Preplanned Studies

Genomic Characterization of *Clostridium botulinum* Isolates from Soil and Soybean Samples in High-Incidence Regions — Xinjiang, Inner Mongolia, and Qinghai PLADs, China, 2024

Yongsi Zhan[1,2]; Xin Ma[3]; Xuebin Guo[4]; Meng Zhang[5]; Chunxia Cui[6]; Weiwei Li[2]; Shaofei Yan[2]; Shenghui Cui[7]; Xingfen Yang[1,#]; Yunchang Guo[1,2,#]

Summary

What is already known about this topic?

Foodborne botulism is prevalent in northwestern China, linked to traditional homemade foods. Recently, some cases have been linked to commercial vacuum-packaged ready-to-eat meat products. Soil is a potential contamination source, yet genomic information on environmental isolates from high-incidence regions remains scarce.

What is added by this report?

This study presents the first genomic characterization of 23 *C. botulinum* isolates obtained from soil and soybean samples in Northwest China. Four botulinum neurotoxin subtypes, A5(B3), B2, B3, and B4, were identified, each demonstrating notable geographic and metabolic diversity. Subtype-specific genomic adaptations, transposase insertions, and an incomplete prophage carrying *bont* in one isolate were observed, suggesting historical horizontal gene transfer.

What are the implications for public health practice?

Soils in high-incidence regions may act as persistent reservoirs of *C. botulinum*, emphasizing the need for targeted evidence-based public health interventions. Strengthening hygiene and sanitation practices during food processing, along with enhanced surveillance of both traditional and commercial food products, are essential to prevent future foodborne botulism outbreaks in endemic regions.

ABSTRACT

Introduction: *Clostridium botulinum* (*C. botulinum*) produces botulinum neurotoxins (BoNTs), the causative agents of botulism, a severe neuroparalytic disease prevalent in northwest China. Recent foodborne botulism outbreaks linked to commercially produced, vacuum-packaged, ready-to-eat foods were traced to raw meat contaminated with *C. botulinum* spores originating from soil, highlighting an emerging public health concern related to environmental reservoirs. However, genomic information on environmental isolates from high-incidence regions remains limited.

Methods: A total of 23 *C. botulinum* strains isolated from soil and soybean samples in northwest China were sequenced in 2024. Genomes were analyzed for plasmids, prophages, antibiotic resistance genes, virulence factors, and *bont*. Evolutionary relationships and adaptive features were investigated via phylogenetic and functional analyses.

Results: The 23 isolates were classified into four BoNT subtypes [A5(B3), B2, B3, B4] and clustered according to subtype and geographic origin. Isolates from Qinghai formed distinct branches. Functional annotation revealed subtype-specific metabolic variations, particularly in carbohydrate metabolism. Although all isolates contained conserved *bont* clusters, some exhibited transposase insertions. One subtype A5(B3) isolate harbored *bont* within an incomplete prophage.

Conclusion: These preliminary insights into environmental *C. botulinum* virulence, ecological adaptation, and evolutionary characteristics in northwest China provide a foundation for targeted surveillance and the development of preventive strategies against botulism in endemic regions.

Clostridium botulinum (*C. botulinum*) is a Gram-positive, spore-forming, and anaerobic bacillus that is ubiquitously distributed in soils, aquatic sediments, and animal feces, posing a potential risk for foodborne and environmental exposure. Botulism is a severe neuroparalytic disease caused by botulinum neurotoxin (BoNT) produced by *C. botulinum*. BoNTs are among

the most potent biological toxins and are classified into serotypes A–G, with types A, B, E, and F primarily associated with human disease (*1*).

