

黄色ブドウ球菌における病原性遺伝子の多様化メカニズム： 同一種内多数ゲノム比較によるゲノム進化解析

東京大学新領域創成科学研究科および東京大学医科学研究所

小林 一三

Staphylococci are normal inhabitants on skin and mucous membrane of humans and can become pathogenic. Especially, *Staphylococcus aureus* can cause severe inflammation in various tissues including skin which sometimes results in serious disorder. In addition, this bacterium has developed resistance to practically all types of antibiotics. Due to its clinical importance, whole genome sequences of multiple *S. aureus* strains have been decoded. To date, complete genome was sequenced for more than ten strains and has become publicly available. In their genome, several tandem clusters of paralogous genes, likely pathogenicity-related, have been found in genomic islands. Intergenomic comparison with respect to these clusters revealed polymorphisms in them. In the case of *lpl* gene cluster, encoding lipoprotein homologues, the variation was very extensive. Our multiple sequence alignment revealed presence of a region highly conserved not only at the amino acid sequence level but also at the nucleotide sequence level and regions to its 5' and 3' sides, which are more variable. The highly conserved nucleotide sequences are likely to have provided a site for homologous recombination generating the variation of this region. Comparison of phylogenies of the 5'-variable region and the 3'-variable region revealed significant incongruence within the same ORF. By contrast, pairs of 3'-variable region of an ORF and 5'-variable region of its downstream ORF gave more congruent phylogenies with groups of conserved pairs, which suggested their linkage. An intergenic region sandwiched by such a pair of variable regions seemed to have co-evolved with it. These lines of observations supports our hypothesis that homologous recombination at the central conserved region have played a major role in generating variations in this cluster. This model explains not only formation of various types of rearrangements through multiple crossing-over events but also generation of a novel ORF with different sets of two variable regions. The crossing-over events caused extensive shuffling of the two variable regions in one ORF, but maintained a conserved unit comprising 3'-variable region, intergenic region, and 5'-variable region spanning adjacent ORFs. This characteristic mode of tandem paralogue diversification, maintaining 3'-part of a gene, intergenic region, and 5'-part of its downstream gene as a unit of evolution, is unique among previously studied paralogue rearrangements, in which an ORF tended to have been considered as the unit of evolution.

1. 緒文

ブドウ球菌属細菌 *Staphylococcus* はヒトの皮膚や粘膜に常在する細菌として重要である。この中で特に病原性が問題となるのは黄色ブドウ球菌 *Staphylococcus aureus* で、ニキビや吹き出物等の傷口から皮膚に侵入し化膿性疾患を引き起こす。とびひ等の深刻な皮膚疾患の原因菌でもあるほか、時に血管やリンパ管を介して感染が拡大すると更に重篤な疾患をも引き起こす。黄色ブドウ球菌はまた、MRSA、VRSA 等、抗生物質が効かない系統の病院内感染・市中感染が社会的な問題ともなっている。

こうした臨床上の重要性から、黄色ブドウ球菌は本稿執筆時点で 10 を超える菌株の全ゲノム配列が解読され、公開されている (<http://www.ncbi.nlm.nih.gov/Genomes/>)。これらのゲノムには、抗生物質耐性に関わる遺伝子群、種々

の毒素遺伝子群、宿主細胞への接着や抗原抗体反応に関わる因子の遺伝子群、加水分解酵素の遺伝子群など、多様な病原性因子が見出されてきた¹⁻⁷⁾。また、既に病原性に関わることが知られていた遺伝子に加えて、これらのホモログである病原性関連の候補遺伝子も見つかってきている。こうした病原性遺伝子の多くは、「動く遺伝子単位」であるファージやゲノミック・アイランド上に存在し、これらに乗って染色体に組み込まれていることが明らかになった⁸⁾。

著者らはこれまでに、黄色ブドウ球菌のゲノム解読に参加し、抗生物質耐性やその他の病原性因子を乗せたゲノミック・アイランド上に、「動く遺伝子」の侵入・維持を制御する制限酵素の遺伝子が存在することに注目してきた¹⁾。また、やはりゲノミック・アイランド上でよく似た病原性関連遺伝子が縦につながってクラスターを形成している、タンDEM・パラログ・クラスターを複数株で比較し、その再編の分子機構を推定した⁹⁾。

