

特集 最近の医療における感染症対策と研究の進歩**1: 最近話題の感染症 - ゲノム解析から臨床まで -****細菌ゲノムシーケンス**

桑原 知 巳

徳島大学大学院ヘルスバイオサイエンス研究部生体制御医学講座分子細菌学分野

(平成16年10月19日受付)

(平成16年10月26日受理)

細菌の有するさまざまな生物学的性質が食品・医薬品の製造や新たな研究技術の開発など、産業的に応用される一方で、細菌は感染症の病原体としてわれわれの生命を脅かす存在でもある。1940年代における抗生物質の登場で、細菌感染症が制圧されるのは時間の問題とされていたが、新興再興感染症の流行や抗生物質による治療が困難な多剤耐性菌の出現など、21世紀においても細菌感染症の脅威は増すばかりである。細菌の環境適応能力には目を見張るものがあるが、このような表現型の変化には遺伝情報の担い手であるゲノムの変化が基盤にあることは言うまでもない。近年、さまざまな病原細菌の全ゲノムシーケンスが明らかになるにつれ、ゲノム配列に刻み込まれた適応進化の痕跡が新興再興感染症の発生メカニズムについて語り始めている。細菌のゲノムシーケンス解析の現状を述べるとともに、そこから明らかになってきた最新の知見について紹介したい。

1. 細菌ゲノムシーケンスの状況

1995年、細菌としては初めての全ゲノム配列がインフルエンザ菌 (*Haemophilus influenzae*) において決定された (表1)。これは自律増殖できる生物としては初めての全ゲノム配列でもある。これに次いで、1997年に学術研究のモデル生物である大腸菌 (*Escherichia coli*) と枯草菌 (*Bacillus subtilis*) の全ゲノム配列が決定され、全ゲノム情報をベースとした網羅的な細菌研究の礎となった。細菌は菌種ごとに多様な生物学的性質や病原性を示す。個々の菌種を研究する者にとって、菌種間における生物学的性状や病原性の違いがどのようなゲノム構造の違いに由来するのかが極めて重要な問題であり、国内外でさまざまな細菌の全ゲノム配列決定が競って進められている。世界における細菌ゲノムシーケンスの現状は GOLD

Genomes Online Database (<http://www.genomesonline.org/>) で知ることができる。現在、200に近い細菌の全ゲノム配列が誌上発表されており、公開されているだけでも約500の細菌について全ゲノムシーケンスプロジェクトが進行中である。特に病原細菌のゲノムに対する関心は高く、誌上発表された細菌ゲノムのうち、約6割は植物やヒトの病原菌であり、結核菌や腸管出血性大腸菌など重要な病原菌については、ほぼゲノム配列決定が終了している (表1)。しかしながら、環境中や私たちの体の細菌叢を構成する未知の菌や産業的に利用価値の高い細菌のゲノムシーケンス競争が今後も続くと思われる。

2. シーケンス技術の進歩と問題点

1995年のインフルエンザ菌の全ゲノム配列決定から数年の間に次々と細菌ゲノムシーケンスが報告されるようになった背景にはシーケンス技術の著しい進歩がある (表2, 図1)。インフルエンザ菌のゲノムが決定された1995年から現在に至るまでの自動シーケンサーの能力を比較すると総塩基解読数は30倍にも向上している¹⁾。細菌の全ゲノムシーケンス決定にはショットガンシーケンス法という方法がとられている。このショットガンシーケンス法は (1) ショットガン工程と (2) フィニッシング工程から成る (図2)。ショットガン工程とは、まず、細菌のゲノムを超音波などの物理的処理によって1-2 kbの断片にし、これをクローニングベクターに連結後、大腸菌に導入してショットガンライブラリーを作成する。パズルで言うならば、まず、完成図をばらばらなピースに分けるような作業である。作成したライブラリーからランダムにクローンを選択し、鋳型 DNA を調整してシーケンスを行うわけである。

