

病原遺伝子データベースの構築およびゲノム全遺伝子の相互比較による相関解析

●黒川 顕

奈良先端科学技術大学院大学

〈研究の目的と進め方〉

1995年にインフルエンザ菌のゲノム全配列が決定されて以来、ゲノム解析手法の発達により、現在では約300種類の細菌のゲノム全配列が得られるようになった。これらの中には、病原性、非病原性を含めた複数の同属同種異株のゲノム全配列も存在している。新規の病原因子の特定は、一種類のゲノム配列解析からは困難で、他のゲノム配列とゲノムレベルで比較することで初めて明らかとなる。しかし、病原因子は遺伝子の水平伝播等で獲得される場合が多く、株、種や属レベルだけの比較ゲノムでは、病原因子の特定、病原因子獲得のメカニズムを解析するには不十分である。そこで、病原性という細菌の表現型に焦点をあて、種属などの種分化レベルに依存しない全遺伝子比較を行い、表現型に寄与する遺伝子群の網羅的な抽出を試みる。さらに、デジタル情報としての遺伝子配列情報だけではなく、これまで細菌学で培われてきた表現型という情報を組み合わせることで、今まで得られなかった病原性と遺伝子群との関連性を抽出することが可能となる。これにより、ある病原性を表現するための特徴的な遺伝子群の特定が可能になると考えられる。また、それら遺伝子群に対して分子進化解析を行うことにより、どのような進化経路を辿ってそれら遺伝子群が病原細菌に存在するようになったかなど、病原性という観点から細菌ゲノムの進化を論ずることが可能になると考えられる。さらに、現在約300種類もの病原性、非病原性細菌のゲノム全配列が公開されており、病原性だけでなく、非病原性の環境菌に対しても同様の手法により比較、相関解析することで、病原性と遺伝子群との関連性がより詳細に得られるようになるだけでなく、例えば生育環境に依存する表現型と遺伝子群との関連性を明らかにするものと期待できる。

また、病原性という曖昧な表現型にこだわるのではなく、病原遺伝子に着目し、その遺伝子群をターゲットとして比較ゲノム解析をすることで、病原遺伝子の分布や構造を明らかにしようとしている。初めとして、病原性大腸菌や腸炎ビブリオ、さらには植物病原菌において代表的な病原因子として注目されている3型分泌装置(TTSS)に着目した。TTSSでは装置を構成する遺伝子群のみならず、宿主細胞に送り込まれる分泌蛋白こそが病原性として重要であるが、これら分泌蛋白は遺伝子配列レベルでの相同性がほとんどなく、それら遺伝子を予測することは非常に困難であり、世界中で精力的に予測が試みられている。代表者は既知の分泌蛋白遺伝子のN末端50残基のアミノ酸頻度パターンに着目し、ゲノム全遺伝子から多次元尺度構成法(MDS)および自己組織化マップ(SOM)によりクラスタリングすることで、分泌蛋白遺伝子の予測をおこなった。病原性大腸菌O157に本方法を適用し、60個の遺伝子を分泌蛋白遺伝子候補として予測した。これら予測した遺伝子には、既知の分泌蛋白遺伝子だけでなく、共同研究者により実験的に明らかとなった分泌蛋白遺伝子がすべて含まれていることから、本予測法が極めて高いレベルにあることを示唆しており、

それら以外の候補も分泌蛋白遺伝子としての可能性が高いと考えられる。このような高い精度をもった予測法は世界的にも例がない。今後は、本方法で予測した遺伝子の抗体を作成し、共同研究者により実際に分泌されているか否かの検証をすると同時に、実験で確認された分泌蛋白遺伝子情報を本方法に取り入れることにより予測精度を向上させ、他の菌種にも応用していく予定である。

〈研究開始時の研究計画〉

現在公開されているゲノム配列が得られる細菌の、主に病原性に関する表現型データベースを作成し、病原性という表現型による分類を試みる。ここで得られた分類をもとに、1年目で開発した方法論を摘要し、系統関係からみるのではなく、病原性という表現型から見た場合の遺伝子分類を試みる。

- ・主に病原性に焦点をあてた表現型データベースを作成する。
- ・上記で得られたデータベースを多次元尺度構成法などの方法にてクラスタリングし、表現型による細菌の分類を行う。
- ・クラスタリングにより得られた細菌の分類群ごとに、全遺伝子の主に相同性によるデータベースを作成し、分類群ごとに共通な遺伝子群を抽出する。このことで、病原性を示す表現型とそれらの特徴付ける遺伝子群との関連性の解析を行う。

