

機関番号：17601

研究種目：基盤研究(B)

研究期間：2008～2010

課題番号：20310116

研究課題名(和文) 新世代 DNA 塩基配列決定技術を用いた大腸菌 pan-genome の解明

研究課題名(英文) Escherichia coli pan-genome analysis using next-generation DNA sequencing technologies

研究代表者

林 哲也 (HAYASHI TETSUYA)

宮崎大学・フロンティア科学実験総合センター・教授

研究者番号：10173014

研究成果の概要(和文): 様々な病原型あるいは由来をもった大腸菌株を多数収集し、新型シーケンサーを用いた大規模全ゲノム配列決定とゲノム比較解析を行った。その結果、大腸菌の pan-genome 情報を取得するとともに、腸管出血性大腸菌 (EHEC) などの病原型に特徴的な遺伝子レパートリーとその獲得メカニズムを解明した。また、EHEC や腸管病原性大腸菌 (EPEC) として分離されている菌株の中に *E. arbertii* が多数含まれていることを見出し、代表菌株の概要配列を取得した。

研究成果の概要(英文): We collected numerous Escherichia coli strains of various pathotypes and origins and performed a large scale genome sequencing using next-generation sequencers. Genomic comparison of these sequenced strains and already sequenced strains provided the pan-genome information of E. coli and revealed several genomic features characteristic to enterohemorrhagic E. coli (EHEC) or other pathotypes. Furthermore, we found a significant number of E. arbertii strains are included in the strains isolated as EHEC or enteropathogenic E. coli (EPEC) and obtained draft sequences of representative E. arbertii strains.

交付決定額

(金額単位:円)

	直接経費	間接経費	合計
2008年度	4,500,000	1,350,000	5,850,000
2009年度	5,500,000	1,650,000	7,150,000
2010年度	5,100,000	1,530,000	6,630,000
年度			
年度			
総計	15,100,000	4,530,000	19,630,000

研究分野：複合新領域

科研費の分科・細目：ゲノム科学・基礎ゲノム科学

キーワード：ゲノム、微生物、細菌、進化、感染症

1. 研究開始当初の背景

細菌ゲノムの大きな特徴の一つは、大量の外來性遺伝子の存在である。そのため、1つの菌種内でも異なる菌株間で驚くほど高いレベルのゲノム多様性が存在することがあり、様々な菌株が保有するゲノム情報の総和をその菌種の pan-genome と呼ぶ。pan-genome の実体を明らかにするために

は相当数の菌株を対象としたゲノム解析が必要であるが、従来の配列決定技術ではコスト的にほぼ不可能であった。しかし、本研究開始直前に新世代のシーケンサーが登場したことによって、低コストでの re-sequencing が可能となり、微生物ゲノム研究分野では、重要菌種における pan-genome の解明が現実的に挑戦しうる

研究課題となってきた。

種々の細菌種の中で、大腸菌はpan-genome解析の研究対象として最も魅力的かつ重要な菌種の1つである。本菌は哺乳類・鳥類・爬虫類など様々な動物の腸内常在菌であり、基本的にはヒトや動物などに対して非病原性であるが、強い病原性を示す菌株（病原性大腸菌）も存在する。ヒト病原株は腸管内で病原性を示すものから髄膜炎や尿路感染のような腸管外感染症を起こすものまで様々である。さらに腸管内病原性株と腸管外病原性株のどちらにも、数種類の異なる病原型が存在する。常在の大腸菌株も、様々な生化学的性状でバリエーションを示す。さらに重要な点は、ゲノムサイズのバリエーションが非常に激しい（4.4~5.6 Mb）ことであり、大腸菌ゲノムの菌株間多様性が明確に反映されている。このため、大腸菌は生化学・分子生物学などにおける「モデル微生物」としてだけでなく、ゲノム多様性あるいは多様化機構の研究においても非常に魅力的な研究対象である。また、大腸菌では、新世代シーケンサーを用いた re-sequencing には必須となるレファレンスゲノム情報が、本研究開始当初において最も充実していた。