本研究では、黄色ブドウ球菌のゲノムが造り替えられ、多様な病原性が獲得されていく機構を解明すべく、以下の2つの解析を行った。(1) ゲノミック・アイランド上の制限酵素遺伝子の活性の有無を実験的に検討する。(2) 著者らの先行研究⁹⁾で著しい再編が示唆された、ゲノミック・アイランド vSaα 上の病原性関連遺伝子 *lpl* のタンDEM・



A mechanism to generate variation in pathogenicity-related tandem paralogues of *Staphylococcus aureus*

Ichizo Kobayashi

Graduate School of Frontier Sciences,
University of Tokyo & Institute of
Medical Science, University of Tokyo

パラログ・クラスターについて、多数のゲノムで比較を行い、ゲノム再編の機構を明らかにすること。(1) については、ファージアッセイによって制限酵素活性を検討した(H. Yuzawa & I. Kobayashi, unpublished)。本稿では(2) について、以下に詳細な結果を報告する。

2. 解析

2.1 配列抽出

NCBI ゲノム配列データベース (<http://www.ncbi.nlm.nih.gov/Genomes/>) に2007年8月末の時点で登録されていた黄色ブドウ球菌9株のゲノム配列において、ゲノミック・アイランド ν Sa α 上に存在するリポプロテイン様遺伝子 *lpl* のホモログを、BLASTP および TBLASTX (<http://www.ncbi.nlm.nih.gov/blast/>) を用いた相溶性検索によって同定した。

2.2 配列解析に使用したプログラム

マルチプルアラインメントの作成には ClustalW (<http://www.ebi.ac.uk/Tools/clustalw/>) を用いた。Similarity plot は PLOTCON (<http://emboss.sourceforge.net/>) により作成した。系統樹は近隣接合法を用いて MEGA4.0 (<http://www.megasoftware.net/>) により作成した。

3. 結果

3.1 *lpl* 遺伝子の構造

9株 (N315, Mu50, MW2, COL, MRSA252, MSSA476, NCTC8325, USA300, RF122) のゲノム配列から同定した *lpl* 遺伝子全ホモログに対して、塩基配列およびアミノ酸配列を用いてマルチプルアラインメントを作成した。塩基配列レベルおよびアミノ酸配列レベルの similarity をプロットした結果を Fig. 1 に示す。なお、N315株と Mu50株、MW2株と MSSA476株では、該当する遺伝子および遺伝子間領域が完全に一致しており、本解析では N315株、MW2株の配列のみを用いた。ORF の内部において、132 bp (44 a. a.) で gap を含まず、アミノ酸配列レベルでも塩基配列レベルでもよく保存された配列が見出され、この領域を central conserved region と定義した。そして、この領域の 5'-側および 3'-側の可変領域をそれぞれ、5'-variable region、3'-variable region と定義した。2つの可変領域は塩基長や配列組成に多様性が見られた。リポプロテイン共通の、脂質修飾を受けるシステイン残基と、その N 末側のシグナルペプチドモチーフは、5'-variable region に含まれていた (Fig. 1)。

一般に、相同組換えは組換えを行う配列間に十分な長さで高い identity を要求する^{10, 11)}。この過程について詳細な解析が行われている近縁の枯草菌では、最小の長さは 70bp と報告されている¹²⁾。塩基配列レベルでも高い

identity を保っている central conserved region は相同組換えを引き起こしてきたと考えられた。

3.2 可変領域の系統樹比較

仮に ORF 中央の central conserved region を介した相同組換えが繰り返されてきたとすると、ある ORF の 5'-variable region と 3'-variable region は交叉 crossing-over によってその組み合わせを変えてきたと考えられる。一方、ある ORF の 3'-variable region とその下流の ORF の 5'-variable region とでは、この過程において組み合わせが保存されてきたことが予想される。

そこで、二つの可変領域の塩基配列系統樹の比較を試みた (Fig. 2)。1つの ORF を構成する 5'-variable region と 3'-variable region のペアで線を結んだ場合 (Fig. 2 (A))、系統樹間をつなぐ線は頻繁に交差し、予想通りこの組み合わせは進化の過程でシャッフルを起こしてきたことが示唆された。一方で、ある ORF の 3'-variable region とその下流の ORF の 5'-variable region というペアで線を結んだ場合 (Fig. 2 (B))、系統樹間をつなぐ線はほと

