

大規模比較ゲノム解析による病原細菌の進化と病態発症機構の解明

大野 誠之、山口 雅也
大阪大学 大学院歯学研究科

1. はじめに

化膿レンサ球菌 (*Streptococcus pyogenes*) は健康な小児および成人の上気道や皮膚などに存在する細菌である。咽頭炎や膿痂疹 (とびひ) などの非侵襲性感染症の原因菌である、いわゆる溶血性レンサ球菌 (溶連菌) である。一方で、壊死性筋膜炎やレンサ球菌性毒素性ショック症候群などの致死的な侵襲性感染症を引き起こし、「人喰いバクテリア」とも呼ばれる。侵襲性感染症の致死率は 23~81% に及ぶことが報告されており、近年日本を含む先進国では侵襲性感染症患者が増加している [1]。しかしながら、ワクチンなど効果的な予防手段が存在しないのが現状である。

化膿レンサ球菌の分類には、病原因子である Mタンパクをコードする遺伝子の配列に基づく *emm* 分類がなされている。近年、特に *emm89* 型化膿レンサ球菌が侵襲性感染症患者から分離される頻度が高まっている。世界的に侵襲性感染症由来の *emm89* 型株の大部分には、莢膜ヒアルロン酸合成遺伝子群 *hasABC* の欠失と、病原因子 NADase およびストレプトリジン O の発現増加をもたらす *nga-ifs-slo* オペロンのプロモーター領域に生じた変異が検出されている。しかし、少なくとも日本において近年分離された非侵襲性感染症由来株でも同様の変異を有する株の割合が高く、侵襲性感染症特異的な変異ではないことが示唆されている [2]。そこで、侵襲性感染症の発症に関与する因子およびワクチン抗原の同定が急務となっている。本研究では、侵襲性感染症の発症機構の解明を目的として、*emm89* 型化膿レンサ球菌の全ゲノム情報から病態に相関する細菌因子の遺伝統計学的探索を行った。

2. 計算方法

2.1 化膿レンサ球菌の全ゲノム情報の収集

国立感染症研究所および全国の公衆衛生研究所等から、日本で 2016 年から 2021 年にかけて分離された *emm89* 型化膿レンサ球菌株など 207 株の分与を受け、次世代シーケンサーにて全ゲノム情報の解読を行った。さらに、すでに NCBI データベース上に公開されている全世界の *emm89* 型化膿レンサ球菌のゲノムデータを、プログラム *fasterq-dump* v.2.9.6 により SQUID 上にダウンロードした。

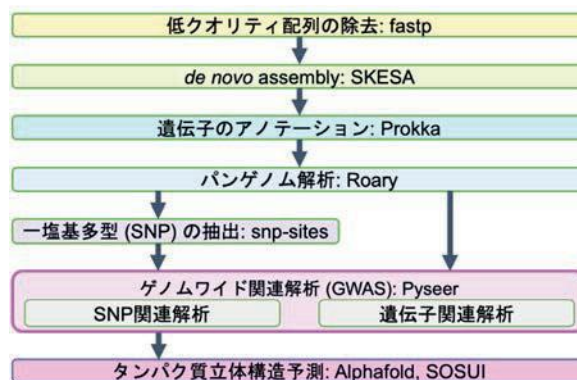


図 1: 全ゲノム解析のパイプライン

2.2 パンゲノム解析

全ゲノム解析のために我々が構築したパイプラインを示す (図 1)。収集したすべてのシーケンスデータについて、プログラム *fastp* v.0.20.1 を用いて低クオリティ配列の除去を行った。本研究で収集した 2016 年から 2021 年にかけて日本で分離された 207 株について、プログラム *emm-typing-tool* v.0.0.1 を用いて *emm* 型を決定した。*emm89* 型株のシーケンスデータについて、プログラム *SKESA* v.2.4.0 による *de novo assembly* を実施した。遺伝子のアノテーションは *Prokka* v.1.14.5 を用いて行った。*Prokka* による解析で得られた GFF