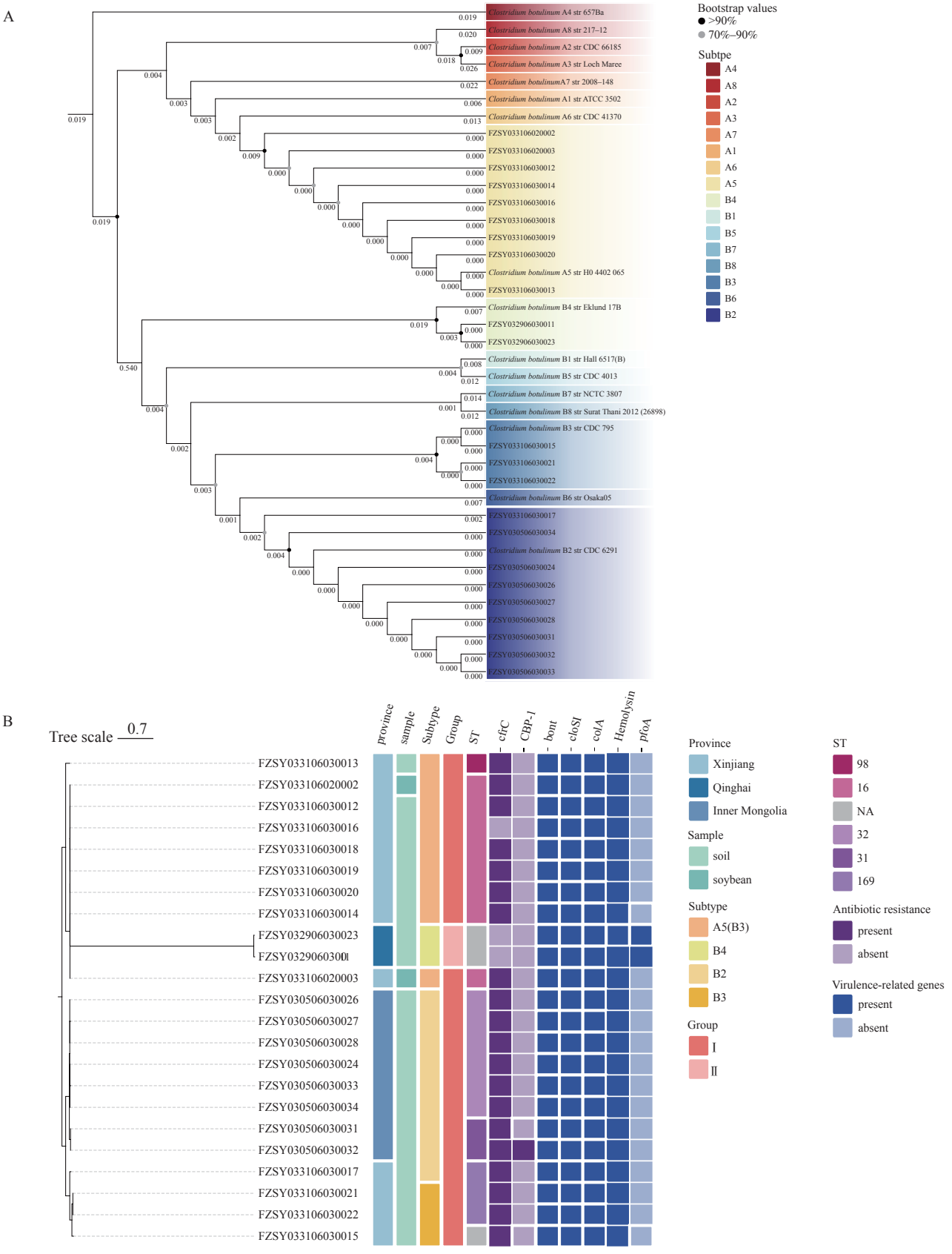
In China, foodborne botulism (FB) exhibits a distinct geographical distribution, with a higher prevalence in northwestern provincial-level administrative divisions (PLADs) such as Xinjiang, Inner Mongolia, and Qinghai. Traditional dietary habits and local environmental conditions promote *C. botulinum* proliferation, and outbreaks are frequently associated with consumption of homemade fermented soybean products and dried meat contaminated with soil-derived spores (*2*–*4*). Additionally, recent FB outbreaks linked to commercial vacuum-packaged ready-to-eat foods were likely caused by contamination of raw meat with *C. botulinum* spores in soil, highlighting an emerging public health concern associated with environmental reservoirs (*5*–*6*). Most previous genomic studies focused on clinical or food isolates, whereas data on environmental isolates from high-incidence regions in China are scarce.

Here, we sequenced 23 *C. botulinum* isolates from soil and soybean samples from high-incidence regions in Northwest China and compared their genetic diversity, evolutionary dynamics, and virulence potential of reservoirs linked to human diseases. This study is critical for assessing public health risks and tailoring region-specific preventive strategies.

Twenty-three *C. botulinum* isolates were collected from Xinjiang (13 isolates, including 11 from soil and two from soybean samples), Inner Mongolia (8 isolates from soil), and Qinghai PLADs (2 isolates from soil) in 2024 (Supplementary Table S1, available at https://weekly.chinacdc.cn/). Genomic DNA was extracted using a Genomic DNA Purification Kit (Promega, Madison, WI, USA). Sequencing was performed by Beijing Novogene Bioinformatics Technology Co., Ltd. (Beijing, China) on an Illumina HiSeq platform (San Diego, CA, USA; 150-bp paired-end; 100× coverage depth). The quality of raw reads was assessed using FastQC (version 0.11.9). Clean reads were assembled *de novo* using SPAdes (version 4.1.0). Assembly quality was evaluated using QUAST (version 5.3.0; Algorithmic Biology Lab, St. Petersburg, Russia). Genome annotation was performed using Prokka (version 1.14.6; University of Melbourne, Melbourne, Australia). Plasmid sequences were identified using PlasmidFinder. A phylogenetic tree of *bont* was constructed using the neighbor-joining method in MEGA (version 11.0; Pennsylvania State University, State College, PA, USA), incorporating sequences from the isolates in this study and representative reference strains of subtypes A and B retrieved from GenBank (Supplementary Table S2, available at https://weekly.chinacdc.cn/). Core genome single-nucleotide polymorphisms (cgSNPs) were identified using Snippy (version 4.6.0; University of Melbourne, Melbourne, Australia). A maximum likelihood tree was generated using FastTree (version 2.1.10, San Diego, CA, USA) on the Galaxy platform (*7*). Sequence types were assigned using the PubMLST database. Phylogenetic trees and figures were generated and visualized using the ChiPlot web server (*8*). Predicted proteins were functionally annotated using the Clusters of Orthologous Genes (COG; https://ngdc.cncb.ac.cn/databasecommons/database/id/37) database, and principal component analysis of the COG category distributions was performed using R (version 4.4.0; The R Project for Statistical Computing, Vienna, Austria). Virulence genes in the Virulence Factor Database (https://www.mgc.ac.cn/VFs/) were identified using BLASTn. Antibiotic-resistance genes were detected using The Comprehensive Antibiotic Resistance Database (https://card.mcmaster.ca/). The 10-kb genomic regions flanking *bont* were compared using Easyfig (version 2.2.5, Brisbane, Australia). Prophage regions were predicted using the PHASTER web server.