〈研究期間の成果〉

1) A群レンサ球菌のゲノム解析に際し、すでに公開されている同属同種異株のゲノム全配列と比較することで、劇症型の株に特徴的なゲノム構造を見いだすとともに、その構造変化がファージの挿入により引き起こされた事を示唆した。そこで代表者は日本のSTSS患者から単離されたM3株(SS1-1)のゲノム全配列解析および、すでに公開されているM1株、M18株と、さらには同じM3株との比較ゲノム解析を行うことで、STSS発症のメカニズムの解明を試みた。SS1-1株のゲノム配列は、全長が約1.9Mbpで、ゲノム全長の約85%にあたる約1.6MbpがSF370およびMGAS8232の両株と保存しており、生命活動に必須の遺伝子やこれまでに知られている病原因子は、付着に関与する遺伝子を除いて4株にほぼ共通であった(遺伝子数は異なる)。また、SF370株およびMGAS8232株には5箇所、SS1-1株には6箇所のプロファージ、ファージ様エレメントがゲノム上に存在しているが、SS1-1株の6箇所のファージ、ファージ様エレメントのうち、2つはM1株、M18株にはなくMGAS315を含むM3株に特異的であった。SS1-1株における6箇所のファージ領域は相互に非常に良く似た構造をもっているものの、上述のO157ゲノム上のプロファージと同様、それらの配列は全域にわたって相同性があるわけではなく、構造単位での相同性が認められるのみである。株間においてゲノム上の相同遺伝子の位置の保存性を調べたところ、非常に興味深い結果が

得られた。SSI-1株では、大規模なゲノムレベルでのリアレンジメントが生じており、1Mbにもおよぶ領域が複製軸を境に、他の3株と比較して逆位を起こしていた。この逆位がSSI-1株に特有のものか、またはどのような株に分布するのかを調べるために、逆位を生じている箇所を挟むようにlong PCRをおこなうべくプライマーを設計し、これまでに単離された122株の*S.pyogenes* M1, M3, M28株に対して逆位の有無の検討をおこなった。その結果、SSI-1株に見られる大規模な逆位は1985年以前の単離株では25%しか確認できなかったが、1990年以降のSTSS患者からの比較的新しい単離株の81%に確認することができた。しかしながら、1990年以降のnon-STSS患者からの単離株では、その57%においてしか逆位が観察されなかった。このことから、SSI-1株で観察されたゲノムレベルの逆位が、劇症型の感染症の出現と何らかの関係があると示唆することができる。この逆位をおこしている領域の複製開始点(ori)に近い部分は、*rrn-comX*領域と呼んでいるribosomal RNAなどがコードされている領域で、oriを境に等距離の位置に存在している。複製終了点(ter)に近い部分には、SSI-1株においては、late領域の配列が似ているSPsP5およびSPsP6の2つのファージがterから等距離の位置に挿入されていた。このことから、SSI-1株に見られた逆位はoriに近い部分は*rrn-comX*領域の相同性で、またterに近い部分はファージのlate領域の配列の相同性により組換えを起こしたことが示唆できる。このlate領域にはスーパー抗原、hyaluronidase、streptodornaseなどの病原性に関与する遺伝子群やholin遺伝子が存在しており、他のファージのlate領域と高い相同性が認められる。2つのM3株を比較したところ、M3株にしか存在しない2つのファージにおけるこの領域で、ファージ配列の組換えが起きていることが明らかとなった。このことは、ファージ領域に存在する病原遺伝子群が、ゲノムの組換えにより交換可能であることを示唆している。SSI-1株に見られるゲノムレベルの大規模な逆位は、terを挟んだ両側にlate領域が相同な2つのファージが取り込まれることで、相同組換えにより生じたと考えられる。さらに、ゲノムレベルの逆位が生じることで、ファージの一部の構造がモザイク状にシャッフルされ、新たなファージがゲノム内で生じるという、病原性大腸菌O157と同様なファージの進化を垣間みることができる。この観察事実はGASゲノムの多様性がファージの獲得のみならず、ファージを獲得することにより生じたゲノムレベルの組換えによってもたらされ、その結果、GASによる感染の多様性に寄与していると考えられる。

- 2) 得られるすべての細菌ゲノム配列について、パラログ遺伝子群の特定および、相同遺伝子と特異的遺伝子の特定をおこなった。
- 3) 文献情報から抽出している病原性のうち、血便に焦点をあてて、相同遺伝子群からこれらに関与することが予想される遺伝子群を抽出した。
- 4) 微生物ゲノムデータベースを対象に、アノテーションが曖昧な遺伝子に関して、現在利用できるバイオインフォマティクスツールを駆使し、自動的にアノテーション可能なHAPPYシステムを構築し、公開した。
- 5) 病原性大腸菌や腸炎ビブリオ、さらには植物病原菌において代表的な病原因子として注目されている3型分泌装置(TTSS)において、装置を構成する遺伝子群のみならず、病原性として重要である宿主細胞に送り込まれる分泌蛋白の予測を試みた。これら分泌蛋白は遺