ファイルに対して、Roary v.3.12.0 を用いてパンゲノム解析を実施した [3]。パンゲノムとは、細菌集団の中に存在する全ての遺伝子を指し、ほぼ全ての株に保存されているコア遺伝子と、一部の集団もしくは菌株特異的に存在するアクセサリ遺伝子に大別される。コア遺伝子群には生存に必要な遺伝子群が含まれている可能性が高く、また全株間で配列比較が可能であるため一塩基多型 (SNP) を抽出できる。対してアクセサリ遺伝子には外来遺伝子などが含まれており、遺伝子の存在自体が病態に関連する可能性が考えられる。Roary による解析のオプションとして、プログラム MAFFT を用いたコア遺伝子アライメントの出力 (-e -mafft)、99 %以上の株が保有する遺伝子をコア遺伝子と定義 (-cd 99)、および 76 スレッドの使用 (-p 76) をそれぞれ指定して計算を実施した。Roary を用いたパンゲノム解析によりコア遺伝子群を連結するとともに、全株について多重配列整列させたコア遺伝子アライメントと、全遺伝子の各株における分布を示した行列を得た。さらにコア遺伝子アライメントから、プログラム snp-sites v.2.5.1 および bcftools v.1.9 にて挿入・欠失を含む一塩基多型 (single nucleotide polymorphism; SNP) の抽出を行った。

2.3 ゲノムワイド関連解析 (GWAS)

侵襲性感染症であるか否かのフェノタイプとジェノタイプとの関連を、ゲノムワイド関連解析 (GWAS) により遺伝統計学的に検討した。プログラム Pyseer v.1.3.4 を用い、SNP もしくは遺伝子の有無をバリエーションとした GWAS を実施した [4]。計算には SNP の情報または遺伝子の分布を示した行列に加えて、各株のフェノタイプの情報、および遺伝学的距離行列の固有値を指定した。固有値は系統によるバイアスの除去のために共変数として追加した。結果はそれぞれ R v.4.0.3 およびパッケージ ggplot2 を用いて描画した。

2.4 タンパク質立体構造予測

GWAS にて有意にフェノタイプと相関する変

異のうち、構造遺伝子領域に存在しアミノ酸の非同義置換を生じる変異を、EMBOSS transeq v.6.6.0.0 により検索した。変異がタンパク質の機能に影響するかを検討するため、プログラム AlphaFold v.2.2.2 を用いてタンパク質立体構造予測モデルを作製した [5]。複合体の形成が既に報告されている分子、あるいは UniProt による検索にてそのタンパク質自身の多量体もしくは他のタンパク質との複合体を形成することが予想される分子については複合体の構造予測を行った。各モデルとも 5 回計算を実行し、最終的にトポロジーの正確度の指標である ipTM+pTM の最大値をとるモデルを採用した。さらに、ウェブツール SOSUI を用いて膜タンパク質であるかの判定および膜貫通領域の予測を実施した。立体構造の描画はソフトウェア PyMOL v.2.5 (Schrödinger, LLC., NY, USA) にて行った。

3. 結果

3.1 化膿レンサ球菌の全ゲノム情報の解読

収集した 207 株の臨床分離株のうち、emm 遺伝子型を決定した結果 150 株が emm89 型と判定された。公開されているデータと合わせて、emm89 型株では侵襲性感染症由来株が 420 株、非侵襲性感染症由来株が 246 株の合計 666 株の全ゲノム情報が収集された。

3.2 パンゲノム解析

パンゲノム解析の結果、666 株の emm89 型株より合計 4,743 遺伝子が検出され、うち 1,327 遺伝子が 99%以上の株が保有するコア遺伝子であった。

3.3 GWAS

GWAS の結果、コア遺伝子の 90 箇所が存在する SNP が有意にフェノタイプと相関した (図 2)。RNA メチルトランスフェラーゼをコードすることが推定される遺伝子 group_589 にはもっとも多い 15 箇所の SNP が集積していた。また、covS 遺伝子には 8 箇所の SNP がそれぞれ集積していた。