表1 これまでに全ゲノム配列が誌上発表された主要な細菌

発表年	細菌	サイズ (Mb)	主な発表施設	掲載誌
1995	<i>Haemophilus influenzae</i> Rd KW20	1.83	TIGR	<i>Science</i>
1995	<i>Mycoplasma genitalium</i> G 37	0.58	TIGR	<i>Science</i>
1996	<i>Mycoplasma pneumoniae</i> M129	0.82	Heidelberg 大学	<i>Nucleic Acids Res</i>
1997	<i>Helicobacter pylori</i> 26695	1.67	TIGR	<i>Nature</i>
1997	<i>Escherichia coli</i> K 12 MG1655	4.64	Wisconsin 大学	<i>Science</i>
1997	<i>Bacillus subtilis</i> 168	4.21	European consortium Japanese consortium	<i>Nature</i>
1997	<i>Borrelia burgdorferi</i> B31	1.23	Brookhaven Natl Lab.	<i>Nature</i>
1998	<i>Mycobacterium tuberculosis</i> H37Rv	4.41	Sanger Institute	<i>Nature</i>
1998	<i>Treponema pallidum</i> subsp. <i>pallidum</i> Nichols	1.14	Texas 大学	<i>Science</i>
1998	<i>Chlamydia trachomatis</i> D/UW 3/CX	1.04	Stanford 大学	<i>Science</i>
1998	<i>Rickettsia prowazekii</i> Madrid E	1.11	Uppsala 大学	<i>Nature</i>
1999	<i>Chlamydomonas reinhardtii</i> CWL029	1.23	Stanford 大学	<i>Nature Genet</i>
2000	<i>Campylobacter jejuni</i> NCTC11168	1.64	Sanger Institute	<i>Nature</i>
2000	<i>Neisseria meningitidis</i> MC58	2.27	TIGR	<i>Science</i>
2000	<i>Vibrio cholerae</i> serotype O1 strain N16961	4.00	TIGR	<i>Nature</i>
2000	<i>Pseudomonas aeruginosa</i> PAO1	6.26	Washington 大学	<i>Nature</i>
2001	<i>Escherichia coli</i> O157:H7 EDL933	5.53	Wisconsin 大学	<i>Nature</i>
2001	<i>Mycobacterium leprae</i> TN	3.27	Sanger Institute Institute Pasteur	<i>Nature</i>
2001	<i>Escherichia coli</i> O157:H7 Sakai	5.59	宮崎大学・北里大学・大阪大学	<i>DNA Res</i>
2001	<i>Streptococcus pyogenes</i> M1 SF370	1.85	Oklahoma 大学	<i>PNAS</i>
2001	<i>Staphylococcus aureus</i> N315(MRSA)	2.81	順天堂大学	<i>Lancet</i>
2001	<i>Streptococcus pneumoniae</i> ATCC-BAA 334	2.16	TIGR	<i>Science</i>
2001	<i>Yersinia pestis</i> CO 92	4.65	Sanger Institute	<i>Nature</i>
2001	<i>Salmonella</i> Typhi CT18	4.81	Sanger Institute	<i>Nature</i>
2001	<i>Salmonella</i> Typhimurium LT2 SGSC1412	4.86	Washington 大学	<i>Nature</i>
2002	<i>Clostridium perfringens</i> 13	3.03	筑波大学・北里大学・九州大学	<i>PNAS</i>
2002	<i>Clostridium tetani</i> Massachusetts E88	2.80	Goettingen Genomics laboratory	<i>PNAS</i>
2003	<i>Vibrio parahaemolyticus</i> RIMD2210633	5.17	大阪大学・北里大学・九州大学	<i>Lancet</i>
2003	<i>Bacteroides thetaiotaomicron</i> VPI 5482	6.26	Washington 大学	<i>Science</i>
2003	<i>Enterococcus faecalis</i> V583	3.21	TIGR	<i>Science</i>
2003	<i>Bacillus anthracis</i> Ames	5.23	TIGR	<i>Nature</i>
2003	<i>Bordetella pertussis</i> Tohama I NCTC13251	4.09	Sanger Institute	<i>Nature Genet</i>
2003	<i>Streptococcus pyogenes</i> M3 SSI 1	1.89	大阪大学・北里大学・九州大学	<i>Genome Res</i>
2003	<i>Porphyromonas gingivalis</i> W83	2.34	TIGR, Forsyth Dental Center	<i>J Bacteriol</i>
2003	<i>Lactobacillus johnsonii</i> NCC533	1.99	Nestle, North Carolina State 大学	<i>PNAS</i>
2004	<i>Legionella pneumophila</i> Philadelphia 1	3.40	Columbia 大学	<i>Science</i>
2004	<i>Bacteroides fragilis</i> YCH46	5.28	徳島大学・北里大学・九州大学	<i>PNAS</i>

TIGR : The Institute of Genome Research

表2 自動シーケンサーの解読能力の進歩

規 格	市販年	解読塩基数 / サンプル	1度に解読できる サンプル数	時間 / サイクル	総解読塩基数 / 台・日
ABI377 (スラブ式)	1995	500	96	8	10万
ABI3700 (キャピラリー式)	1998	600	96	3	46万
ABI3730 (キャピラリー式)	2002	600	96	1	138万
MegaBACE1000 (キャピラリー式)	1998	550	96	1.5	85万
MegaBACE4000 (キャピラリー式)	2001	550	384	1.5	338万