The genomes of 23 isolates sequenced and assembled *de novo* showed sizes of 3.79–4.23 Mb and GC contents of 27.2%–28.3% (Supplementary Table S3, available at https://weekly.chinacdc.cn/). PlasmidFinder revealed no plasmids. Phylogenetic analyses based on *bont* genes, cgSNPs, and multilocus sequence typing consistently revealed that the 23 isolates clustered primarily according to subtype and sequence type. This included nine subtype A5(B3) and three subtype B3 isolates from Xinjiang; eight subtype B2 isolates from Inner Mongolia and one from Xinjiang; and two subtype B4 isolates from Qinghai (Figure 1). Isolates from Xinjiang and Inner Mongolia formed a closely related genetic cluster (Group I), whereas Qinghai isolates constituted a distinct phylogenetic branch (Group II). Notably, isolates FZSY033106030013 [A5(B3)], FZSY033106030017 (B2), and FZSY033106030015 (B3) exhibited unique sequence types and greater genetic distances from others within the same subtype, suggesting microevolution or distinct ancestral origins.

Functional annotation assigned the predicted proteins to 23 categories. Excluding proteins with


A
Clostridium botulinum A4 str 657Ba
Clostridium botulinum A8 str 217–12
Clostridium botulinum A2 str CDC 66185
Clostridium botulinum A3 str Loch Maree
Clostridium botulinumA7 str 2008–148
Clostridium botulinum A1 str ATCC 3502
Clostridium botulinum A6 str CDC 41370
FZSY033106020002
FZSY033106020003
FZSY033106030012
FZSY033106030014
FZSY033106030016
FZSY033106030018
FZSY033106030019
FZSY033106030020
Clostridium botulinum A5 str H0 4402 065
FZSY033106030013
Clostridium botulinum B4 str Eklund 17B
FZSY032906030011
FZSY032906030023
Clostridium botulinum B1 str Hall 6517(B)
Clostridium botulinum B5 str CDC 4013
Clostridium botulinum B7 str NCTC 3807
Clostridium botulinum B8 str Surat Thani 2012 (26898)
Clostridium botulinum B3 str CDC 795
FZSY033106030015
FZSY033106030021
FZSY033106030022
Clostridium botulinum B6 str Osaka05
FZSY033106030017
FZSY030506030034
Clostridium botulinum B2 str CDC 6291
FZSY030506030024
FZSY030506030026
FZSY030506030027
FZSY030506030028
FZSY030506030031
FZSY030506030032
FZSY030506030033
Bootstrap values
>90%
70%–90%
Subtpe
A4
A8
A2
A3
A7
A1
A6
A5
B4
B1
B5
B7
B8
B3
B6
B2
B
Tree scale 0.7
province
sample
Subtype
Group
ST
cftC
CBP-1
bont
closI
colA
Hemolysin
pfoA
FZSY033106030013
FZSY033106020002
FZSY033106030012
FZSY033106030016
FZSY033106030018
FZSY033106030019
FZSY033106030020
FZSY033106030014
FZSY032906030023
FZSY032906030011
FZSY033106020003
FZSY030506030026
FZSY030506030027
FZSY030506030028
FZSY030506030024
FZSY030506030033
FZSY030506030034
FZSY030506030031
FZSY030506030032
FZSY033106030017
FZSY033106030021
FZSY033106030022
FZSY033106030015
Province
Xinjiang
Qinghai
Inner Mongolia
Sample
soil
soybean
Subtype
A5(B3)
B4
B2
B3
Group
I
II
ST
98
16
NA
32
31
169
Antibiotic resistance
present
absent
Virulence-related genes
present
absent

