium botulinum	ILLUMINA	B1
IDR2600019603-01-02	SAMN63663781	2026	public	USA: NY\, Richmond County	stool	Clostridium botulinum	ILLUMINA	B1
IDR2600008692-01-02	SAMN63663780	2026	public	USA: NY\, Kings County	stool	Clostridium botulinum	ILLUMINA	B1
IDR2500091709-02	SAMN63663779	2025	public	USA: MI	stool	Clostridium botulinum	ILLUMINA	B1
IDR2500091709-01	SAMN63663778	2025	public	USA: MI	stool	Clostridium botulinum	ILLUMINA	B1
IDR2500090795-01-02	SAMN63663777	2025	public	USA: MA	stool	Clostridium botulinum	ILLUMINA	A2
IDR2500090794-01-02	SAMN63663776	2025	public	USA: MA	stool	Clostridium botulinum	ILLUMINA	A2

BoNTax: QC and ANI results

Sample	Total length	Contigs	N50	QC	AMR	VIRULENCE	ST	Reference match	ANI (%)	Species match
Cbot_gp1_CDC67190	3954777	1	3954777	Pass	cfr,cplR	bont_A1,bont_B5	4	Cbotulinum_group1_CDC67190	100	Clostridium botulinum group1
Cbot_gp1_Hall_80	3812159	64	175053	Pass	cfr,cplR	bont_B1	30	Cbotulinum_group1_Hall_80	100	Clostridium botulinum group1
Cbot_gp1_Langeland	3995387	1	3995387	Pass	cfr,cplR	bont_F1	88	Cbotulinum_group1_Langeland	100	Clostridium botulinum group1
Cbot_gp1_LochMareeA3	4259691	2	3992906	Pass	cfr,cplR	bont_A3	3	Cbotulinum_group1_LochMareeA3	100	Clostridium botulinum group1
Csporo_ATCC_15579	4102325	2	2710723	Pass	cfr,cplR	0	70	Csporogenes_ATCC_15579	100	Clostridium sporogenes
Csporo_DSM_795	4106655	16	722007	Pass	cplR	0	-	Csporogenes_DSM_795	100	Clostridium sporogenes
Csporo_Osaka05	4408349	3	4004744	Pass	cfr,cplR	bont_B6	35	Csporogenes_Osaka05	99.99	Clostridium sporogenes
Csporo_Prevot_594	4334551	2	4077214	Pass	cplR	bont_B2	54	Csporogenes_Prevot_594	100	Clostridium sporogenes
IDR2500090794-01-02	4622411	38	675493	Flagged	cfr,cplR	bont_A2	90	Csporogenes_CDC38839	98.99	Clostridium sporogenes
IDR2500090795-01-02	4630086	27	679098	Flagged	cfr,cplR	bont_A2	90	Csporogenes_CDC45459	98.96	Clostridium sporogenes
IDR2500091709-01	4113953	20	949626	Pass	cfr,cplR	bont_B1	34	Cbotulinum_group1_CDC64197	98.18	Clostridium botulinum group1
IDR2500091709-02	4112508	25	658983	Pass	cfr,cplR	bont_B1	34	Cbotulinum_group1_CDC64197	98.3	Clostridium botulinum group1
IDR2600008692-01-02	4297366	35	363923	Pass	cfr,cplR	bont_B1	96	Csporogenes_CDC31846	99.77	Clostridium sporogenes
IDR2600019603-01-02	4312369	34	289182	Pass	cfr,cplR	bont_B1	96	Csporogenes_CDC31846	99.98	Clostridium sporogenes
IDR2600022776-01-02	4297762	41	433935	Pass	cfr,cplR	bont_B1	96	Csporogenes_CDC31846	99.77	Clostridium sporogenes
IDR2600023425-01-02	4117996	41	366163	Pass	cfr,cplR	bont_B1	-	Cbotulinum_group1_Langeland	98.35	Clostridium botulinum group1
IDR2600031989-01-02	4067084	29	602142	Pass	cfr,cplR	bont_B7	34	Cbotulinum_group1_CDC64197	98.32	Clostridium botulinum group1
IDR2600036479-01-02	4293120	44	241293	Pass	cfr,cplR	bont_B1	96	Csporogenes_CDC31846	99.97	Clostridium sporogenes

ANI heatmap


BoNTax
92
94
96
98
100
C.sporogenes
C.botulinum group1
IDR2600036479_01_02
IDR2600019603_01_02
IDR2600008692_01_02
IDR2600022776_01_02
Csporo_ATCC_15579
IDR2500090795_01_02
IDR2500090794_01_02
Csporo_Osaka05
Csporo_Prevot_594
Csporo_DSM_795
IDR2500091709_02
IDR2500091709_01
IDR2600031989_01_02
IDR2600023425_01_02
Cbot_gp1_Langeland
Cbot_gp1_Hall_80
Cbot_gp1_CDC67190
Cbot_gp1_LochMareeA

Strengths and limitations

Strengths

Standardized genome-based species classification

Species identity and toxin profiles reported separately

Limitations

Results depend on reference coverage and assembly quality

Borderline ANI and QC flags require further review

Future directions

Reference coverage Expand to additional botulinum groups and newly described taxa.

Ambiguous cases Add phylogenomic analysis for borderline species assignments.

Surveillance links Explore matching to NCBI Pathogen Detection isolates and broader clostridial surveillance.

BoNTax: species identity and toxin profiles

Classify	Curated ANI-based species assignment
Characterize	Separate toxin profiles and genomic markers
Review	Integrated QC and results in GalaxyTrakr ensures consistent, reproducible analysis across external surveillance programs

Access BoNTax

GalaxyTrakr | GitHub: BoNTax | Step-by-step protocol

Acknowledgments

Dr. Narjol Gonzalez-Escalona
Dr. Shashi Sharma

Office of Applied Microbiology and Technology (OAMT)

GalaxyTrakr

Questions?