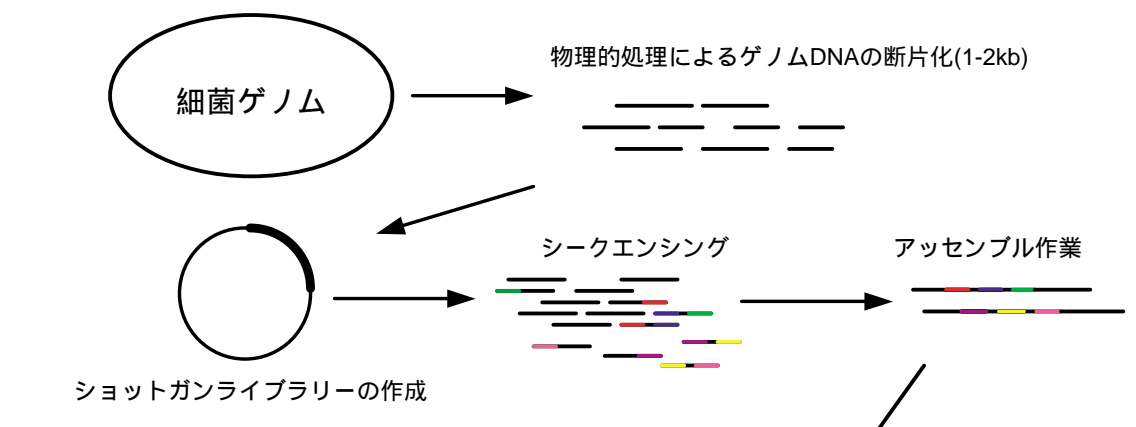


図1 Auto multi-capillary array sequencer, MegaBACE1000 および4000

配列を決定しようとする細菌のゲノムサイズの約5倍から10倍の長さに対応するシーケンスを集めるとほぼゲノム全体をカバーする塩基配列情報を得ることができる。1回のシーケンス反応で約500塩基しか解読できないので、小さな1 Mb (100万塩基対) のゲノムを決定する場合でも、最低10,000という膨大な数のシーケンス反応が必要になる。得られたシーケンスを相同な配列をもとにつないでいく作業をアッセンブルと言い、この段

階で1本の環状配列になれば完了となるが、現実はその簡単ではなく、多数のギャップ領域が残ることになる。このギャップ領域をPCRなどによって埋めていき、1本につながった後も、本当に決定された配列が正しいのかを実験的に確認しなければならない。また、配列の質が悪い領域については再シーケンスを行うなど、他の研究者が利用するに足る正確な配列であることを検証しなくてはならない。この過程をフィニッシング工程と呼ぶ。しかし、実際に登録されている幾つかの細菌ゲノム配列を見てみると A, G, C, T 以外の不確定塩基が数多く認められる。2001年、腸管出血性大腸菌 O157の全ゲノム配列が米国と日本の2つの研究グループから独立に発表されているが、先に Nature に掲載された EDL933株の配列には未だ決定されないまま放置された4 kbのギャップと2,600もの不確定塩基が残されている²⁾。一方、日本の研究グループから発表された腸管出血性大腸菌 O157堺株のゲノム配列には不確定塩基が存在せず、極めて精度の高い配列決定がなされている³⁾。ゲノム配列決定によって同定された遺伝子の機能をすでに登録されている遺伝子との相同性により予測し、遺伝子に名前を付与していく作業をアノテーションと呼ぶが、アノテーションの根拠に乏しい場合が少なくない。膨大な量

(1) ショットガン工程



(2) フィニッシング工程

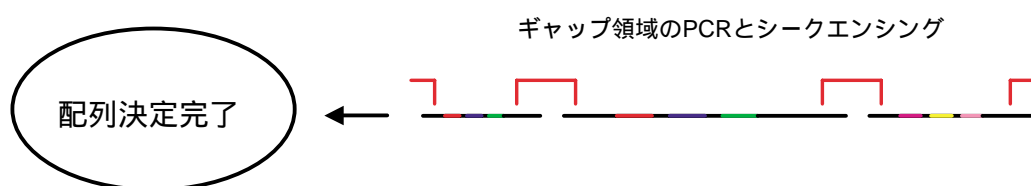


図2 ショットガンシーケンス法の流れ

に及ぶ細菌ゲノム配列情報は研究者にとっては大きな情報源であるが、ゲノム配列を利用する側においては、それぞれのゲノム配列のクオリティーや各遺伝子のアノテーションの根拠について十分な注意を払う必要がある。

3. 細菌ゲノムの特徴

細菌の染色体は1本の環状2本鎖DNAであると考えられていたが、全ゲノム配列決定によってライム病の病原体である *Borrelia burgdoferi* の線状染色体やコレラ菌などの *Vibrio* 属細菌の有する環状2本染色体など、細菌の染色体構造には菌種間で多様性があることが明らかになった。また、染色体の大きさにも大きな違いがあり、 *Mycoplasma genitalium* の0.58 Mbから *Streptomyces avermitilis* の9.12 Mbまで様々である。共通祖先からの長い適応進化の過程で遺伝子の水平伝播、遺伝子重複、遺伝子の欠失やゲノムの再構成が頻繁に生じた結果、このようなゲノム構造の多様性が生み出されたと考えられる(図3)。細菌が感染を成立させるためには、宿主への定着・侵入、宿主内での栄養源の獲得や宿主免疫からの回避など多くの機能が必要であり、そのための遺伝子群が必要となる。