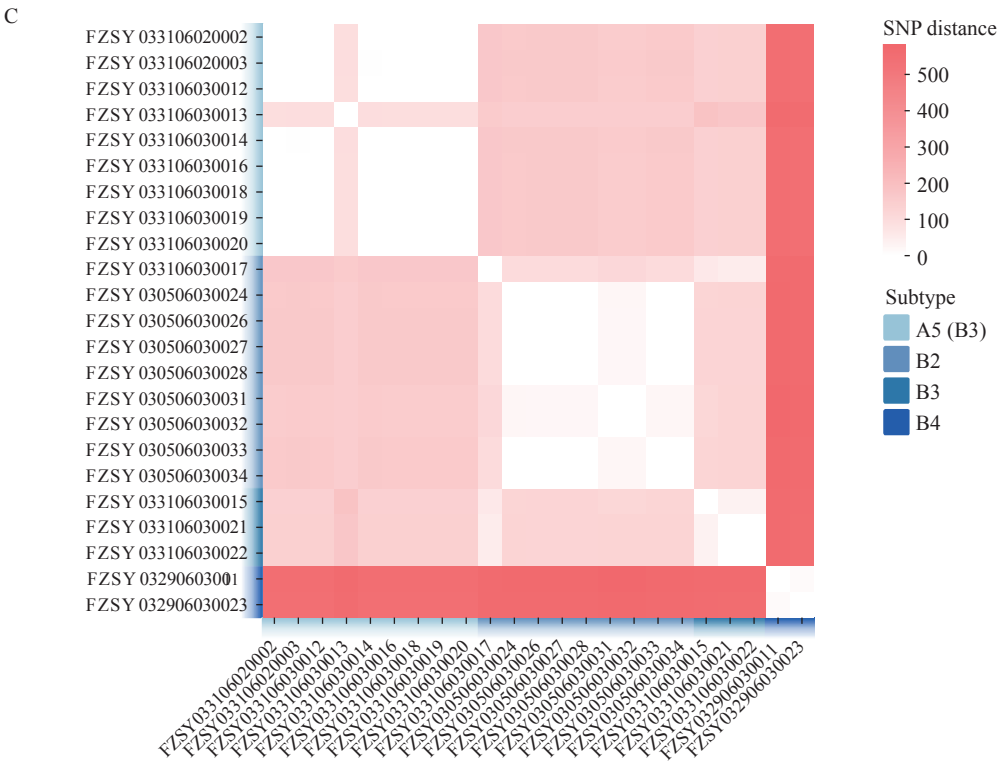
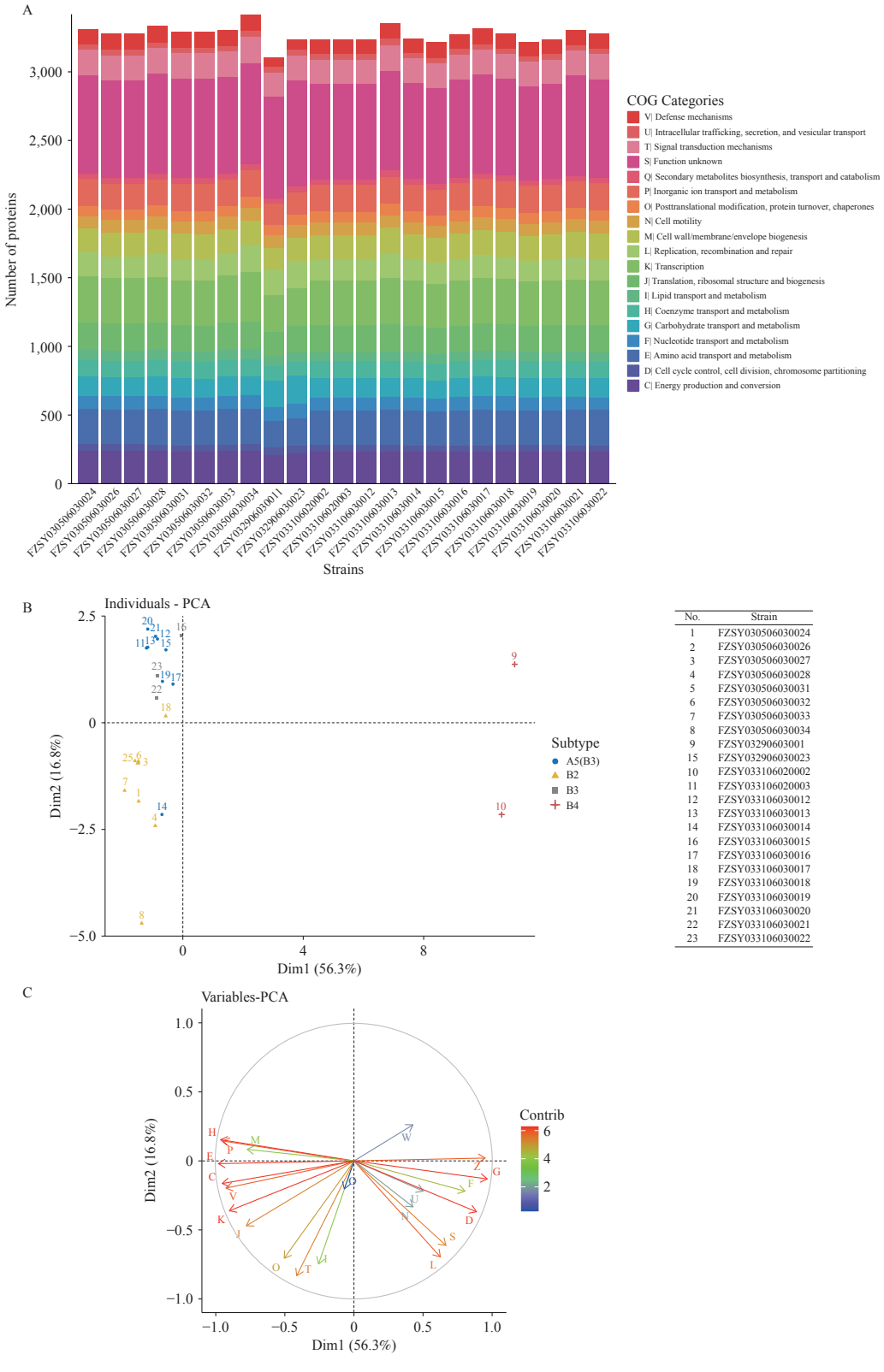


FIGURE 1. Phylogenetic analysis of 23 *Clostridium botulinum* isolates from China. (A) Neighbor-joining phylogenetic tree based on *bont* nucleotide sequences. (B) Core-genome SNP-based maximum-likelihood phylogenetic tree with a heatmap showing PLAD, sample type, subtype, multilocus sequence type, group, virulence factors, and antibiotic resistance genes. (C) SNP distance matrix illustrating pairwise genetic distances among isolates, with color gradients indicating SNP differences and subtypes.
Note: For (A) Bootstrap values (1,000 replicates) are indicated, with colors representing subtypes.
Abbreviation: SNP=single-nucleotide polymorphism; PLAD=provincial-level administrative division.

unknown functions, the most abundant categories were transcription, amino acid transport and metabolism, and cell cycle control, cell division, and chromosome partitioning (Figure 2A). Principal component analysis based on COG annotations revealed that the 23 isolates clustered into three major groups. A5(B3) and B3 exhibited functional similarities and were grouped, whereas B2 and B4 formed distinct clusters (Figure 2B). This separation was mainly driven by the categories carbohydrate transport and metabolism, transcription, and amino acid transport and metabolism (Figure 2C). Importantly, FZSY033106030013 [A5(B3)] was positioned closer to the B2 cluster, whereas FZSY033106030017 (B2) was closer to the B3 cluster.

Five virulence factors and two antibiotic-resistance genes were identified across the isolates (Figure 1B). All isolates carried *bont*, *cloSI*, *colA*, and hemolysin, whereas *pfoA* was detected only in B4 isolates. *cfrC*, encoding resistance to the antibiotics phenicol, oxazolidinone, lincosamide, and streptogramin, was present in 20 isolates but absent from both B4 isolates and one A5(B3) isolate (FZSY033106030016); CBP-1, encoding resistance to penicillin β-lactam antibiotics, was found exclusively in the B2 isolate FZSY030506030032.

