

厚生労働科学研究費補助金（新型インフルエンザ等新興・再興感染症研究事業
（新興・再興感染症及び予防接種政策推進研究事業））

分担研究報告書

超高速病原体ゲノム解読システムの構築と包括的な核酸迅速診断法の確立

研究分担者	黒田 誠	国立感染症研究所 病原体ゲノム解析研究センター センター長
研究協力者	関塚剛史	国立感染症研究所 病原体ゲノム解析研究センター第三室
	山下明史	国立感染症研究所 病原体ゲノム解析研究センター第三室
	見理 剛	国立感染症研究所 細菌第二部

研究要旨

未知病原体や変異病原体による新興感染症の汎発流行、そしてそれらを利用したバイオテロなどの危険性は近年社会的不安の一つとして認識されつつある。その危険性に対し的確な対処法を立案・整備する上で、バイオテロ病原体を網羅的かつ迅速に配列解読することは最も確かなアプローチの一つと考える。次世代ゲノムシーケンサー（Next-generation DNA sequencer: NGS）のパフォーマンスを用いて WHO 指定バイオテロ病原体のゲノム配列及び変異情報データベースを充実させ、有事において迅速に対応出来る体制を整えることを本研究課題は目標としている。

炭疽菌 (*Bacillus anthracis*) は炭疽症の原因となる細菌で、第二次世界大戦以降、生物兵器として各国の軍事機関に研究され、1993 年のオウム真理教によるテロ未遂や 2001 年のアメリカのテロ事件にも利用された事がある。そのため、炭疽症のアウトブレイクが起きた際に、原因菌の出自を詳細に調査することはバイオセキュリティ上の重要な意味を持つ。出自を調査するにはコアゲノムを用いた分子系統解析を行うのが有効であると考えられるが、コアゲノム情報を用いた分子系統解析を容易に行えるツールは存在していなかった。そのため我々は *B. anthracis* の次世代シーケンス (next-generation sequencing: NGS) データをアップロードすると single nucleotide polymorphisms (SNPs) を抽出し、*B. anthracis* を含む *B. cereus* グループ中の系統的位置関係を推定するウェブアプリケーション GcoGSA-BA (Global core Genome SNP Analysis for *Bacillus anthracis*) を開発した。GcoGSA-BA はコアゲノム系統解析だけでなく、pX01、pX02 といった病原性を規定するプラスミドや pX01 上に存在する三つの炭疽菌毒素遺伝子 (lethal factor, LF; edema factor, EF; protective antigen, PA) を検出する機能も有しており、炭疽菌 (*B. anthracis* Ames Ancestor) ばかりでなく炭疽菌毒素遺伝子を例外的に保持している *B. cereus* G9241 の系統的位置関係も正しく推定し、更に炭疽菌毒素遺伝子を検出することができた。テロを疑われる炭疽症のアウトブレイクが起きた際に GcoGSA-BA を用いることで、バイオテロリズムのために用意された株なのか、環境由来の病原菌なのかを迅速かつ容易に判断することができ、国内外のバイオセキュリティに大きく貢献することができると考えられる。順次、他カテゴリー A 病原体であるペスト菌、野兔病菌、コクシエラ、類鼻疽菌、ボツリヌス菌のゲノム情報を用いた GcoGSA を展開していく予定である。

A . 研究目的

未知病原体や変異病原体による新興感染症の汎発流行、そしてそれらを利用したバイオテロなどの危険性は近年社会的不安の一つとして認識されている。最先端技術を駆使した次世代ゲノムシーケンサーにより、今までは数年を要した全ゲノム解読が数週間で終了することが可能となった。最先端の革新技術を応用し、効率のかつ安定的に病原体検査システムを運用する体制を整え、WHO 指定バイオテロ病原菌および未知病原体をも検査対象とする網羅的解析法を構築することを目的としている。

B . 研究方法

国立感染症研究所・病原体ゲノム解析研究センターで構築した網羅的病原体検査法の情報解析のみをパイプライン化し、インタラクティブに誰でも利用できるソフトを開発した。詳細は研究結果を参照。

(倫理面への配慮)