申請者らは、代表的なヒト病原性大腸菌である腸管出血性大腸菌（以下EHEC）0157のゲノム解読を行い、非病原性大腸菌K-12とのゲノム比較から、0157のゲノムが K-12と共通の 4.1 Mbの配列と0157特異的な 1.5 Mbの配列からなることを明らかにしていた（Hayashi et al., DNA Res., 2001）。その後、海外のグループから数株の大腸菌のゲノム配列が報告されたが、いずれも大腸菌固有と思われる 4 Mb 程度の共通配列と様々なサイズの菌株特異配列から構成されていた。我々のグループも 0157 以外に、他の血清型の EHEC、ヒト常在株など全配列解析を進めており、同様の結果が得られていた。このような菌株特異配列には、特定の菌株間において、ある程度の重複が存在することも認められていたが、大腸菌の pan-genome 情報を解明するためには、由来や系統の異なる菌株の全ゲノム配列を相当数解析する必要があると予想された。

一方、同一菌種内でのゲノム多様性を多数の菌株について解析する場合には、従来はマイクロアレイを用いた遺伝子組成解析などが行われてきたが、既知の遺伝子の有無に関する情報が得られないという致命的な問題があった。また、従来のシーケンス法では多数の菌株のゲノムを解読することは、コスト的に不可能であった。しかし、冒頭でも述べたように、Roche 454 や Solexa などの新世代シーケンサーの登場で、この状況は一変し、基盤 (B) の予算内

でも多数の菌株の re-sequencing ができる可能性が出てきた。

新世代シーケンサーの中で、Solexa は短い配列（25~35 bp；研究開始当初）しか得られないものの、配列の正確性は高く、大量の配列データが低コストで一度に取得できることから、多菌株の re-sequencing に適していると考えられていた。したがって、レファレンス配列が充実していれば、Solexa データを複数のレファレンス配列にマッピングすることで、解析対象菌株の全ゲノム配列のかなりの部分が取得できると考えられた。また、その菌株にのみ存在するようなゲノム領域については、サイズが小さい場合には PCR に対応でき、大きな菌株特異領域については、Fosmid library を作成し、その end-sequence をレファレンス配列へマッピングすること（Fosmid mapping）により、大きな菌株特異領域の同定と標的を限定した配列解析が可能である。そこで、これらの手法に Solexa を組み合わせることで、多数の菌株のほぼ全長配列を取得できると考えた。

2. 研究の目的

本研究では、新世代のシーケンサーの中で最も re-sequencing に適していると考えられた Solexa を利用して、まだゲノム解析が行われていないタイプの大腸菌株（ヒト腸管病原性株、ヒト腸管外病原性株、動物病原性株、腸管常在株など；16 株程度を想定）の re-sequencing を行う。Solexa のデータに Fosmid マッピングや PCR を利用した菌株特異領域の配列決定を組み合わせることによって、これらの菌株の完全配列に近いゲノム情報を新たに取得する。最終的に、これらのデータと既に解読されている大腸菌株のゲノム情報（研究開始時点で 12 株）と統合することにより、大腸菌の pan-genome の実体とその形成機構の解明を目指した。

なお、本研究開始後、Roche 454 の性能が向上し（400 bp の配列の取得が可能となったこと、またペアエンドシーケンスも可能になったこと）、解析コストも低下したこと、さらに、本シーケンサーが我々の研究室に導入され、自由に使用できる状況になったこと、などの状況の変化があったため、本研究では、Roche 454 を用いてゲノム配列解析を行うことに方針を変更した。

3. 研究の方法

(1) 菌株の収集と選定：これまでゲノム解析が行われていない病原型・宿主特異性・進化系統に属する大腸菌（まだゲノム解析が行われていない non-0157 タイプの EHEC などのヒト腸管病原性株、ヒト腸管外病原性株、動物病原性株、腸内常在株）を対象として菌株の収集と解析対象菌株の選定を行った。菌株の

選定はMLSTのデータ等を加味して行い、さらに選定した菌株については、必要に応じて、PFGE解析によるゲノムサイズの推定や保有するプラスミドの有無とそのサイズの検定等を行った。