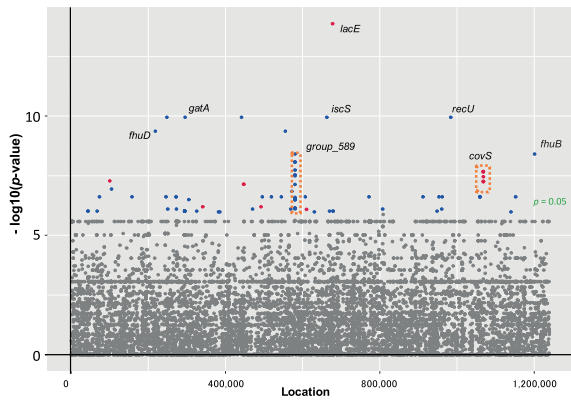


図 2 : SNP に関する GWAS の結果

SNP の集積する遺伝子領域の多くはその塩基配列の多様性が高く、欠失や挿入が存在していた。もっとも p 値が低い SNP は *lacE* 遺伝子に存在した。*lacE* 遺伝子はラクトースやガラクトース代謝に関与するホスホリボシルトランスフェラーゼシステム (PTS) を構成する、EIICB コンポーネントをコードする。この SNP により LacE タンパクの 554 番目のアミノ酸がグリシンからバリンに置換される。効果量は 3.47 と正の値を示し、侵襲性により相関することが示唆された。またこの変異は主にアメリカ合衆国で分離された株に認められた (図 2)。

Fe^{3+} -フェリクローム複合体結合タンパク質をコードする *fhuD* 遺伝子およびシデロフォア輸送パーミアーズをコードする *fhuB* 遺伝子において有意にフェノタイプと相関した SNP は、それぞれ日本の株においてのみ認められる変異であった。各変異により FhuD の 180 番目のアミノ酸がバリンからイソロイシンへ、FhuB の 73 番目のバリンがアラニンに置換された。それぞれ効果量は負の値を示し、両 SNP の存在は非侵襲性に関連することが示された。

侵襲性感染症の発症に関与する遺伝子を特定するため、検出された全 4,743 遺伝子をバリエーションとした GWAS を実施した。その結果、140 遺伝子の存在が有意に病態と相関し、そのすべてがコア遺伝子に含まれないアクセサリ遺伝子であった (図 3)。140 遺伝子中、トランスポザーゼやリコンビナーゼなど可動性遺伝因子に関連する遺伝子群は 14 遺伝子、ファージタンパク質をコードする遺伝子は 25 遺伝子が検出された。

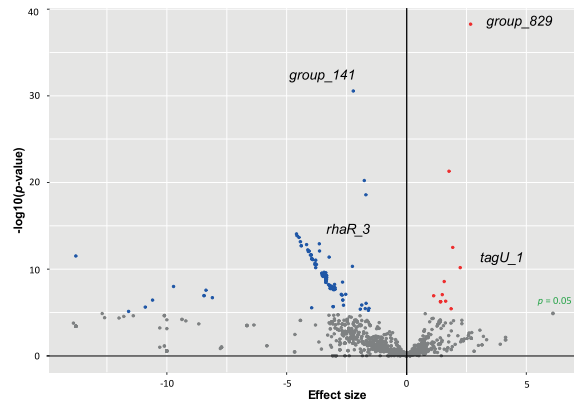


図 3 : 遺伝子に関する GWAS の結果

3.4 タンパク質立体構造解析

fhuB 遺伝子および *fhuD* 遺伝子に存在する SNP は日本においてのみ認められる変異であり、それぞれアミノ酸の置換を起こすことが示唆されたため、AlphaFold による立体構造予測により変異による機能への影響を検討した。

fhuB 遺伝子における SNP である T218C は、FhuB タンパクにおける 73 番目のバリンをアラニンに置換する。FhuADBG は、シデロフォアの種類であり、 Fe^{3+} と錯体を形成してその運搬を担うフェリクロームの取り込みに関与する ATP-binding cassette トランスポーターシステムである [6, 7]。FhuB は細胞膜上に局在し、FhuG とヘテロ二量体を形成してチャンネルを構成する。細胞外ではフェリクロームと結合するタンパク質 FhuD からフェリクロームを受け取り、細胞内では ATP 結合活性を有するタンパク質 FhuA と結合する。SOSUI と AlphaFold により、膜貫通領域と立体構造の予測を行ったところ、FhuB および FhuG はそれぞれ 9 回膜貫通型タンパク質であることが示唆された (図 4A)。一般的な ABC トランスポーターでは、FhuA のような ATP 結合タンパク質は 2 分子が会合しているため、FhuA が 2 分子、および FhuD、FhuB、FhuG がそれぞれ 1 分子の五量体として立体構造予測モデルを構築した。その結果、FhuB の 73 番目のアミノ酸残基は細胞外の FhuD と隣接する領域に存在することが示唆された (図 4B、赤矢頭)。SNP によるバリンからアラニンへの置換によって疎水性側鎖が短縮し、分子周囲の疎水性が低下することで FhuD との相互