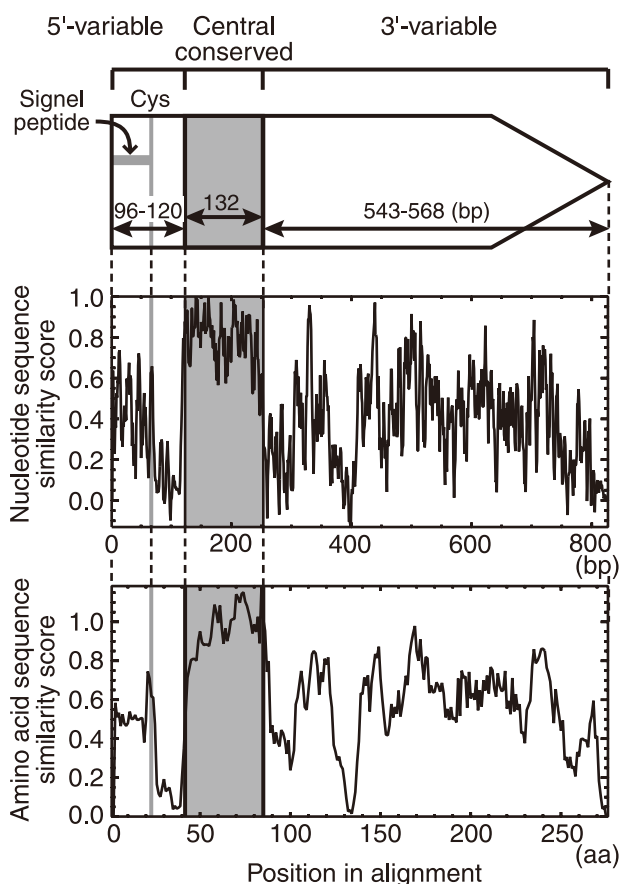


Fig. 1 *lpl* 遺伝子の構造 (上)、塩基配列に基づく similarity plot (中) およびアミノ酸配列に基づく similarity plot (下)。ORF 中央の保存領域を灰色で示す。(Tsuru & Kobayashi, unpublished)