Comparative analysis of the 10-kb flanking regions upstream and downstream of *bont* revealed a conserved *ha70–ha17–ha33–botR–ntnh–bont* cluster in all isolates (Figure 3). A5(B3) isolates harbored complete *bont/A5* and a truncated *bont/B3*. Two distinct gene contexts occurred within the A5(B3), B2, and B3 subtypes, with FZSY033106030013 [A5(B3)], FZSY033106030017 (B2), and FZSY033106030015 (B3) showing arrangements that differed from those of other isolates of the same subtype (Figure 3). Specifically, FZSY033106030013 [A5(B3)] contained a transposase gene and divergent downstream region, whereas FZSY033106030017 (B2) harbored a transposase gene upstream of the cluster. PHASTER analysis identified prophage sequences in all 23 genomes (Supplementary Table S4, available at



No.	Strain
1	FZSY030506030024
2	FZSY030506030026
3	FZSY030506030027
4	FZSY030506030028
5	FZSY030506030031
6	FZSY030506030032
7	FZSY030506030033
8	FZSY030506030034
9	FZSY03290603001
15	FZSY032906030023
10	FZSY033106020002
11	FZSY033106020003
12	FZSY033106030012
13	FZSY033106030013
14	FZSY033106030014
16	FZSY033106030015
17	FZSY033106030016
18	FZSY033106030017
19	FZSY033106030018
20	FZSY033106030019
21	FZSY033106030020
22	FZSY033106030021
23	FZSY033106030022

FIGURE 2. Functional annotation and PCA clustering of 23 *Clostridium botulinum* isolates based on COG categories. (A) Stacked bar chart showing the distribution of predicted proteins across 23 isolates, with colors representing COG categories. (B) PCA scatter plot of isolates along the first two principal components, with points colored by subtype. (C) PCA correlation circle plot of COG categories as variables, where arrow length and direction indicate correlation strength and sign, and color intensity indicates contribution magnitude.
Abbreviation: PCA=principal component analysis; COG=clusters of orthologous genes.

https://weekly.chinacdc.cn/), with only one isolate, FZSY033106030013 [A5(B3)], containing an incomplete prophage element that carried *bont* (Figure 4).

DISCUSSION

In China, FB cases are predominantly reported north of 30°N, with clear regional variations in the