該当なし

C . 研究結果

1) ネットワーク経路によるバイオテロ病原体検査のための解析システム

現在、感染研・ゲノムセンターにベンチトップ型・次世代シーケンサーと情報解析サーバーが整備されている。感染研に臨床検体が到着後、数日で解析・検査結果を報告できるようにはなっているものの、感染研に検体が送付されるまでには現場で数多くの微生物検査等が行われた後になってしまうケースが少なくない。このような現状の中、各地方自治体にも本研究課題で構築したシステムを導入して頂き、検査体制の一助となるのであれば非常に有効かつ迅速なバイオテロ対策に資するものと考えている。

炭疽菌が候補として浮上した場合、コアゲノム SNPs を利用した炭疽菌・菌株の類縁関係を推定するためのゲノムワイド SNPs 解析システム Global core genome SNP analysis for *Bacillus anthracis* (GcoGSA-BA) を開発した(図1と2)。炭疽菌ゲノム 5.23 Mb 全体に渡り特徴的な塩基アレル 657,183 箇所を用いたゲ

ノム分子系統樹を作成できる(図3)。従来の MLST 法では分解能が非常に悪く、由来特定に難渋した菌株においても高精度に分類することを可能にした。

現在、関係者のみ運用可能としている。GcoGSA-BA は、公開されている炭疽菌ゲノムと比較したゲノムワイド系統樹の生データまで作成し、適当な系統樹ビューワーで閲覧可能なデータのダウンロードを可能にする(図2)。炭疽菌は *Bacillus cereus* group に属し、セレウス菌との鑑別間違いを起こさぬよう、Lethal factor (LF), Edema factor (EF), Protective antigen (PA) の炭疽菌の病原性に必須な毒素因子の特定も可能にした(図2)。

2) 国内分離ボツリヌス菌のゲノム分子疫学解析

国内分離ボツリヌス菌のゲノム分子疫学解析を行った(感染研・細菌第二部・見理先生との共同研究)。公開ゲノム配列とのコアゲノム比較解析から、144,896 箇所の SNPs サイトを抽出し、ゲノム分子系統樹を作成した(図5)。従来の分子疫学解析を精度よく評価することに成功し、茨城県と岩手県に分離県と年代が異なる分離株においても由来特定と同一性を議論する基盤情報を収集解析することができた。

D/E . 考 察・結 論

これまで分担研究として、WHO 指定バイオテロ病原体の配列データベース化を進めてきた。ゲノム情報を活用することにより有効なトレーサビリティに役立つ目的である。構築データベースを有効に活用するためには、迅速な解読リード配列の取得が望ましく、そのためには多くの難題が残っている。本年度の課題として、情報解析の NGS - MePIC - MEGAN - GcoGSA の解読・解析パイプラインの運用と公開を目指した。検査現場からのデータ転送等、迅速にゲノム比較解析を行うために未だ解消されていないボトルネックが残っているが、誰もが簡便に利用できるシステムが先行していけば、シーケンサー等のインフラ整備が後々追いついてくるものと考えている。

F . 健康危険情報

特になし

G . 研究発表

1) 論文発表

- Kenri T, Sekizuka T, Yamamoto A, Iwaki M, Komiya T, Hatakeyama T, Nakajima H, Takahashi M, Kuroda M, Shibayama K. Genetic characterization and comparison of *Clostridium botulinum* isolates from botulism cases in Japan between 2006 and 2011. Appl Environ Microbiol. 2014 Nov;80(22):6954-64. doi: 10.1128/AEM.02134-14. Epub 2014 Sep 5. PubMed PMID: 25192986; PubMed Central PMCID: PMC4249013.
- Yamashita A, Sekizuka T, Kuroda M. GcoGSA-BA: A Global Core Genome SNP Analysis for *Bacillus anthracis*. Biosecur Bioterror. Volume 13, Number 1, 2015. DOI: 10.1089/hs.2014.0076 in press.