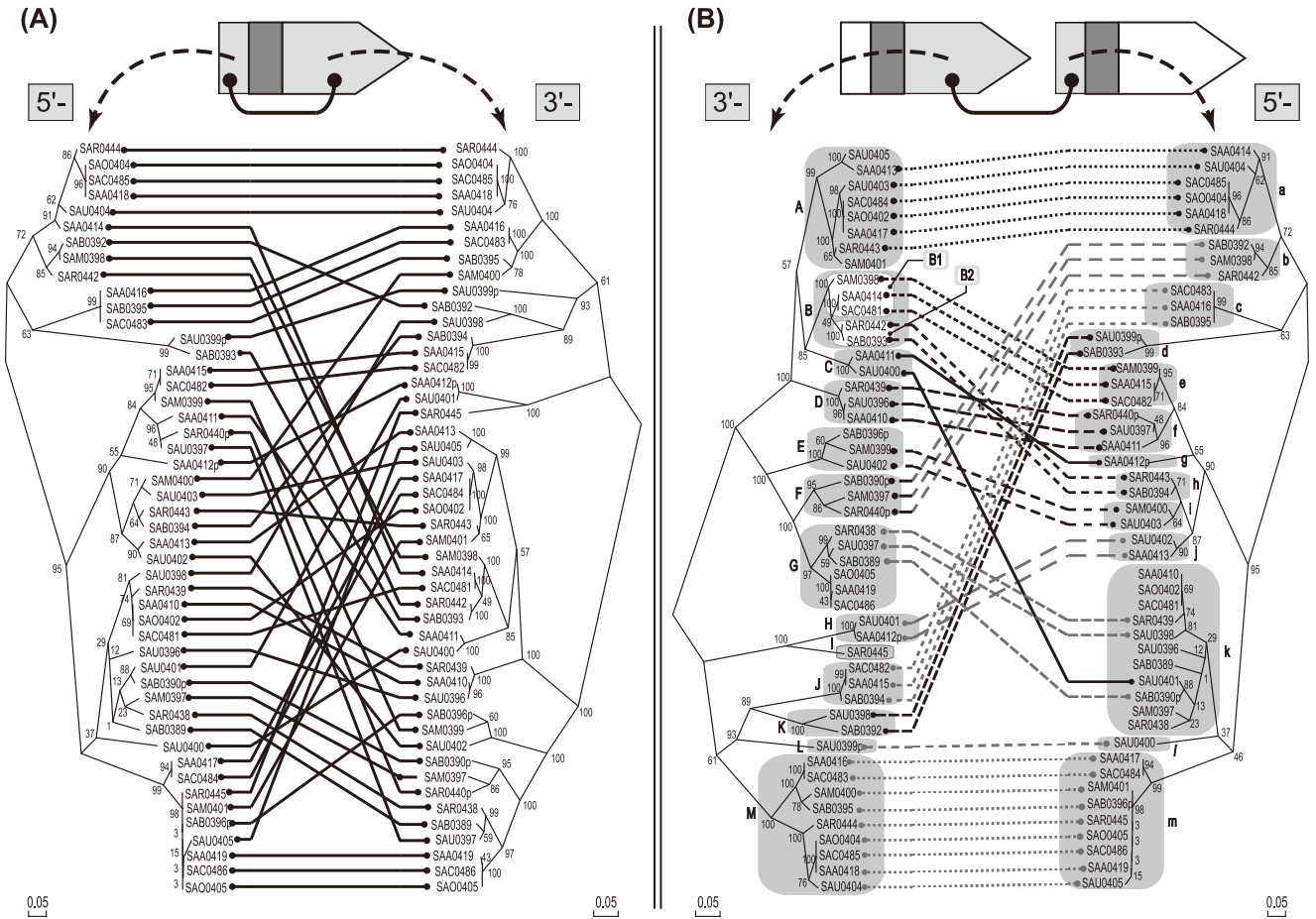


Fig. 2 5'-variable region と 3'-variable region の系統樹比較。塩基配列を用いて近隣接合法により作成したそれぞれの系統樹を、一つの ORF の 5' 側と 3' 側という組み合わせ (A)、およびある ORF の 3' 側とその下流の ORF の 5' 側という組み合わせ (B) で比較したものを示す。(B) では、配列が保存されたペアー同士の種類が、それぞれ異なる破線のパターンで示されている。このペアーの分類に対応する、各系統樹での単系統群・側系統群の分類が、長方形で囲ってある。(Tsuru & Kobayashi, unpublished)

んどの場合、互いに平行な線でつながれるグループに分類できることが見出された。加えて、この平行な線のグループは、それぞれの可変領域の系統樹における internal branch の短い単系統群ないしは側系統をつなぐように形成されていた。これは、お互いに配列が保存されたペアの存在を示しており、やはり予想と一致する結果であった。

3・3 遺伝子間領域の比較

Fig. 2 (B) において保存されていることが示された、「ある ORF の 3'-variable region とその下流の ORF の 5'-variable region」という組み合わせのペアについて、それらの間に挟まれた遺伝子間領域の配列を比較すると、遺伝子間領域でも配列が保存されていることが明らかとなった。従って、このタンDEM・パラログ・クラスターでは、隣り合う 2 つの ORF をまたいで、「3'-variable region、遺伝子間領域、5'-variable region」を一つの単位とする、保存された配列のグループが存在することが示された。

3・4 ゲノム間比較とゲノム再編素過程の再構築

二つの可変領域にたいする分類を各ゲノムの Map 上に表示したものを Fig. 3 (A) に示す。ゲノム間を比較すると、遺伝子の順序は保存しておらず、またどの分類の配列セットを持つかがゲノムごとに異なっており、この領域が激しく再編成を起してきてきたことが示唆された。その中で、ORF を単位としてみると、1 つの ORF の中では 5'-variable region と 3'-variable region との組み合わせがシャッフルされており、多様なホモログが存在している。他方、「ある ORF の 3'-variable region とその下流の ORF の 5'-variable region」という組み合わせの保存されたペアは、このような再編成を経た中でも維持されてきた様子が見取れる。

