

二種の大腸菌 (K12 株・O157 株) における α 配列の比較解析

政策メディア研究科 鶴野 レイナ

Abstract

大腸菌において、 α 配列 (5'-GCTGGTGG-3') は RecBCD により認識され相同組換えを促進するホットスポットである。K-12 株 (MG1665) においては 1009 個の α 配列が大腸菌ゲノム内に存在し、このうち 989 個もの α 配列が ORF 内に存在している。

コンプリートゲノムが公開された O157 株と K12 株を用いて、 α 配列を遺伝子名により対応させ、数や位置、保存状況などの比較を行なった。O157 では、ゲノムサイズの増幅に伴い、 α 配列の数も多くなっていたが、全体を通じて α 配列の分布はこの 2 種の大腸菌で非常に保存されている。その中で注目すべき点としては、複製終結地点周辺に 0.5M にわたるゲノムの逆位領域が見受けられ、この逆位に伴い同様に α 配列も綺麗に対応して逆位している。

さらにこの 2 種の大腸菌において共通な起源を持つ遺伝子領域と、それぞれの特異的な領域とに分け、これらの領域での α 配列の数と出現頻度を求めた。その結果、それぞれの大腸菌に特異的な領域においては α 配列の出現頻度が低く、 α 配列は自己と他者との認識配列であるという仮説に基づくと、外来 DNA 由来の領域であることが示唆された。

1 はじめに

■ α 配列とは？

大腸菌において、 α 配列とは 8bp の配列で、次のような配列を指す名称である。

```
5'-- GCTGGTGG --3'
3'-- CGACCACC --5'
```

α 配列は、一塩基ずつ切断していく酵素「エキソヌクレアーゼ V」としての役割を果たすタンパク「RecBCD」の標的部位である。

エキソヌクレアーゼ V が α 配列の下流の DSB(double strand break) に結合すると DNA を 3' から 5' の方向に一塩基ずつ切りながら α 配列に向かって移動し、 α 配列に達すると、その配列の 4~6 塩基右側でヘリカーゼの機能に変化する。

```

           $\alpha$  部位  -----3'
5'-----GCTGGTGG  ↑ ↑ * * (RecBCD が螺旋をほどこしながら、
3'-----CGACCACC  ↓ ↓ * * センス鎖を一塩基ずつ切る)
          -----5'
```

さらにその後、ヘリカーゼとしての RecBCD が二重らせんをほどこき、一本鎖領域を広げていく。この 3' 末端一本鎖領域が、組換えの際に他の相同性のある二重鎖と鎖の交換をひき起こす引き金となる。

このモデルによると、組換えの際に必要な遊離した 3' 末端を作り出すための一連の過程で α 配列の存在が必要なのである。つまり、 α 配列が意味するものは、組換えの促進である [1] [2] [3] [4]。

また、大腸菌 α 配列はそのほとんどが遺伝子内に存在しており、アミノ酸に翻訳されている。大腸菌ゲノムでは、元々遺伝子領域が多く、8割が遺伝子領域であるが、それ以上に遺伝子領域における α 配列の出現頻度が高いと言える [7]。

またもう一つの特徴として、大腸菌の α 配列は期待値以上に出現する配列であるということだ。マルコフ解析を用いて 2塩基頻度を元に解析した結果でも、大腸菌 α 配列の出現頻度は非常に高いことがわかる。

■ 大腸菌 2 種と α 配列の特徴

大腸菌 2 種 (K12 株・O157 株) の特徴と α 配列の数に関する情報を以下にまとめた。O157 株も K12 MG1665 の特徴であるように、L-V に翻訳される割合が多いことがわかる。

	K12 株	O157 株
ゲノムの大きさ	4639221 bp	5528970 bp
遺伝子数	4405 個	5349 個
α 配列の数 (遺伝子内)	1009 (989)	1125 (1097)
L-V に翻訳される割合	68.5 %	66.2 %