2) 学会発表

- 山下明史、関塚剛、黒田誠. GcoGSA-BA NGS データから炭疽菌のコアゲノム系統解析を行うウェブアプリケーション。第88回日本細菌学会総会 岐阜市 平成 27 年 3 月 26-28 日 (ポスター発表)
- Kenri T, Sekizuka T, Yamamoto A, Iwaki M, Takahashi M, Kuroda M, Shibayama K. Whole genome single-nucleotide polymorphism (SNP) Analysis of *Clostridium botulinum* isolates from botulism cases in Japan. 51th Interagency Botulism Research Coordinating Committee (IBRCC 2014), Philadelphia PA, USA. October 26-29, 2014. (Poster presentation).

H . 知的財産権の出願・登録状況 (予定を含む)

該当なし

炭疽菌 *Bacillus anthracis* のゲノム分子疫学 (関係者のみ運用予定)

GcoGSA BA

Global core Genome SNP Analysis for *Bacillus anthracis*

Project name:

Input your sequence reads (Only .fastq.gz up to 2GBytes in total are acceptable.)

Strain Name	Read file 1	Read file 2
1	選択... ファイルが選択されていません。	選択... ファイルが選択されていません。
2	選択... ファイルが選択されていません。	選択... ファイルが選択されていません。
3	選択... ファイルが選択されていません。	選択... ファイルが選択されていません。
4	選択... ファイルが選択されていません。	選択... ファイルが選択されていません。
5	選択... ファイルが選択されていません。	選択... ファイルが選択されていません。

Send an e-mail after finished analyzing.

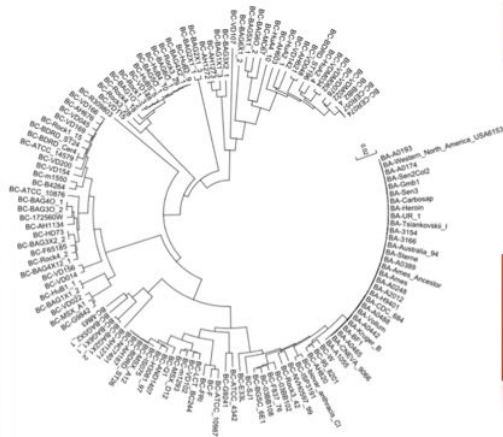
Mail address:

Send

Change number of samples

Reference genome: *Bacillus anthracis* str. 'Ames Ancestor' chromosome, complete genome. (gil50196905|refl|NC_007530.2)

Phylogenetic tree of the original data set.



[Traditional tree shape](#)
[Newick formatted data](#)
[MEGA formatted data](#)
[SNP allere table \(compressed 13MBytes\)](#)

LF
EF
PA

MiSeq



便 血液、髄液 咽頭ぬぐい液



~~GSA junior~~



SNP判定のエラーが多い

Mapping status

Strain	Raw reads	After Trimming	Mapping
S6 sequence	7631281	4926893 (64.6%)	4896388 (64.2%)
S5 sequence	6850274	4814255 (70.3%)	4775247 (69.7%)

Mapped region

Sequence or region	Whole length	Mapped bp (%) (depth >=5)	
		S5 sequence	S6 sequence
gil50196905 refl NC_007530.2	5227419	4914301 (94.0%)	4827010 (92.3%)
gil20520075 gbl AE011190.1	181677	163651 (90.1%)	161559 (88.9%)
gil50118566 gbl AE017335.3	94830	84018 (88.6%)	81139 (85.6%)
lethal factor (AE011190_149357-151786)	2429	1865 (76.8%)	1767 (72.7%)
edema factor (AE011190_122608-125010)	2402	1946 (81.0%)	2187 (91.0%)
protective antigen (AE011190_143779-146073)	2294	2225 (97.0%)	2161 (94.2%)

Number of SNPs used: 636038 / 657183 (96.8%)

[Download SNP data in tabular format. \(8.6 MBytes\)](#)
[Download SNP data in fasta format. \(83.3 MBytes\)](#)
[Download phylogenetic tree in Newick format. \(5.5 KBytes\)](#)
[Download phylogenetic tree in PDF format. \(9.3 KBytes\)](#)