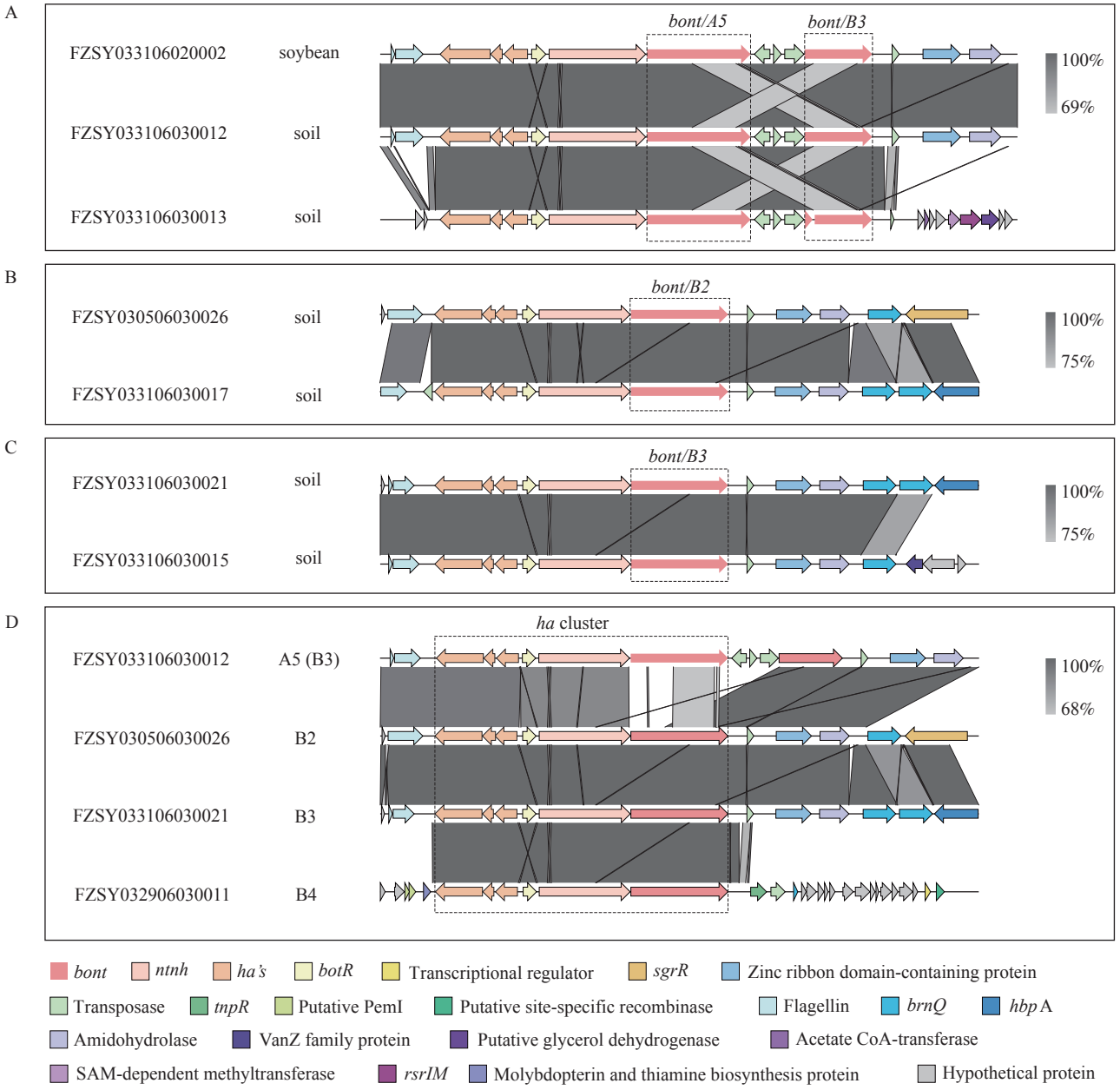


FIGURE 3. Comparative genomic context of the 10 kb upstream and downstream flanking regions of *bont* cluster in 23 *Clostridium botulinum* isolates. (A) Two distinct gene contexts of subtype A5(B3) isolates. (B) Two distinct gene contexts of subtype B2 isolates. (C) Two distinct gene contexts of subtype B3 isolates. (D) Representative gene contexts of subtypes A5(B3), B2, B3, and B4.

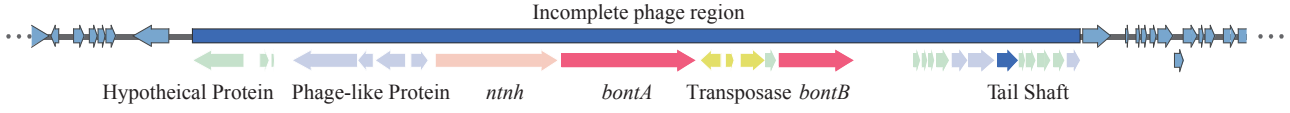


FIGURE 4. Incomplete prophage element carrying *bont* in the A5(B3) isolate FZSY033106030013.

causative foods (*9*). Homemade fermented soybean products are the primary vehicles in Xinjiang and Inner Mongolia, whereas homemade dried meats are frequently consumed in the Qinghai Plateau (*2*–*3*). Notably, recent cases of botulism were linked to commercial vacuum-packed meat and meat products, reflecting a growing concern alongside traditional sources (*5*–*6*). These products may be contaminated with soil-derived *C. botulinum* spores present on raw meat, with anaerobic packaging facilitating spore germination and toxin production. Thus, studies are needed to explore the genomic characteristics of *C. botulinum* in high-risk regions.

We collected 23 *C. botulinum* isolates from soil and soybean samples obtained from three Chinese PLADs associated with a high incidence of botulism. Phylogenetic analyses based on *bont* and cgSNPs revealed four subtypes [A5(B3), B2, B3, and B4] with distinct subtype-specific clusters. Isolates from Xinjiang and Inner Mongolia were genetically distant from those from Qinghai, suggesting geographic differentiation and adaptation to distinct ecological niches influenced by high-altitude, low-oxygen environments (*10*). Interestingly, atypical sequence types and unexpected cgSNP distances in several isolates indicate recombination or horizontal gene transfer. Importantly, the predominance of BoNT/A5(B3) in soils was consistent with the serotypes reported in historical FB cases from the same regions (*11*), suggesting that local soils act as environmental reservoirs, contributing to soil-to-food transmission. This concordance highlights the need for stricter hygiene practices in food processing, particularly during the traditional fermentation and preservation of soybean products and meat in these high-risk regions.

Functional annotation further suggested niche-specific metabolic adaptations, particularly differences in carbohydrate metabolism between subtypes, which may influence their persistence in local environments (*12*). The functional proximity of some A5(B3) and B2 isolates to other subtypes further suggested shared metabolic traits or transitional evolutionary states, possibly facilitated by genetic exchange. These ecological features, coupled with the conserved *bont* cluster and sporadic acquisition of resistance genes, support the evolutionary stability and genomic plasticity of environmental isolates (*13*). Moreover, identification of an incomplete prophage carrying *bont* suggests historical phage-mediated dissemination of toxin genes, although its current mobility remains uncertain (*14*).

This study had several limitations. First, the relatively small number of strains from the three PLADs may not fully represent the genetic diversity of *C. botulinum* across endemic regions in China. In addition, the lack of paired food and clinical isolates from outbreaks limits our ability to establish direct transmission pathways from soil to food. Finally, reliance on genomic data without phenotypic validation restricts inferences regarding toxin expression.

We performed genomic characterization of *C. botulinum* toxin subtypes from soil reservoirs in Northwest China, revealing their genomic diversity and potential ecological adaptations in high-incidence regions. Tailored surveillance and preventive strategies are needed to mitigate foodborne botulism in traditional and industrial settings.

Conflicts of interest: No conflicts of interest.

Funding: Supported by National Key Research and Development Program of China (2022YFC2303900) and Henan Medical Science and Technique Foundation (SBGJ202402091).

doi: 10.46234/ccdcw2026.038

[#] Corresponding authors: Yunchang Guo, gych@cfsa.net.cn; Xingfen Yang, yangalice79@smu.edu.cn.

[1] Food Safety and Health Research Center, School of Public Health, Southern Medical University, Guangzhou City, Guangdong Province, China; [2] China National Center for Food Safety Risk Assessment, Beijing, China; [3] Xinjiang Uygur Autonomous Region Center for Disease Control and Prevention, Urumqi City, Xinjiang Uygur Autonomous Region, China; [4] Qinghai Province Center for Disease Control and Prevention, Xining City, Qinghai Province, China; [5] Henan Province Center for Disease Control and Prevention, Zhengzhou City, Henan Province, China; [6] Inner Mongolia Center for Disease Control and Prevention, Hohhot City, Inner Mongolia Autonomous Region, China; [7] China National Institute for Food and Drug Control, Beijing, China.