事実、ヒトに病原性を示す菌種や常在菌のゲノムには病原遺伝子や宿主内での生存に有利に働くと推測される多数の遺伝子群が水平伝播によって挿入されている。

4. 細菌ゲノムのダイナミックな変化

前述したように、細菌のゲノム構造には菌種または菌株間で多様性が見られ、この違いがそれぞれの生物学的特徴の源になっている。特に感染症を引き起こす病原細菌は、進化の過程で非病原菌にはない毒素遺伝子などの病原遺伝子を水平伝播によって獲得している。病原細菌の比較ゲノム解析から、細菌ゲノムは現在もなお進化を続けており、細菌ゲノムのダイナミックな変化が新たな感染症の出現に密接に関連している可能性が示されている。1996年7月、堺市で10,000人も感染者を出した腸管出血性大腸菌と1980年代後半から症例が報告されるようになった劇症型A群溶連菌感染症の病原体である *Streptococcus pyogenes* のゲノム解析から明らかになってきた最新の知見について紹介する。

腸管出血性大腸菌 O157堺株と実験室内で研究によく用いられる非病原性の大腸菌 K12株の比較解析から、

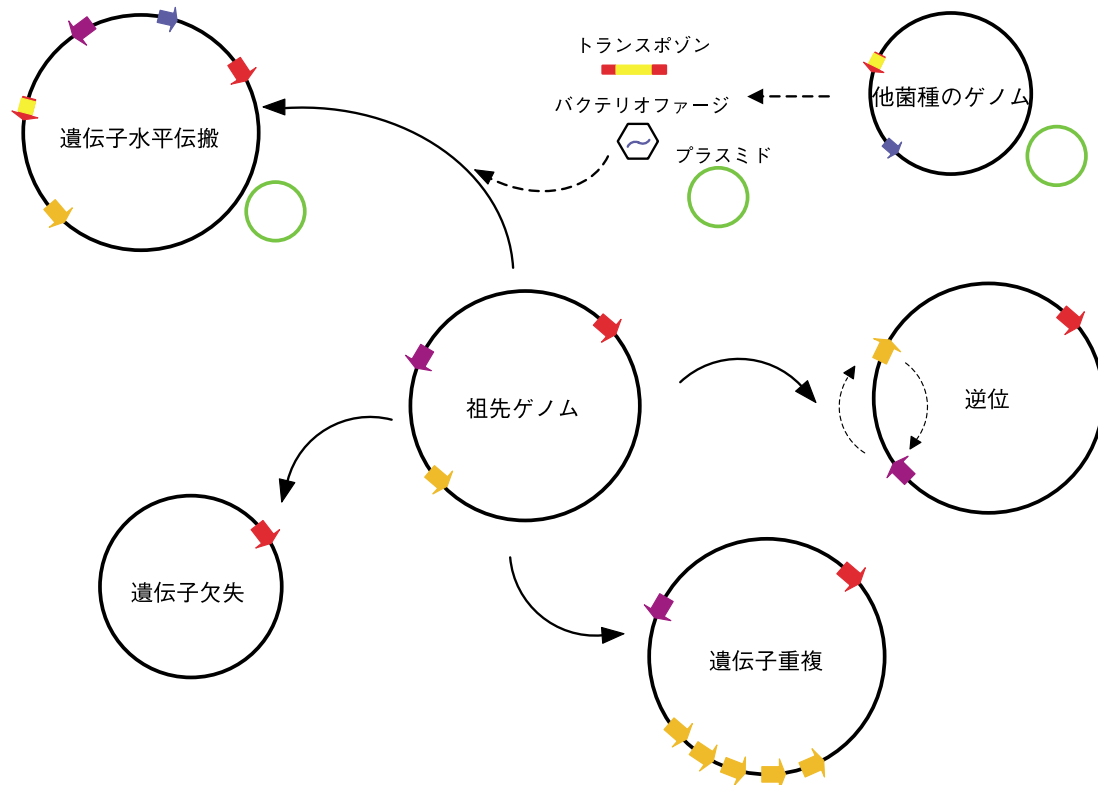


図3 細菌ゲノムの進化

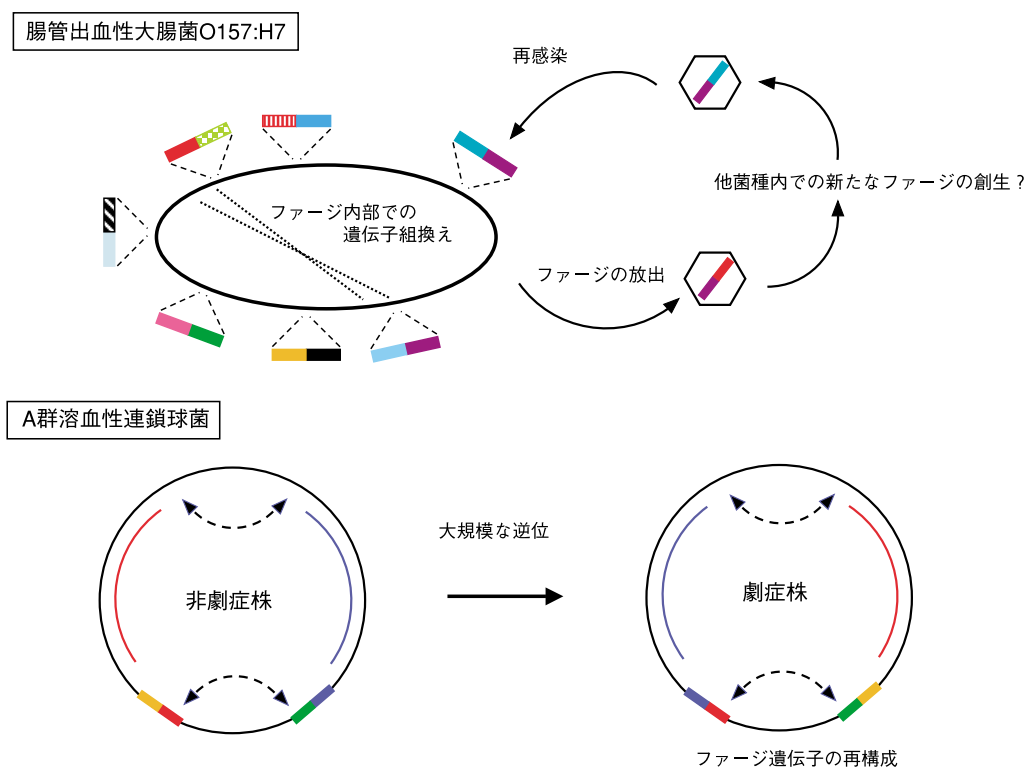


図4 細菌ゲノムの多様性を生み出すファージ遺伝子の再構成

そのゲノムサイズには1 Mb もの違いがあり、この大部分は志賀毒素遺伝子を有する Stx ファージなど24種類ものバクテリオファージの挿入によって作り上げられていることが明らかとなった³⁾。腸管出血性大腸菌のゲノム構造は菌株間においても多様性があり、この違いがパルスフィールド電気泳動による疫学調査に利用されている。この菌株間におけるゲノム構造の違いの大部分は挿入されたファージの遺伝子の構造多様性に起因しており、このファージの多様性は腸管出血性大腸菌 O157のゲノムに挿入された多数のファージ間での組換えによって生み出されていると考えられている⁴⁾(図4)。