図1 MePIC - MEGAN による病原体検索の結果、炭疽菌が候補として浮上した場合、先に使用した解読リードを用いたゲノムワイドSNPs解析システム Global core genome SNP analysis for *Bacillus anthracis* (GcoGSA-BA)を開発した。現在、関係者のみ運用可能としている。公開されている炭疽菌ゲノムと比較したゲノムワイド系統樹の生データまで作成し、適当な系統樹ビューワーで閲覧可能なデータをダウンロード可能である。炭疽菌は *Bacillus cereus* group に属し、セレウス菌との鑑別間違いを起さぬよう、Lethal factor (LF), Edema factor (EF), Protective antigen (PA)の炭疽菌の病原因子の特定も可能にした。

GcoGSA *Bacillus anthracis* based on phylocoregenomics

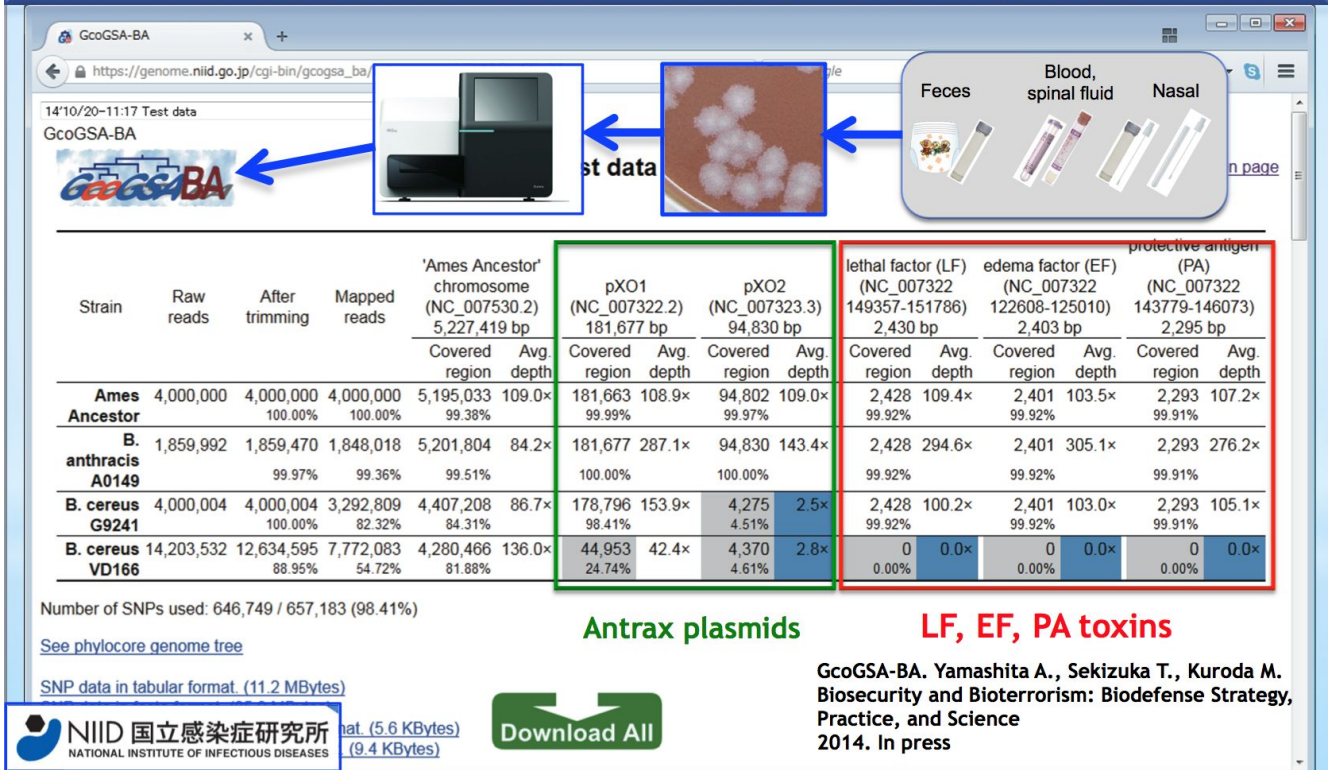


図2 GcoGSA を用いたコアゲノム SNPs 分子系統解析と炭疽菌プラスミド・毒素因子(LF, EF, PA 毒素)の特定

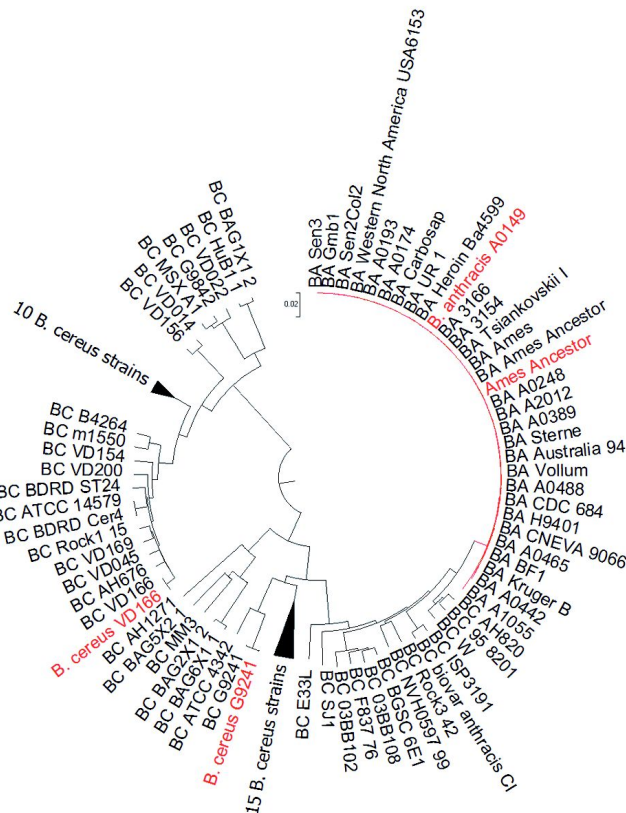
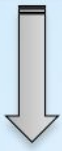