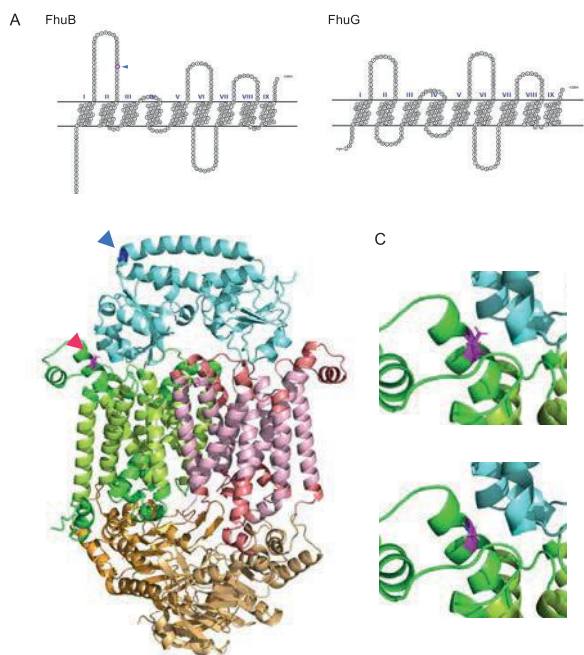


図4：AlphaFoldによるFhuAADBGB複合体の立体構造予測

作用が変化し、鉄輸送機能に影響を及ぼす可能性が示された(図4C)。

一方で、*fhuD* 遺伝子における SNP である G1661A は、FhuD タンパクにおける 554 番目のアミノ酸残基をバリンからイソロイシンに置換する。AlphaFold による構造予測から、554 番目のアミノ酸残基は活性部位や FhuB との相互作用が行われる可能性のある部位とは離れた α -ヘリックス中に存在することが示唆された(図4B、青矢頭)。バリンとイソロイシンのいずれも炭化水素側鎖を持つ疎水性アミノ酸であり、変異による FhuD の構造や機能への影響は極めて小さいものと考えられた。

4. おわりに

本研究では、侵襲性および非侵襲性感染症由来の化膿レンサ球菌の全ゲノム情報を 666 株分解読・収集し、侵襲性感染症の発症に関与する細菌因子の探索を行った。その結果、コア遺伝子群に存在する 90 箇所の SNP および 140 遺伝子の存在が病態と有意に関連した。なかでも、日本で分離された株においてのみ存在する *fhuB* 遺伝子における SNP はアミノ酸の置換を生じ、立体構造予測の結果から鉄輸送機能に影響することが示唆

された。

さらに我々は GWAS で得られた病態と関連する変異を導入した、遺伝子変異株の作製を行っており、分子生物学的実験を通して病原性の証明および治療標的としての可能性の検討を行っている。*fhuB* 遺伝子における SNP を導入した遺伝子変異株は、ヒト血液中における増殖が減弱し、変異と病原性との関連が証明された。今後は他の因子についても同様の解析を実施する予定である。

参考文献

- (1) Walker, M.J. *et al.*, Disease manifestations and pathogenic mechanisms of Group A *Streptococcus*. *Clin Microbiol Rev*, 2014. **27**(2): p. 264-301.
- (2) Hirose, Y. *et al.*, Genetic Characterization of *Streptococcus pyogenes emm89* Strains Isolated in Japan. *Infectious Microbes & Diseases*, 2020. **2**(4): p. 160-166.
- (3) Page, A.J. *et al.*, Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, 2015. **31**(22): p. 3691-3.
- (4) Lees, J.A. *et al.*, pyseer: A comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics*, 2018. **34**(24): p. 4310-4312.
- (5) Jumper, J. *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* 2021 **596**:7873, 2021. **596**(7873): p. 583-589.
- (6) Hanks, T.S. *et al.*, Differential regulation of iron- and manganese-specific MtsABC and heme-specific HtsABC transporters by the metalloregulator MtsR of group A *Streptococcus*. *Infect Immun*, 2006. **74**(9): p. 5132-9.
- (7) Hanks, T.S. *et al.*, ABC transporter FtsABCD of *Streptococcus pyogenes* mediates uptake of ferric ferrichrome. *BMC Microbiology*, 2005. **5**: p. 62-62.