(2) Roche 454データの取得：ゲノムDNAはGenomic-tip 100/G (Qiagen) を用いて調整し、一部をRoche 454による配列解析に用い、残りは必要に応じてFosmid libraryの作成などに用いた。Roche 454データのアセンブリは本器機付属のプログラムを用いて起こった。

(3) Roche 454データのレファレンス配列へのマッピング：当初の予定では、取得したSolexa データの既知大腸菌ゲノム配列(12株)に対するマッピングが解析の基本であった。しかし、使用機種をバージョンアップされたRoche 454に変更したため、*de novo*に近い形での配列決定が可能となった。したがって、12株のレファレンスゲノムへの454配列(contig配列)のマッピングは、genome rearrangementの有無の検定やgap fillingのためのガイドとして実施した。

(4) Fosmid mappingおよびPCRを用いたgap filling：解析菌株のFosmid libraryを作成した。当初の計画では、作成したライブラリーごとに、約1,300クローン(x10の重複度)の末端配列を決定し、12株のレファレンスゲノムに対してそれぞれマッピングを行う予定であった。しかし、Roche 454によるペアエンドシーケンスが可能になったため、システムティックなFosmid mappingは不要となり、必要なFosmidクローン(極めて良く似たプロファージ等に由来する長い繰り返し配列を含む領域)をPCRによってスクリーニングし、当該Fosmidクローンの配列(またはその部分配列)のみを決定することとした。短いギャップに関しては、PCRとprimer walkingを用いてgap fillingを行った。

(5) SolexaあるいはSOLiDを用いた配列解析：必要に応じて、SolexaあるいはSOLiDを用いた配列解析を行い、Roche 454で得られたゲノム配列のクオリティー(精度)の向上や近縁菌株での配列多型解析を実施した。

(6) 大腸菌pan-genome情報の取得：得られたゲノム配列の総当たり相同性解析(基本的にはBLAST解析)を行い、大腸菌に高度に保存されている遺伝子群(core-genome情報)と各菌株に特異的遺伝子群をそれぞれのゲノムで同定し、解析時点での大腸菌群全体としてのゲノム情報(pan-genome情報)を取得した。また、系統群(phylogroup A, B1, B2, D1, D2, E)や病原型等に特異的な遺伝子の検索・同定を試みた。

4. 研究成果

(1) EHECの解析：本研究開始前から解析を

行っていたEHEC 026, 0111, 0103 については、全ゲノム配列の取得に成功し、既知大腸菌ゲノム配列(解析時点で論文発表されていた21菌株; 赤痢菌を含む)との比較解析を行った。その結果、1) EHEC 026, 0111, 0103は、EHEC 0157と同様に大きなゲノムサイズ(5.5 Mbから5.9 Mb)を有すること、2) いずれも、EHEC 0157と同様に多数のプロファージやintegrative element(IE)を保有しており、これらのファージあるいはIEやプラスミドによって大量の外來性遺伝子が持ち込まれていること、3) 全遺伝子レポトリを他の大腸菌株と比較すると、EHEC 0157, 026, 0111, 0103はそれぞれ異なる進化系統に属するにもかかわらず、極めて良く似た病原遺伝子セットを保有していること、4) 同一の病原遺伝子をコードするファージ・IE・プラスミドの基本骨格は、各EHEC間で有意に異なり、それぞれが異なる由来・進化過程を経たものであると考えられること、などが明らかとなった。これらの結果は、異なる血清型をもつEHECがそれぞれ独立して類似した病原遺伝子セットを獲得することにより、同じ病原型へと進化したこと、またその原動力はファージを中心とした可動性遺伝因子を介した遺伝子の水平伝播であることを示している。したがって、今回の我々の解析によって、以前から提唱されていたEHECの平行進化の実体とメカニズムが明らかになったと言える。また、この解析を行った時点での大腸菌における全遺伝子レポトリ(pan-genome; 遺伝子グループとして解析)は、12,940 遺伝子グループであり、3株のnon-EHECのそれぞれから、新規の遺伝子(または遺伝子グループ)が100から300種類同定された。後者の結果は、大腸菌のpan-genomeがopen genomeであり、新規菌株のゲノム解析により、100以上の新規遺伝子が発見できることを意味する。