Pairwise のゲノム間比較において、次のような再編成の素過程を見出すことが出来た。(1) 挿入 / 欠失が USA300 株 (k-D-f-C-g-H-j-A-a-B1-) と COL 株 (k-B1-) の間に見られた。(2) 転位が N315/Mu50 株 (-D-f-G-k-) と MSSA252

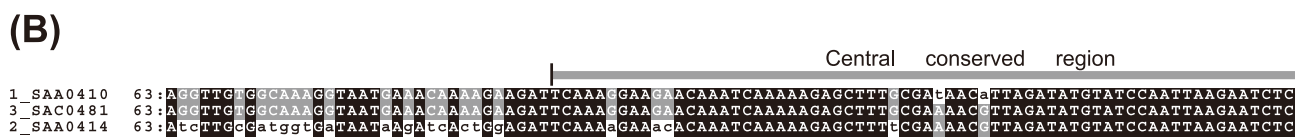
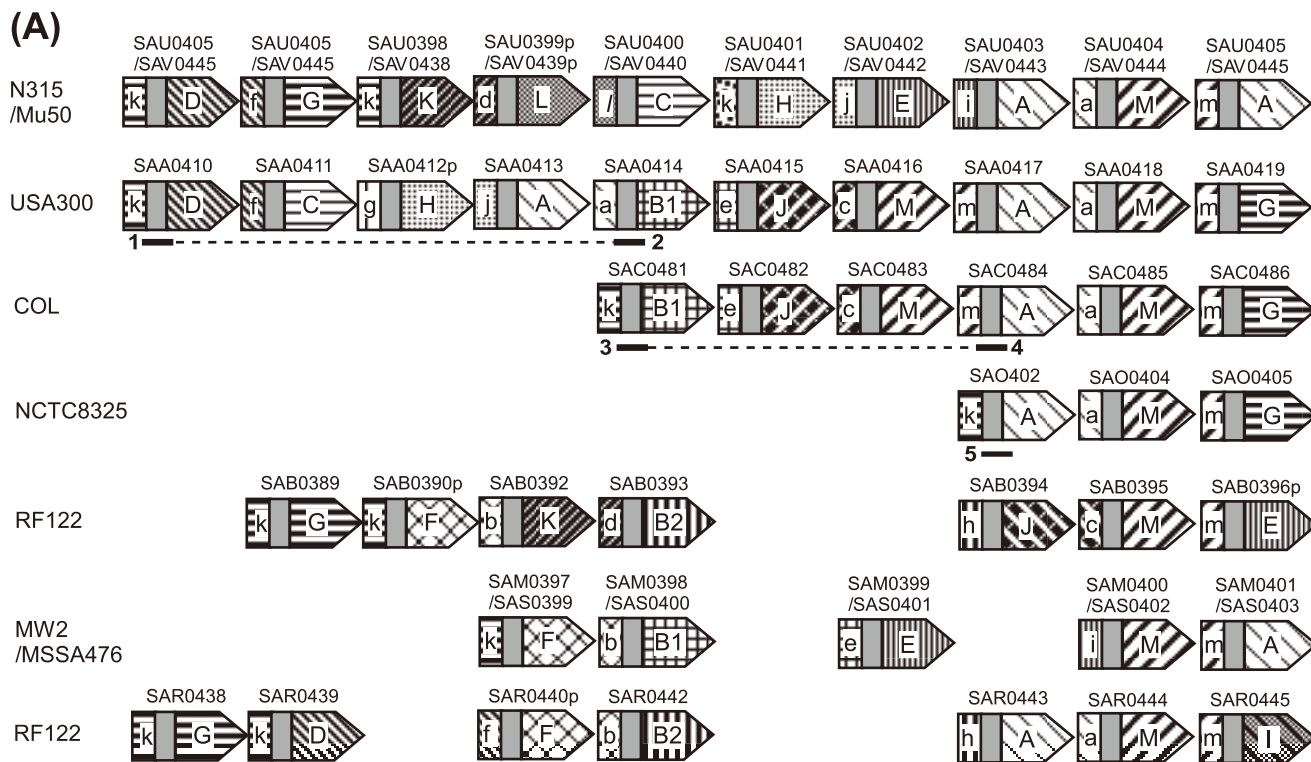