病原遺伝子を乗せたファージが細菌ゲノム内での遺伝子組換えによって新たな構造を有するファージへと進化し、菌株または菌種間を伝播する間にさらに構造を変化させて行く。このようなファージなどの外来遺伝子が細菌ゲノムのダイナミックな構造変化の原動力となっている可能性を示唆している。

2003年に劇症型 A 群溶血性連鎖球菌感染症 (TSLs) 患者から分離された *Streptococcus pyogenes* SSI 1株の全ゲノム配列が報告された⁵⁾。非劇症株との比較解析の結果、TSLs に見られる激しい症状を説明できるような病原遺伝子は認められていない。しかしながら、劇症株のゲノ

ムには大規模な逆位が認められ、この逆位はスーパー抗原やヒアルロニダーゼといった A 群溶連菌の病原遺伝子を乗せたファージ挿入部位で起こっていることが明らかになった(図4)。このことは、A 群溶連菌のゲノム上でも新たな構造を有するファージが生み出されており、また、外来性遺伝子であるファージが大規模なゲノムの再構成に関係していることを示している。TSLs は1980年代後半から我が国でも症例が報告されるようになった再興感染症であるが、このようなゲノムの大規模な逆位が1990年以降に TSLs 患者から分離された劇症株の81%に認められ、1985年以前に分離された菌株においてはわずか25%にしか認められないと報告されている⁵⁾。このことは、ゲノム構造のダイナミックな変化が病原遺伝子群の発現に何らかの影響を及ぼし、新たな症状を示す感染症の出現に関与している可能性を示唆している。

5. ゲノム配列が語る腸内菌の宿主への適応戦略

最後にわれわれの研究グループが行った腸管内常在菌の一つである *Bacteroides* のゲノム解析について少し紹介したい⁶⁾。われわれの腸内細菌叢は様々な菌種によって構成されており、その種類は400を超えと言われてい