Submitted: September 10, 2025
Accepted: January 12, 2026
Issued: February 27, 2026

REFERENCES

1. Rawson AM, Dempster AW, Humphreys CM, Minton NP. Pathogenicity and virulence of *Clostridium botulinum*. Virulence 2023;14(1):2205251. https://doi.org/10.1080/21505594.2023.2205251.
2. Li HQ, Guo YC, Tian T, Guo WH, Liu CQ, Liang XC, et al. Epidemiological analysis of foodborne botulism outbreaks - China, 2004-2020. China CDC Wkly 2022;4(35):788 – 92. https://doi.org/10.46234/ccdcw2022.114.

3. Wang YL, Guo XB. Epidemiological analysis of foodborne botulism in Qinghai Province from 1959 to 2022. Chin J Food Hyg 2024;36(2): 207 – 11. https://doi.org/10.13590/j.cjfh.2024.02.015.
4. Wang LJ, Li KC, Qian SY, Gao HM, Liu J, Li Z, et al. Clinical characteristics and prognosis of 8 cases of severe infant botulism. Chin J Pediatr 2024;62(3):218 – 22. https://doi.org/10.3760/cma.j.cn112140-20230908-00169.
5. Min M, Bai LL, Peng XB, Guo L, Wan K, Qiu ZW. An outbreak of botulinum types A, B, and E associated with vacuum-packaged salted fish and ham. J Emerg Med 2021;60(6):760 – 3. https://doi.org/10.1016/j.jemermed.2020.12.006.
6. Cui W, Ma CM, Liu M, Li Y, Zhou L, Shi YW, et al. Foodborne botulism caused by *Clostridium botulinum* subtype A5(b3) by self-packaged vacuum spicy rabbit heads. Microorganisms 2025;13(7):1662. https://doi.org/10.3390/microorganisms13071662.
7. Blankenberg D, Coraor N, Von Kuster G, Taylor J, Nekrutenko A, Galaxy T. Integrating diverse databases into an unified analysis framework: a Galaxy approach. Database (Oxford) 2011;2011:bar011. https://doi.org/10.1093/database/bar011.
8. Xie JM, Chen YR, Cai GJ, Cai RL, Hu Z, Wang H. Tree Visualization By One Table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. Nucleic Acids Res 2023;51(W1): W587 – 92. https://doi.org/10.1093/nar/gkad359.
9. Gao QY, Liu HD, Huang YF, Wu JG, Xia HQ. Foodborne botulism in China. Chin J Food Hyg 1989;(3):45-50. http://dx.doi.org/10.13590/j.cjfh.1989.03.021. (In Chinese).
10. Fu SW, Wang CH. An overview of type E botulism in China. Biomed Environ Sci 2008;21(4):353 – 6. https://doi.org/10.1016/S0895-3988(08)60054-9.
11. Shi Y, Zhao SY. Botulism in China. Mod Med J 1986;14(3):179-81. https://d.wanfangdata.com.cn/periodical/CiBQZXJpb2RpY2FsQ0hJU29scjkyMDI1MTIyNDE1NDU1NRINdGR5eDE5ODYwMzAyMhoIbzl1NjJoa2E%3D. (In Chinese).
12. Li GM, Wang HW, Zhang Y, Yan J, Duan ZW, Pang L, et al. Genomic characterisation and traceability analysis of a *Clostridium botulinum* strain involved in a food poisoning incident. BMC Infect Dis 2025;25(1):323. https://doi.org/10.1186/s12879-025-10700-4.
13. Ma X, Li KX, Li F, Su J, Meng WW, Sun YM, et al. Tracing foodborne botulism events caused by *Clostridium botulinum* in Xinjiang Province, China, using a core genome sequence typing scheme. Microbiol Spectr 2022;10(6):e0116422. https://doi.org/10.1128/spectrum.01164-22.
14. Smith TJ, Hill KK, Raphael BH. Historical and current perspectives on *Clostridium botulinum* diversity. Res Microbiol 2015;166(4):290 – 302. https://doi.org/10.1016/j.resmic.2014.09.007.

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY TABLE S1. 23 *C. botulinum* isolates in this study.

Isolate ID	Origin	PLAD	Year	Subtype
FZSY033106020002	soybean	Xinjiang	2024	A5(B3)
FZSY033106020003	soybean	Xinjiang	2024	A5(B3)
FZSY033106030012	soil	Xinjiang	2024	A5(B3)
FZSY033106030014	soil	Xinjiang	2024	A5(B3)
FZSY033106030016	soil	Xinjiang	2024	A5(B3)
FZSY033106030018	soil	Xinjiang	2024	A5(B3)
FZSY033106030019	soil	Xinjiang	2024	A5(B3)
FZSY033106030020	soil	Xinjiang	2024	A5(B3)
FZSY033106030013	soil	Xinjiang	2024	A5(B3)
FZSY033106030017	soil	Xinjiang	2024	B2
FZSY030506030024	soil	Inner Mongolia	2024	B2
FZSY030506030026	soil	Inner Mongolia	2024	B2
FZSY030506030027	soil	Inner Mongolia	2024	B2
FZSY030506030028	soil	Inner Mongolia	2024	B2
FZSY030506030031	soil	Inner Mongolia	2024	B2
FZSY030506030032	soil	Inner Mongolia	2024	B2
FZSY030506030033	soil	Inner Mongolia	2024	B2
FZSY030506030034	soil	Inner Mongolia	2024	B2
FZSY033106030015	soil	Xinjiang	2024	B3
FZSY033106030021	soil	Xinjiang	2024	B3
FZSY033106030022	soil	Xinjiang	2024	B3
FZSY032906030011	soil	Qinghai	2024	B4
FZSY032906030023	soil	Qinghai	2024	B4

SUPPLEMENTARY TABLE S2. Reference *C. botulinum* genomes used for phylogenetic analysis.