Fig. 3 (A) 各株における *lpl* タンデム・パラログ・クラスターの map。アルファベットで示した 5'-variable region (小文字) および 3'-variable region (大文字) に対する分類は、Fig. 2 (B) での分類に基づく。USA300 株・COL 株間および COL 株・NCTC8325 株間の挿入 / 欠失について想定される欠失イベントについて、欠失領域を破線で、その境界領域を 1 から 5 の太線で示す。(B) USA300 株・COL 株間で想定される欠失イベントでの、境界領域のアラインメント。境界領域 3 は、境界領域 1 と境界領域 2 の間の組換えによって生じたことが示唆される。Central conserved region をアラインメントの上に表示した。(C) COL 株・NCTC8325 株間で想定される欠失イベントでの、境界領域のアラインメント。(B) と同様。こちらでは、境界領域 5 が境界領域 3 と境界領域 4 の組換えによって生じたことが示唆される。(Tsuru & Kobayashi, unpublished)

株 (-G-k-D-f) 間で見られた。(3) 置換は RF122 株 (-b-K-d-B2-h-J-c-M-) と MW2/MSSA476 株 (-b-B1-e-E-i-M-) の間で見られた。また、(4) USA300 株に存在する離れた二つの A-a は遺伝子変換 gene conversion を示唆している。

このタンデムパラログクラスターで見られる多様性は、これらの再編素過程の繰り返しによって生成されてきたと考えられる。USA300 株・COL 株間および COL 株・NCTC8325 株

間の挿入/欠失について想定される、欠失イベントの境界のアラインメントを Fig. 3 (B), (C) に示す。ここでは、3本の配列のうち中央の配列が上流で1番目の配列と完全一致（3本目とは部分一致）していたものが、下流では3番目の配列と完全一致（1本目とは部分一致）するようになる、という変遷がみられる。組換え点と推定されるこの領域は、central conserved region によく一致していた。この結果は、これらのイベントの分子機構が central conserved region での相同組換えであった可能性を支持している。次節のモデルで示すように、その他の再編素過程も、同様に central conserved region での相同組換えという分子機構によって生じてきたとして説明できる。

3・5 Central conserved region での相同組み換えによる再編・多様化モデル

Fig. 4 に、今回の解析に基づくタンデムパラログ遺伝子の再編・多様化モデルを示す。配列レベルの多様性が生じる過程において、ORF 内部の central conserved region が特異的に保存され、この部分が二つの可変領域に挟

まれるような構造が形成されたと考えられる。Central conserved region は相同組換えの足場となり、更なる再編を生じてきたのだろう。すなわち、交叉を通じ ORF 内で2つの可変領域の組み合わせを交換し、ORF という遺伝子発現の単位での多様性の創出に寄与してきた。この過程で、隣接する ORF をまたいで、「上流 ORF の 3'-variable region、遺伝子間領域、下流 ORF の 5'-variable region」という単位は保存されてきたと考えられる (Fig. 4 (A))。

交叉の繰り返しはまた、Fig. 4 (B) で示したように、Fig. 3 (A) で観察されたような様々な再編の素過程を生じうる。現在のゲノムのこの領域に見られる著しい多型の生成はこれらの再編素過程の積み重ねによって生じたと考えられる。

4. 考 察

今回のモデルにおいては ORF 内部の配列を介した相同組み換えによって、2つの可変領域の交換による ORF レベルの多様性の創出と、ゲノム間のパラログ・クラスター比較で見られる遺伝子順序やホモログ・セットの多様性の