現在、その他のEHEC(0121, 0165, 全配列決定済みの026とは別系統の026)のゲノム解析も進行中である。いずれもドラフト配列は取得済みであり、0121と0165についてはfinishingも終了間近である。これらのゲノム解析から、EHECという特定の病原型内でのゲノムの多様性とEHECに共通したゲノム特性や共通遺伝子セットなどに関して、より詳細な情報が得られるものと期待している。

なお、EHEC 0157については、菌株間でのゲノム比較から、以前の解析から明らかになったファージのバリエーションに加えて、ISエレメント(特にIS629)が0157菌株間でのゲノム多様性の原動力であることを明らかにし、IS629の多様性を利用した0157の迅速菌株識別システムの開発にも成功した。さらに、IS629の多様性が生じるメカニズムを解析した結果、IS629などのIS3ファミリーや

IS1 ファミリー・IS30 ファミリーなどのゲノムからの脱落と、それに伴うゲノムの欠失を誘導する新規蛋白質 (IEE) を発見した。また、EHEC 026 についても、EHEC 0157 と同様の菌株間比較ゲノム解析を行い (3 株の SOLiD による re-sequencing を含む) IS621 のパリーションが菌株識別に有効であることを見出したため、このパリーションを利用した 026 の迅速菌株識別システムの開発を行った (論文投稿中)。026 については、ゲノム配列情報を利用した MLVA 解析系の改変により、その解像度の向上にも成功した。

このほか、EHEC 0157 あるいは他の EHEC のゲノム配列中から見出された病原性関連遺伝子の機能解析や EHEC 0157 ゲノム上に存在する多数のプロフェージ群の機能や相互作用の解析においても、様々な成果を得た。

(2) その他の病原性大腸菌の解析: これまでに、腸管病原性大腸菌 (EPEC) 1 株の全ゲノム解析に成功し、論文発表を行った (英国・Sanger 研究所との共同研究)。その他、赤痢菌 1 株と腸管侵入性大腸菌 (EIEC) 4 株 (4 系統でそれぞれ 1 株) の概要配列を取得 (EIEC の 1 株はフィニッシュ中) している。EIEC に関しては、海外においても解析が進んでおらず、大腸菌の pan-genome あるいは EIEC と赤痢菌との違いなどを考える上で、極めて重要なゲノム情報が得られると期待される。

また、国内外の LEE 陽性大腸菌 (ヒトおよび各種動物由来) を 240 株収集し、MLST 解析による系統解析と LEE のタイピングを行った。この結果、これらの菌株の中には大腸菌の近縁種である *E. arbertii* が多数 (26 株) の含まれていること、また *E. arbertii* の中には Stx_f 産生株も存在することが明らかとなり、4 株を選択して概要配列を取得した (1 株については finishing 中)。これらの *E. arbertii* 株のゲノム解析は、大腸菌 (特に LEE を有する EHEC と EPEC) の進化や pan-genome 解析において、極めて有用な情報をもたらすものと期待される (有用な比較解析の対象)。なお、この収集において得られた他の LEE 陽性大腸菌 (主に EPEC) についても、今後の EPEC 解析、特に動物由来の EPEC とヒト由来の EPEC の比較解析において貴重な研究リソースとなる。

(3) その他の大腸菌の解析: 腸内常在株としては、他のグループと共同で 2 株の全ゲノム解読に成功した (SE11 株と SE15 株)。いずれも、腸管への定着因子や鉄の獲得系にかかわる遺伝子群等を除くと、明らかな病原性関連遺伝子はコードされていなかった。この結果は、腸内常在菌株が非病原性または弱毒性であるという従来の推測を裏付けるものである。また腸内常在株と血液由来株との関連等を明らかにするため、全国各地の医療機関