伝子配列レベルでの相同性がほとんどなく、それら遺伝子を予測することは非常に困難であり、世界中で精力的に予測が試みられている。代表者は既知の分泌蛋白遺伝子のN末端50残基のアミノ酸頻度パターンに着目し、ゲノム全遺伝子から多次元尺度構成法(MDS)および自己組織化マップ(SOM)によりクラスタリングすることで、分泌蛋白遺伝子の予測をおこなった。病原性大腸菌O157に本方法を適用し、60個の遺伝子を分泌蛋白遺伝子候補として予測した。これら予測した遺伝子には、既知の分泌蛋白遺伝子だけでなく、共同研究者により実験的に明らかとなった分泌蛋白遺伝子がすべて含まれていることから、高い精度で予測できているものと考えられる。そこで、本予測法を用いて予測した遺伝子が、本当に菌体外に分泌されているかどうかを、現在共同研究者らが実験による確認をおこなっている。

<国内外での成果の位置づけ>

ファージのダイナミクスによるゲノムの構造変化と病原性との関連づけた報告は世界的にも稀である。バクテリアゲノムの進化とファージなどによる遺伝子の水平伝播が大きく関連していることが明らかとなったことは、バクテリアゲノム研究のひとつの大きな成果と言える。また血便という病原性から抽出した3型分泌装置に関与する遺伝子群は、世界的に注目が集まっており、特に分泌されるタンパク質をコードする遺伝子の特徴抽出は、実験およびバイオインフォマティクスの両分野において克服すべき課題となっている。

<達成できなかったこと、予想外の困難、その理由>

病原性データベースに関しては、特徴的な病原性に関するデータベースは作成完了し、平成15年度の合同班会議にて発表した。しかしながら、文献から抽出した全情報をデータベース化しようと試みたものの、非常に多くの表現型から構成される点、および病原性という曖昧な定義が原因で、効率の良いデータベース作成が困難であった。

<今後の課題>

病原性という曖昧な表現型にこだわるのではなく、病原遺伝子に着目し、その遺伝子群をターゲットとして比較ゲノム解析をすることで、病原遺伝子の分布や構造を明らかにしようとしている。その第一段階として、2成分制御系の遺伝子群に着目し、枯草菌をはじめとするグラム陽性菌のゲノム全配列を比較解析し、これら遺伝子群が、種間において遺伝子の相同性が低いものの、関連遺伝子群のゲノム上での並び順においては種を超えて非常に良く保存されている事が明らかとなった。現在、これら遺伝子群の構造をさらに詳細に解析中である。さらに、3型分泌装置の遺伝子群においても同様の解析を実施しており、現在入手できる細菌のゲノム配列のアノテーション情報から抽出した遺伝子群のデータベースを構築し公開している。

- 1) 病原遺伝子データベースの第一段階として、特に3型分泌装置および2成分制御系の遺伝子群を遺伝子データベースより抽出し、病原遺伝子データベースとして蓄積する。
- 2) それら遺伝子群を相同性検索により細菌のゲノムより網羅的に抽出し、さらに病原遺伝子データベースとして蓄積する。
- 3) オルソログ遺伝子データベースに、それら遺伝子群を

照会することで、細菌種ごとの病原遺伝子群の構造や、保存性、特異性を明らかにする。

- 4) 上記 3 で明らかになった遺伝子群の保存性と特異性をもとに、主に文献情報から抽出した病原性に関する情報との関連性を検討する。

〈研究期間の全成果公表リスト〉

1) 論文／プロシーディング

1. 0404121150
Nakagawa, I., Kurokawa, K., Nakata, M., Yamashita, A., Tomiyasu, Y., Okahashi, N., Kawabata, S., Yamazaki, K., Shiba, T., Yasunaga, T., Hattori, M., Hayashi, H., and Hamada, S., Genome sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic arrangement in invasive strains and phage evolution. *Genome Res.*, 13, 1042-1055 (2003).
2. 0404121155
Murakami, T., Najima, M., Ogawa, M., Kurokawa, K., and Yasunaga, T., HAPPY: Hypothetical and putative protein database system. *Genome Informatics*, 14, 653-654 (2003).
3. 0602171808
Kurokawa, K., Bioinformatics analysis of bacterial genomes: -One example- repeat patterns in bacterial genomes. *Infections* 2004, 112-123, (2004).
4. in press
Iida, T. and Kurokawa, K., Comparative genomics: genome configuration and the driving forces in the evolution of vibrios. *The Biology of Vibrios*, in press, (2006).