Type	Name	Accession
A1	*Clostridium botulinum* str. ATCC 3502	NC_009495
A2	*Clostridium botulinum* str. CDC 66185	KM875565
A3	*Clostridium botulinum* str. Loch Maree	ABA29017
A4	*Clostridium botulinum* str. 657Ba	EU341307
A5	*Clostridium botulinum* str. H04402065	NC_017299
A6	*Clostridium botulinum* str. CDC 41370	FJ981696
A7	*Clostridium botulinum* str. 2008-148	JQ954969
A8	*Clostridium botulinum* str. 217-12	KF667385
B1	*Clostridium botulinum* str. Hall 6517(B)	EF028399
B2	*Clostridium botulinum* str. CDC 6291	EF028401
B3	*Clostridium botulinum* str. CDC 795	EF028400
B4	*Clostridium botulinum* str. Eklund 17B	X71343
B5	*Clostridium botulinum* str. CDC 4013	GU271943
B6	*Clostridium botulinum* str. Osaka05	AB302852
B7	*Clostridium botulinum* str. NCTC 3807	JN120760
B8	*Clostridium botulinum* str. Surat Thani 2012(26898)	KC714045

SUPPLEMENTARY TABLE S3. Assembly quality metrics of the 23 *C. botulinum* isolates.

Isolate ID	# contigs	Largest contig	Total length	GC (%)	N50	N90	auN	L50	L90	# N's per 100 kbp
FZSY033106020002	22	1,064,051	3,820,350	27.96	956,595	296,752	785,899.1	2	5	7.54
FZSY033106020003	26	980,537	3,816,116	27.96	948,223	140,245	639,391.4	2	7	7.49
FZSY033106030012	38	1,063,744	3,820,004	27.96	450,112	91,876	494,569.3	3	12	12.46
FZSY033106030014	30	955,337	3,841,179	27.94	542,381	91,939	533,920.8	3	9	17.6
FZSY033106030016	34	613,154	4,036,687	27.78	456,366	117,664	409,258.5	4	10	2.40
FZSY033106030018	27	1,063,876	4,033,829	27.81	957,218	140,375	688,982.3	2	7	11.80
FZSY033106030019	91	389,991	3,798,600	27.97	99,889	19,645	120,484.4	13	42	2.58
FZSY033106030020	25	1,063,877	3,818,330	27.96	956,713	140,335	723,286.4	2	6	12.54
FZSY033106030013	59	657,641	4,144,238	27.92	294,433	80,006	350,612.4	5	15	27.97
FZSY033106030017	20	1,619,332	3,954,786	27.93	763,374	359,843	960,363.8	2	5	12.21
FZSY030506030024	23	1,632,096	3,919,871	28.04	694,263	331,506	947,743.1	2	5	4.92
FZSY030506030026	31	1,622,688	3,899,509	28.08	697,890	102,244	928,959.6	2	6	14.95
FZSY030506030027	20	2,060,240	3,885,110	28.03	2,060,240	331,506	1,313,771.1	1	4	12.41
FZSY030506030028	22	1,622,773	4,093,349	27.90	734,398	208,608	910,252.0	2	6	7.11
FZSY030506030031	24	1,559,889	3,868,044	28.02	613,544	140,522	872,302.2	2	6	2.48
FZSY030506030032	26	1,560,066	3,868,479	28.02	575,595	140,522	861,060.0	2	6	5.01
FZSY030506030033	19	2,085,851	3,897,311	28.04	2,085,851	331,506	1,331,857.4	1	4	4.98
FZSY030506030034	80	2,135,297	4,230,770	28.29	2,135,297	109,580	1,227,985.7	1	7	7.00
FZSY033106030015	22	2,069,139	3,786,509	27.99	2,069,139	288,694	1,338,038.6	1	4	7.61
FZSY033106030021	32	1,121,985	4,041,323	27.83	551,067	96,717	578,130.9	3	9	16.65
FZSY033106030022	20	1,931,186	3,875,983	28.00	1,246,838	414,779	1,413,864.6	2	3	12.44
FZSY032906030011	32	675,997	3,872,241	27.16	358,211	91,935	402,042.0	4	11	4.86
FZSY032906030023	11	1,759,473	3,945,336	27.36	1,109,829	797,469	1,262,569.9	2	3	5.45

SUPPLEMENTARY TABLE S4. Summary of prophage prediction across 23 isolates.

strain	prophage regions	intact	incomplete	questionable
FZSY030506030024	5	2	2	1
FZSY030506030026	6	1	3	2
FZSY030506030027	5	1	2	2
FZSY030506030028	7	2	4	1
FZSY030506030031	3	1	2	0
FZSY030506030032	3	1	2	0
FZSY030506030033	3	1	1	1
FZSY030506030034	4	1	1	2
FZSY032906030011	7	2	3	2
FZSY032906030023	8	3	3	2
FZSY033106020002	2	1	1	0
FZSY033106020003	2	1	1	0
FZSY033106030012	2	0	1	1
FZSY033106030013	3	2	1	0
FZSY033106030014	3	1	1	1
FZSY033106030015	2	1	0	1
FZSY033106030016	4	2	0	2
FZSY033106030017	4	2	2	0
FZSY033106030018	5	0	4	1
FZSY033106030019	1	1	0	0
FZSY033106030020	2	0	1	1
FZSY033106030021	7	1	5	1
FZSY033106030022	5	2	3	0