等の協力を得て血液分離株を 120 株、また常在株 50 株を収集した。これらの菌株については、7 種類の house-keeping 遺伝子の配列を決定し、この情報を使った MLST 解析により系統解析を行っている。菌株収集と MLST 解析の双方で予想以上の時間を費やしたため、解析が終了していないが、この菌株セットは今後の解析において重要なリソースとなる。なお、爬虫類 (ヘビ・トカゲ) の腸管由来大腸菌の収集も試みたが、収集に関する季節的な制約等のため、収集を中止した。

5. 主な発表論文等

[雑誌論文] (計 34 件) 査読無の記載があるもの以外は全て査読有

1. M. Kusumoto, T. Ooka, Y. Nishiya, Y. Ogura, (他4名), T. Hayashi: IS-excision enhancer removes transposable elements from bacterial genomes and induces various genomic deletions. Nat. Commun. 2:152, DOI:10.1038/ncomms1152. 2011.
2. H. Yen, T. Ooka, A. Iguchi, T. Hayashi, N. Sugimoto, T. Tobe: NleC, a type III secretion protease, compromises NF- κ B activation by targeting p65/RelA. PLoS Pathog. 6 (12): e1001231, 2010.
3. H. Izumiya, (他3名), T. Hayashi, S. Iyoda and H. Watanabe: New system for multilocus variable-number tandem-repeat analysis of the enterohemorrhagic *Escherichia coli* strains belonging to three major serogroups: O157, O26, and O111. Microbiol. Immunol. 54: 569-577, 2010.
4. H. Toh, K. Oshima, A. Toyoda, Y. Ogura, T. Ooka, (他3名), K. Kurokawa, (他3名), T. Hayashi, M. Hattori: Complete genome sequence of the wild-type commensal *Escherichia coli* strain SE15, belonging to phylogenetic group B2. J. Bacteriol. 192(4): 1165-6, 2010.
5. 大岡唯祐, 林哲也: 細菌ゲノム解析の進歩. 日本臨床, 査読無, 68(8):144-149, 2010.
6. Y. Ogura, T. Ooka, (他7名), K. Kurokawa, T. Tobe, M. Hattori, T. Hayashi: Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc. Natl. Acad. Sci. USA 106(42): 17939-17944, 2009.
7. R. Nobe, (他6名), T. Hayashi, E. Oswald: Enterohaemorrhagic *Escherichia coli* serogroup O111 inhibits NF- κ B-dependent innate responses in a manner independent of a type III secreted OspG orthologue. Microbiology.