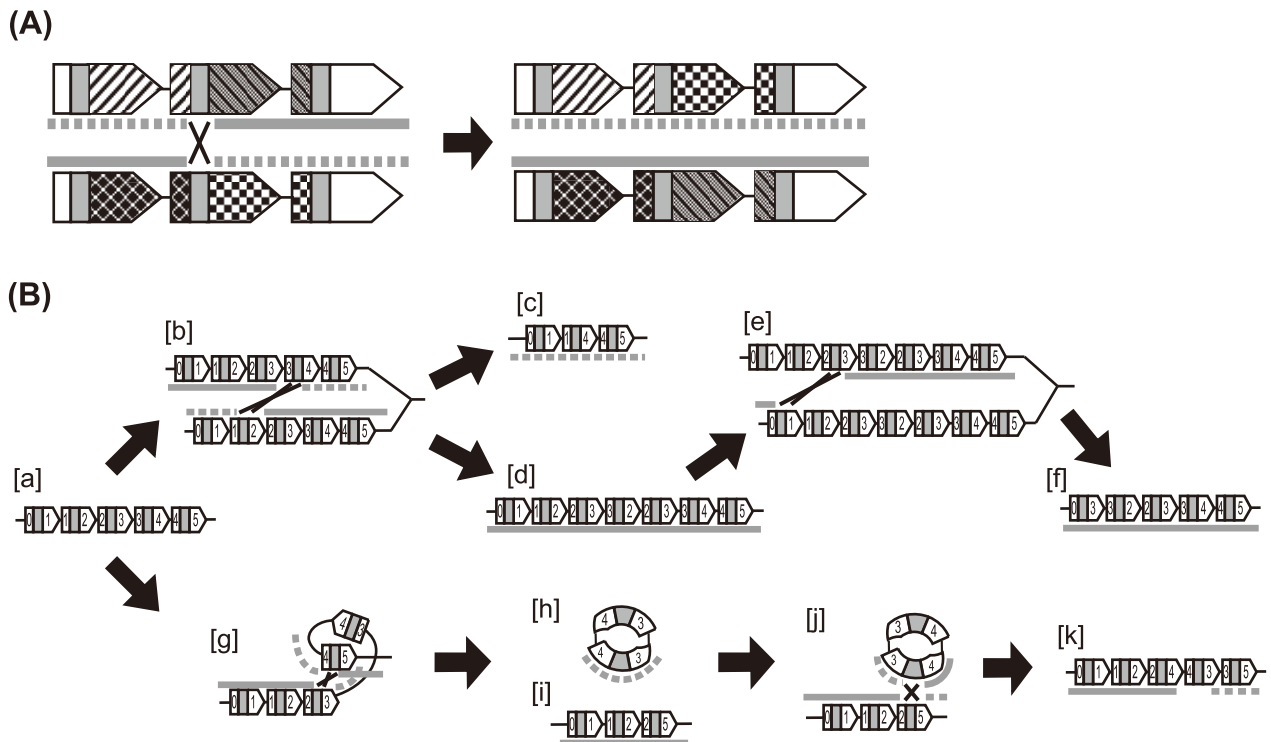


Fig. 4 ORF 内部の保存配列を介した相同組換えによる多様化モデル。(A) 相同組換えに伴う交叉によって、一つの ORF の 5'-variable region と 3'-variable region の組み合わせが交換されるが、ある ORF の 3'-variable region と隣接する下流の ORF の 5'-variable region の組み合わせは保存される。(B) このような交叉の繰り返しによって様々な再編が生じうる。姉妹染色体間の不等交叉 [b] により、欠失 ([a] から [c]) と重複 ([a] から [d]) が生じる。さらに不等交叉 [e] が繰り返されると、遺伝子変換 ([a] から [f]) が生じうる。染色体内の不等交叉 [g] は、環状 DNA [h] の切り出しと欠失 ([a] から [i]) を生じる。環状 DNA が再挿入される [j] と、転位 ([a] から [k]) が生じる。二通りの欠失 ([a] から [c]、[a] から [i]) によって、置換 ([c] と [i]) が生じうる。

創出の両方を説明することができた。この多様性の基となる、(異なるグループの間で見られる) 配列レベルの多様性は、今回のモデルとは異なる進化的次元で形成されてきたと考えられる。lpl 遺伝子ホモログを載せたゲノミック・アイランド vSa α は、ブドウ球菌属細菌の中で黄色ブドウ球菌に種特異的に見られる。従って多様な配列を有するホモログ・セットは vSa α が黄色ブドウ球菌の祖先型に挿入される前に既に存在し、その後、本モデルの機構によって種内での多様性を更に蓄積してきたのかもしれない。