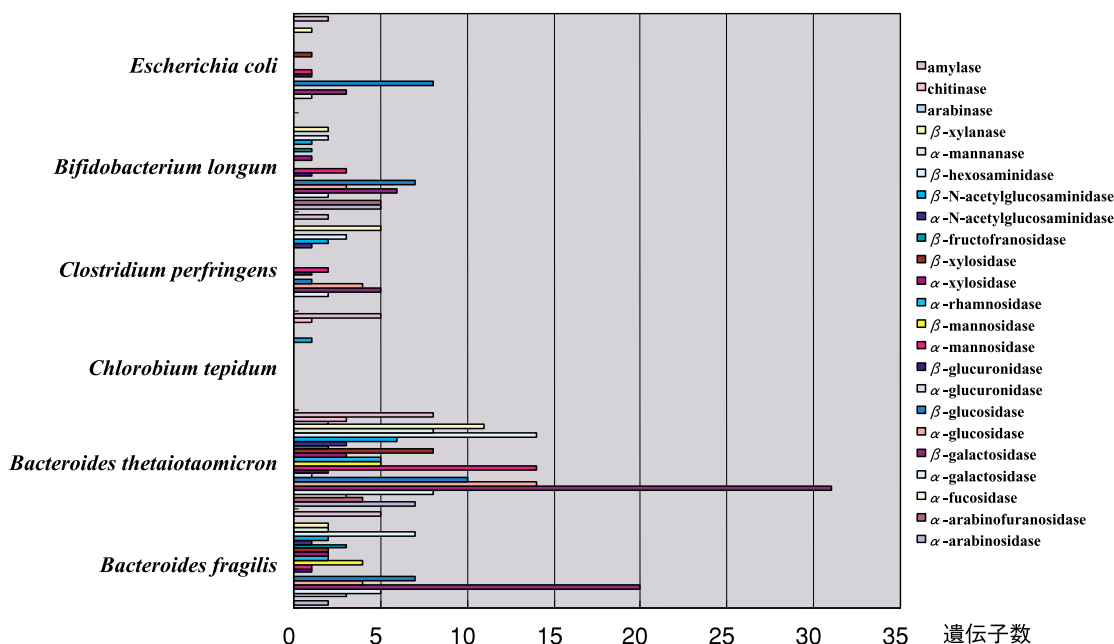


図5 腸管内常在菌のゲノム上に見られる糖鎖分解酵素遺伝子群の重複

る。また、腸内容物 1 g あたり 10^{11} 個もの細菌が存在する。腸内菌のうち、人工培地での培養が可能な菌種は約 3 割程度であるが、その中で最も数が多いのが *Bacteroides* 属である。*Bacteroides* は主として大腸に常在する偏性嫌気性グラム陰性桿菌であるが、他の属と同様に様々な菌種を含んでおり、各菌種によって腸内での定着状態や病原性に違いがある。*Bacteroides* の中で最も病原性が強いのが *B. fragilis* であり、大腸粘膜に密に付着している。現在、*Bacteroides* の中で、*B. fragilis* と *B. thetaiotaomicron* の 2 菌種について全ゲノム配列が報告されている。*B. thetaiotaomicron* は *B. fragilis* よりも病原性は弱いですが、大腸内での菌数は *B. fragilis* の約 100 倍にも及び、図 5 は *Bacteroides* とその他の腸内菌である大腸菌、*Bifidobacterium longum*、*Clostridium perfringens* と腸内常在菌ではないが、*Bacteroides* の近縁菌種である *Chlorobium tepidum*

のゲノム上に存在する多糖分解酵素遺伝子の種類と数を比較したものである。大腸に常在する *Bacteroides* や *Bifidobacterium*、*Clostridium* のゲノム上では数多くの多糖分解酵素遺伝子群が存在している。これらに比較して上部消化管常在菌である大腸菌のゲノム上には糖鎖分解酵素遺伝子群は比較的少ない。宿主が摂取した栄養素のうち、単糖や二糖類などの低分子栄養素は上部消化管において宿主や常在菌に吸収され、下部消化管にはわずかな量しか到達しない。したがって、下部消化管に常在する菌は宿主や上部消化管の常在菌が吸収できない難分解性の多糖を栄養素として利用しなければならない。多数の糖鎖分解酵素遺伝子群はこのような環境に適応するための大腸常在菌のゲノム進化をよく現している。特に *Bacteroides* において糖鎖分解酵素遺伝子群の遺伝子重複は顕著であり、*Bacteroides* が大腸最優勢菌であること