- 155: 3214-3225, 2009.
8. T. Ooka, (他3名), K. Kurokawa, Y. Ogura, (他6名), T. Hayashi: Development of a multiplex PCR-based rapid typing method for enterohemorrhagic *Escherichia coli* O157 strains. *J. Clin. Microbiol.* 47(9): 2888-2894, 2009.
 9. N. Yoshii, Y. Ogura, T. Hayashi, (他5名), M. Akiba: Pulsed-field gel electrophoresis profile changes resulting from spontaneous chromosomal deletions in enterohemorrhagic *Escherichia coli* O157:H7 during passage in cattle. *Appl. Environ. Microbiol.* 75(17): 5719-5726, 2009.
 10. T. Ooka, Y. Ogura, (他5名), T. Hayashi: Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in *Escherichia coli* O157 genomes. *Genome Res.* 19: 1809-1816, 2009.
 11. S.R. Leopold, (他5名), Y. Ogura, A. Iguchi, T. Hayashi, (他5名), P.I. Tarr: A precise reconstruction of the emergence and constrained radiations of *Escherichia coli* O157 portrayed by backbone concatenomic analysis. *Proc. Natl. Acad. Sci. USA.* 106 (21): 8713-8718, 2009.
 12. Md. Asadulghani, Y. Ogura, T. Ooka, (他4名), T. Hayashi: The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *PLoS Pathog.* 5(5) e1000408, 2009.
 13. A. Miyahara, N. Nakanishi, T. Ooka, T. Hayashi, N. Sugimoto, T. Tobe: Enterohemorrhagic *Escherichia coli* effector EspL2 induces actin microfilament aggregation through annexin 2 activation. *Cell Microbiol.* 11(2): 337-350, 2009.
 14. N. Nakanishi, (他2名), T. Hayashi, N. Sugimoto, T. Tobe: Regulation of virulence by butyrate sensing in enterohemorrhagic *Escherichia coli*. *Microbiology* 155: 521-530, 2009.
 15. A. Iguchi, N.R. Thomson, Y. Ogura, D. Saunders, T. Ooka, (他3名), K. Kurokawa, (他5名), T. Hayashi, J. Parkhill, G. Frankel: The complete genome sequence and comparative genome analysis of enteropathogenic *E. coli* O127:H6 strain E2348/69. *J. Bacteriol.* 191(1): 347-354, 2009.
 16. 小椋義俊, 大岡唯祐, 林哲也: ゲノム解析から見えてきた腸管出血性大腸菌の毒素及び病原因子の新知見, 化学療法の領域, 査読無, 25(5):767-777, 2009.
 17. K. Oshima, H. Toh, Y. Ogura, (他3名), T. Ooka, (他2名), T. Hayashi, K. Itoh, M. Hattori: Complete genome sequence and comparative analysis of the wild-type commensal *Escherichia coli* strain SE11 isolated from a healthy adult. *DNA Res.* 15(6): 375-386, 2008.
 18. Y. Ogura, (他2名), K. Kurokawa, (他2名), T. Ooka, (他4名), T. Hayashi: Systematic identification and sequence analysis of the genomic islands of the enteropathogenic *Escherichia coli* strain B171-8 by the combined use of Whole Genome PCR Scanning and fosmid mapping. *J. Bacteriol.* 190(21): 6948-6960, 2008.

[学会発表](計60件)

1. T. Hayashi (Invited Speaker): Genomic view on the diversification and evolution of enterohemorrhagic *Escherichia coli* (EHEC). The third China-Japan Science Forum, Diseases Prevention and Control. Mar. 14-16, 2010, Wuhan, China.
2. T. Hayashi (Invited Speaker): Genomic view on the parallel evolution of enterohemorrhagic *Escherichia coli* strains. The 4th Nagasaki Symposium on Tropical and Emerging Infectious Diseases, Nov. 26, 2009. Nagasaki.
3. T. Hayashi (Invited Speaker): Whole genome sequencing analysis of O26, O111, and O103 EHEC. VTEC2009 (7th International Symposium on Shiga Toxin (Verocytotoxin)-Producing *Escherichia coli* Infections), May 12, 2009, Buenos Aires.
4. T. Hayashi (Invited Speaker): Genomic view on the parallel evolution of enteropathogenic *Escherichia coli* strains. INRA-JSPS Workshop, Jan. 28-31, 2009, Paris.
5. T. Hayashi (Invited Speaker): Genome analysis of O157 and non-O157 EHEC strains. The 9th Korea-Japan International Symposium on Microbiology, Oct. 17, 2008, Seoul, Korea.

[図書](計1件)

1. 林哲也：細菌ゲノムと遺伝子構造、変異と修復、発現調節(第1編 微生物学総論 第2章 細菌学総論)，薬学領域の病原微生物学・感染症学・化学療法学 第2版，西島正弘，後藤直正 編集，pp.46-74，廣川書店、全416頁，2009.

〔産業財産権〕

出願状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕(計0件)

6. 研究組織

(1)研究代表者

林 哲也 (HAYASHI TETSUYA)
宮崎大学・フロンティア科学実験総合センター・教授
研究者番号：10173014

(2)研究分担者

小椋 義俊 (OGURA YOSHITOSHI)
宮崎大学・フロンティア科学実験総合センター・助教
研究者番号：40363585

大岡 唯祐 (OHOKA TADASUKE)
宮崎大学・医学部・助教
研究者番号：50363594

黒川 顕 (KUROKAWA KEN)
東京工業大学・生命理工学研究科・教授
研究者番号：20343246

(3)連携研究者

研究者番号：