多くの細菌では、様々な環境にตอบสนองして DNA レベルの再編成を引き起こし、表現型を変化させて生存を計る、相変異 phase variation と呼ばれる過程が知られている¹³⁾。宿主の免疫系の抗原となる分子の多様性を生み出す場合は特に抗原変異 antigenic variation と呼ばれる¹³⁾。時にそれらは細菌の病原性にも関与する。これらの過程にパラログ群が関わる例が多く知られている¹⁴⁾。Lpl 遺伝子産物がどのような生理学的な意義を担っているのか、相変異や抗原変異に関わる因子なのかについては不明である。しかし、もし他の細菌で見られるような抗原変異に関わるリポタンパク質なのであれば、なぜこのような多様な配列を持つ遺伝子がいくつも維持されているのか、なぜ central conserved region が塩基配列レベルで非常に高い identity で保存されているのかを説明することができよう。今後の実験的研究の進展が待たれる。

「隣接する ORF をまたいだ配列を再編成の単位として遺伝子レベルの多様化を創出する」という本研究のモデルは、ORF という単位に基づいていたこれまでのパラログ研究・多重遺伝子族研究にない、新たな視点の導入である。

謝辞

本研究に対して終始ご支援いただいたコスメトロジー財団に感謝いたします。本報告書の中の進化解析は、東京大学大学院博士課程学生の鶴剛史君が中心になって進めました。

(参考文献)

- 1) Kuroda M, Ohta T, Uchiyama I, et al., :Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*, Lancet, 357, 1225-1240, 2001.
- 2) Baba T, Takeuchi F, Kuroda M, et al., :Genome and virulence determinants of high virulence community-acquired MRSA, Lancet, 359, 1819-1827, 2002.
- 3) Ohta T, Hirakawa H, Morikawa K, et al., :Nucleotide substitutions in *Staphylococcus aureus* strains, Mu50, Mu3, and N315, DNA Res., 11, 51-56, 2004.
- 4) Holden MT, Feil EJ, Lindsay JA, et al., :Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance, Proc. Natl. Acad. Sci. U. S. A., 101, 9786-9791, 2004.
- 5) Gill SR, Fouts DE, Archer GL, et al., :Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain, J. Bacteriol., 187, 2426-2438, 2005.
- 6) Diep BA, Gill SR, Chang RF, et al., :Complete genome sequence of USA300, an epidemic clone of community-acquired methicillin-resistant *Staphylococcus aureus*, Lancet, 367, 731-739, 2006.
- 7) Herron-Olson L, Fitzgerald JR, Musser JM, et al., :Molecular Correlates of Host Specialization in *Staphylococcus aureus*, PLoS ONE, 2, e1120, 2007.
- 8) Lindsay JA, Holden MT, :*Staphylococcus aureus*: superbug, super genome?, Trends Microbiol., 12, 378-385, 2004.
- 9) Tsuru T, Kawai M, Mizutani-Ui Y, et al., :Evolution of paralogous genes: reconstruction of genome rearrangements through comparison of multiple genomes within *Staphylococcus aureus*, Mol. Biol. Evol., 23, 1269-1285, 2006.
- 10) Fujitani Y, Yamamoto K, Kobayashi I, :Dependence of frequency of homologous recombination on the homology length, Genetics, 140, 797-809, 1995.
- 11) Fujitani Y, Kobayashi I, :Effect of DNA sequence divergence on homologous recombination as analyzed by a random-walk model, Genetics, 153, 1973-1988, 1999.
- 12) Khasanov FK, Zvingila DJ, Zainullin AA, et al., :Homologous recombination between plasmid and chromosomal DNA in *Bacillus subtilis* requires approximately 70 bp of homology, Mol. Gen. Genet., 234, 494-497, 1992.
- 13) Villemur R, Deziel E, :Phase variation and antigenic variation, Mullany P (ed) :The Dynamic Bacterial Genome, Cambridge University Press, New York, 2005, 277-322.
- 14) Hughes D, Norstrom T, :Biological consequences for bacteria of homologous recombination, Mullany P (ed) :The Dynamic Bacterial Genome, Cambridge University Press, New York, 2005, 351-384.