図3 コアゲノム SNPs 分子系統解析の1例。炭疽菌と類縁のセレウス菌との鑑別を容易にするだけでなく、炭疽菌株グループの中の詳細な分子疫学を可能とした。

illumina .fastq 生解読リード



- De novo assembly
- RNA finding
- 16S-rRNA phylogeny
- Species identification

GcoGSA :
Global core Genome SNPs Analysis

- GcoGSA-BA (炭疽菌、セレウス菌)
- YP (ペスト菌)
 - FT (野兎病菌 Type A, B)
 - CR (コクシエラ、リケッチア)
 - BPM (類鼻疽菌)
 - CB (ボツリヌス菌)

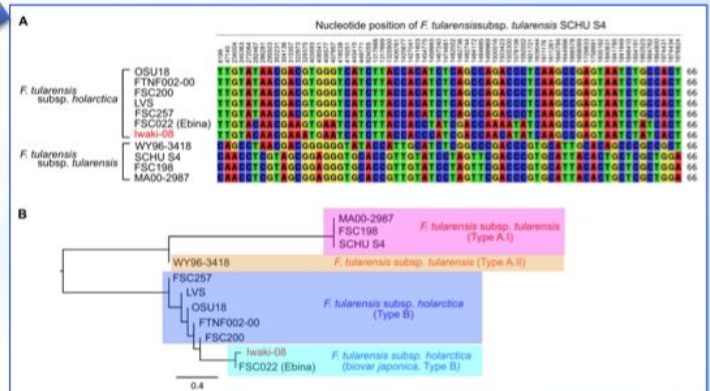
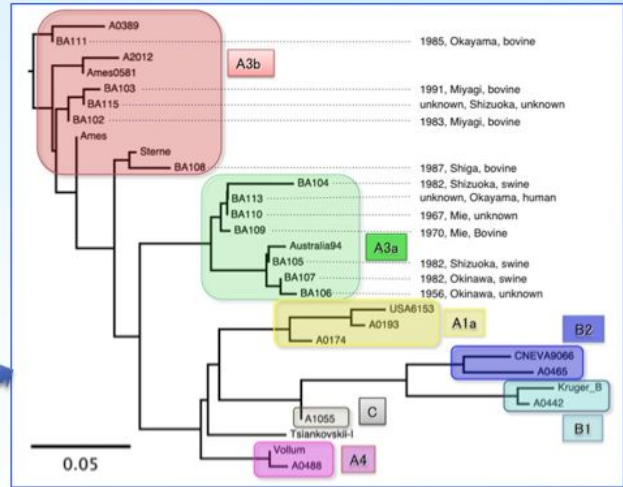


図4 今後、ペスト、野兎病、コクシエラ、類鼻疽、ボツリヌスと順々に構築予定である。

Genetic characterization and comparison of *Clostridium botulinum* strains isolated from botulism cases in Japan between 2006 and 2011.