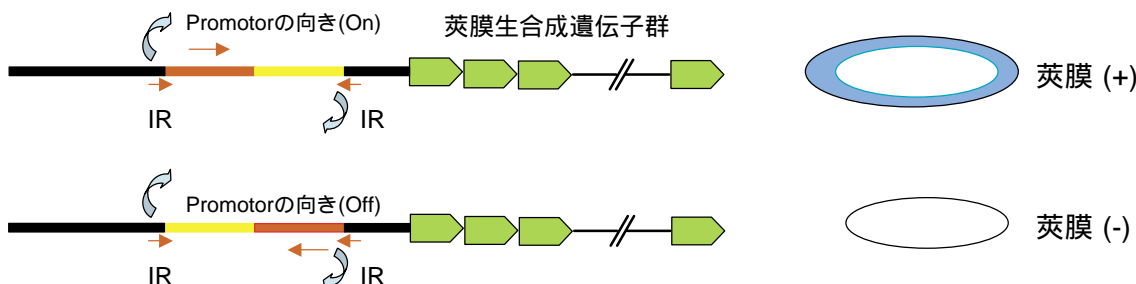


図6 DNA inversion による遺伝子の発現調節
DNA inversion は inverted repeat (IR) 配列と呼ばれる互いに逆向きの配列内で起きる。

とよく相関している。一方, *Bacteroides* の近縁種で湖などの環境水中に常在する *Chlorobium tepidum* のゲノムではこのような遺伝子重複は認められない。生息環境の違いが細菌ゲノムの進化にどれほど大きな影響を及ぼしているのかを再認識させられる。また, *Bacteroides* は他の菌種に類を見ないほど多くの莢膜生合成遺伝子領域をゲノム上に有している (*B. fragilis* で9箇所, *B. thetaiotaomicron* で7箇所)。さらに, *Bacteroides* ではプロモーターの向きを変化させる DNA inversion (逆位) によって多くの莢膜生合成遺伝子の発現が調節されている (図6)。ゲノム解析の結果, *Bacteroides* のゲノム上には莢膜生合成遺伝子領域以外にも外膜蛋白質などの表層構造の構築に関与する遺伝子群の発現を DNA inversion によって on-off 制御している領域が多数存在することが明らかになった。このような DNA inversion を起こす領域の数は *B. thetaiotaomicron* よりも *B. fragilis* において圧倒的に多く, *B. fragilis* はより宿主免疫に認識されにくい複雑な表層抗原性を作るシステムを持っている。この違いが両者の大腸内での局在 (*B. fragilis* はより粘膜面に近いところに常在している) や病原性の差を生み出していると考えられる (図7)。同じ大腸という常在部位においても細菌はわずかな環境の違いに独自の戦略で適応

し, 定着部位を分かち合いながらバランスを保っていることが伺える。

おわりに

抗生物質の登場によって細菌研究はすでに終わったと考える人がいるかもしれない。しかし, 地球上に存在する細菌のうち, 私たちが生物学的性状について知識を持っているものは1%に満たないと言われている。私たちの生命活動に密接に関係している腸内菌でさえ, 菌種として認識されているものは約3割である。土壌や糞便の中に存在する培養不能な菌をそのゲノム全塩基配列より理解できる時代になってきており, 彼らが長い時間をかけて進化させてきたゲノムの中には, 私たちの想像を超える有用な遺伝子システムが眠っているに違いない。これまで述べてきたように, 細菌ゲノムは今なおダイナミックに進化を続けており, このゲノムの変化が新たな感染症の発生に密接に関係している。細菌の全ゲノムシーケンスは彼らの驚異の環境適応能力を語っているが, ゲノム進化の最も大きな原動力は生息環境の変化であり, 抗生物質の開発, 食文化の変化, 自然破壊など細菌ゲノムの進化を加速させる機会を与えているのはわれ