Kenri T¹, Sekizuka T², Yamamoto A³, Iwaki M³, Komiya T³, Hatakeyama T⁴, Nakajima H⁵, Takahashi M³, Kuroda M², Shibayama K³.

Figure 5

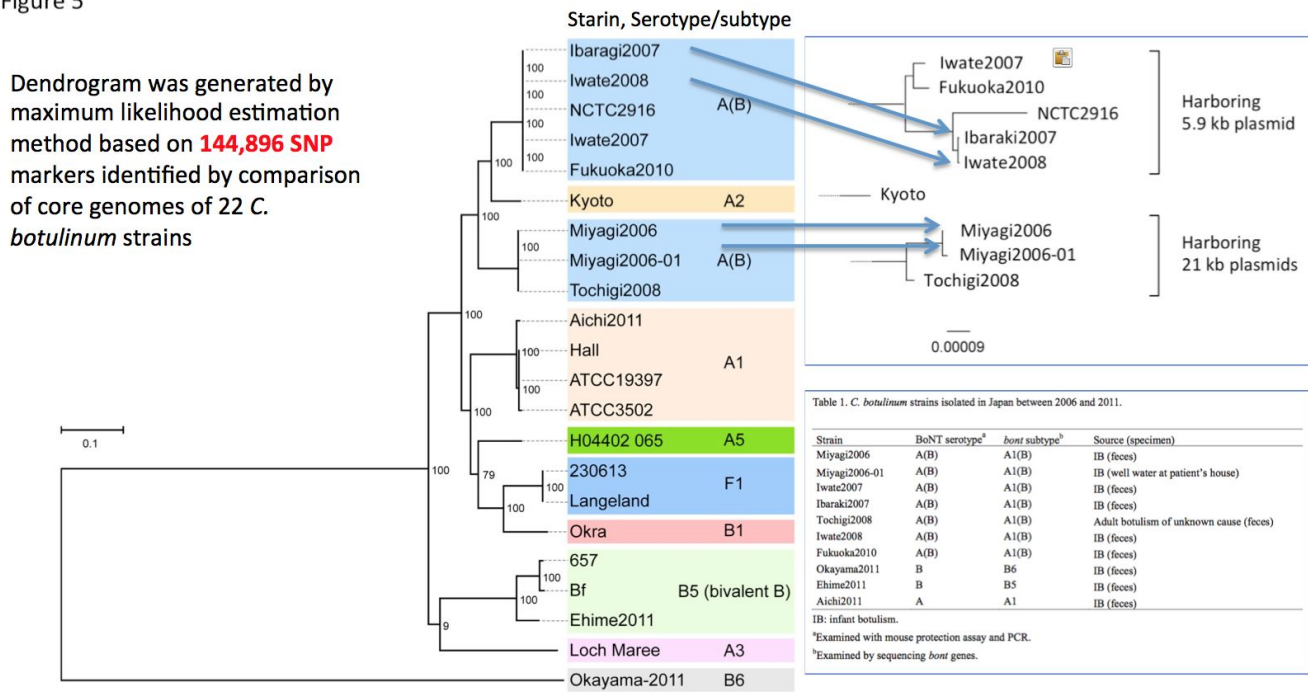


図5 国内分離ボツリヌス菌のゲノム分子疫学解析（感染研・細菌第二部 見理先生との共同研究）。国内分離株と公開ゲノム配列のコアゲノム比較解析から、144,896箇所のSNPsサイトを抽出し、ゲノム分子系統樹を作成した。従来の分子疫学解析を精度よく評価することに成功し、分離県と年代が異なる分離株においても由来特定と同一性を議論する基盤情報を収集解析することができた。