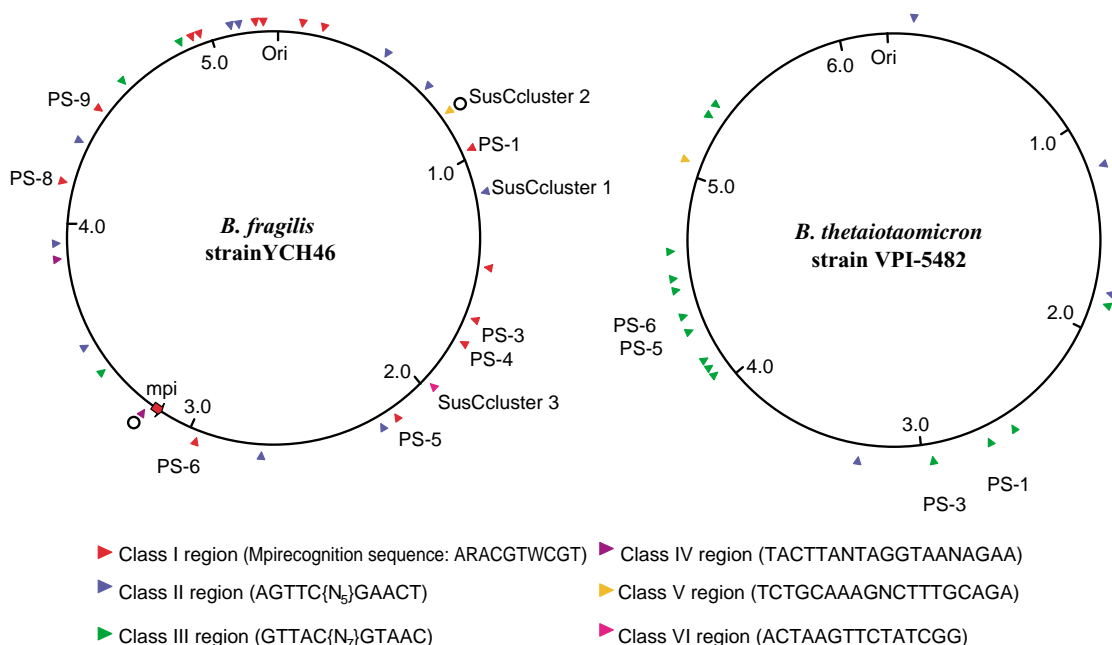


図7 *Bacteroides* のゲノム上に多数存在する DNA inversion を起こす領域

B. fragilis のゲノム上には promoter 領域の DNA inversion によって莢膜生合成遺伝子 (PS) や外膜蛋白質遺伝子 (SusC) など菌体表層構造の構築に関与する遺伝子の発現が on off 制御されている領域が多数存在する (31箇所)。DNA inversion の起点となる inverted repeat 配列内のモチーフ配列によってクラス分類してある。

われかもしれない。

文 献

- 1) 服部正平 : 微生物ゲノムシーケンス決定法 . 第18回「大学と科学」公開シンポジウム講演抄録集 微生物はなぜ病気をおこすか ゲノムの特徴 (林 英生 編), 第1版, 技報堂, 東京, 2003, pp 33-45
- 2) Perna, N.T., Plunkett, G III., Burland, V., Mau, B., *et al.* : Genome sequence of enterohemorrhagic *Escherichia coli* O157 : H7. *Nature*, 409 : 529-533, 2001
- 3) Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., *et al.* : Complete genome sequence of enterohemorrhagic *Escherichia coli* O157 : H 7 and genomic comparison with a laboratory strain K 12 . *DNA Res.*, 8 : 11-22, 2001
- 4) Ohnishi, M., Terajima, J., Kurokawa, K., Nakayama, K., *et al.* : Genomic diversity of enterohemorrhagic *Escherichia coli* O157 : H 7 revealed by whole genome PCR scanning. *Proc. Natl. Acad. Sci. USA.*, 99 : 17043-17048, 2002
- 5) Nakagawa, I., Kurokawa, K., Nakata, M., Yamashita, A., *et al.* : Genome sequence of an M 3 strain of *Streptococcus pyogenes* reveals a large scale genomic rearrangement in invasive strains and new insights into phage evolution. *Genome Res.*, 13 : 1042-1055, 2003
- 6) Kuwahara, T., Yamashita, A., Hirakawa, H., Nakayama, H., *et al.* : Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. *Proc. Natl. Acad. Sci. USA.*, 101 : 14919-14924, 2004

Determination of whole DNA sequences of bacterial genomes

Tomomi Kuwahara

Department of Molecular Bacteriology, Institute of Health Biosciences, The University of Tokushima Graduate School, Tokushima, Japan

SUMMARY

Bacteria exhibit unique biological characteristics at species or strain levels, and it has become possible to understand their diversities by analyzing whole genome sequences. Almost 200 bacterial genome sequences have so far been published, and determination of the complete genomes of nearly 500 bacteria is now in progress. Developments in sequencing technology and improvements in automated sequencers have contributed to the rapid accumulation of genome sequence data.

Comparative analysis of bacterial genomes has revealed that there are extensive diversities in their structures such as linear chromosomes of *Borrelia burgdorferi* and two chromosomes of the genus *Vibrio*. The diversities have been established by a combination of horizontal gene transfer, gene duplication, deletion, and/or genomic rearrangements during the process of adaptation over a long period to each environment. These changes in bacterial genomes, especially in pathogenic bacteria, are related to the occurrence of novel types of infection or multi drug resistance. On the other hand, genomic analysis of a gut commensal, *Bacteroides fragilis*, has revealed that this species dynamically changes the genomic structure within a short period of time by multiple DNA inversions that create diverse surface antigenicities to evade the host immune system. Thus, whole genome sequencing provides important information on adaptation strategies of each bacterium. However, in an environmental ecosystem such as soil, water, and human microflora, a large number of bacteria interact with each other and comprise a functional unit. Community genomics, which targets all of the bacterial genome sequences included in a particular environmental ecosystem, is expected to provide novel insights into microbe microbe and host microbe interactions.

Key words : bacterial genomics, whole genome sequencing, adaptation