■ 組換えの初期のメカニズム

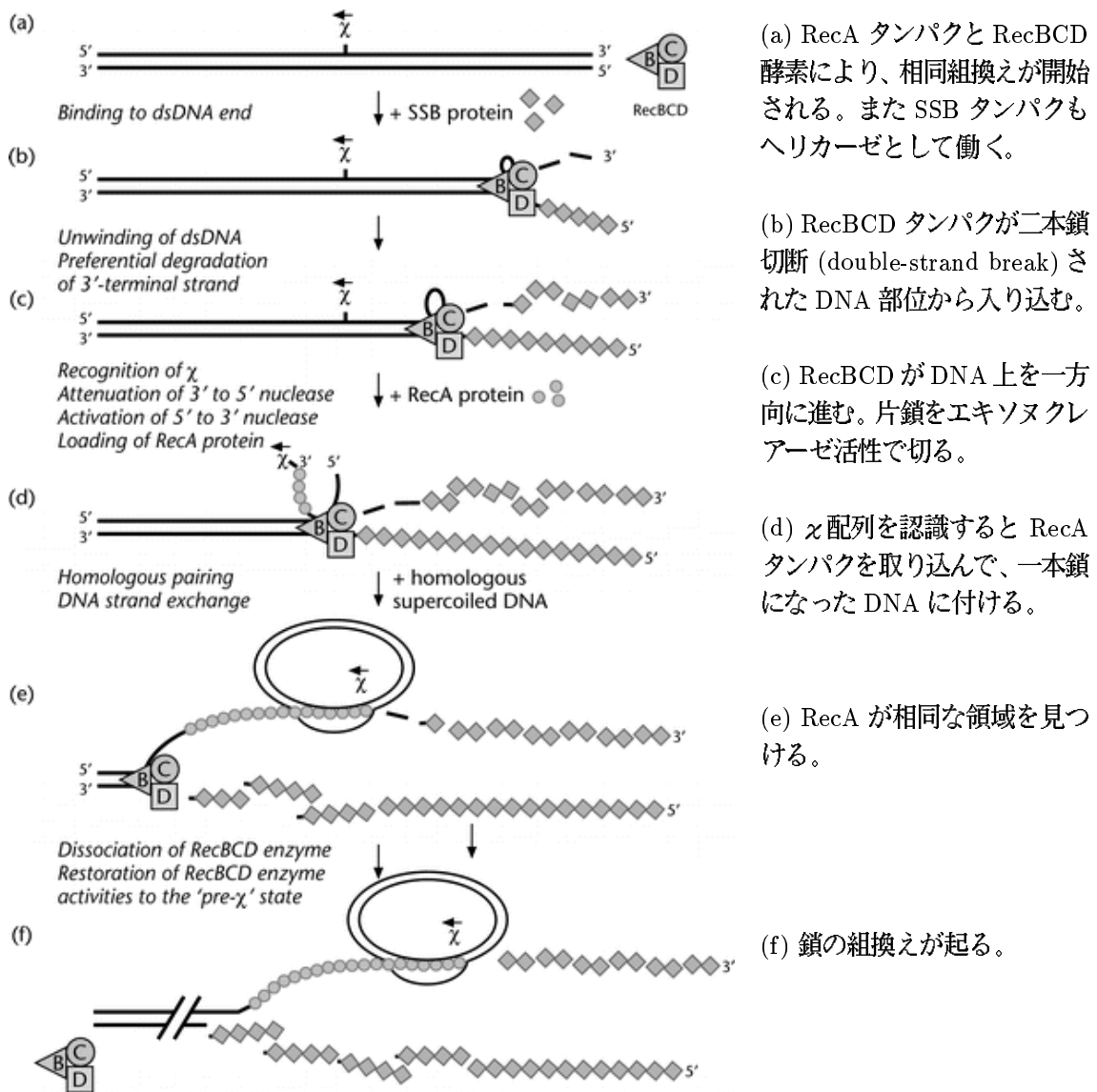


図1: 大腸菌 χ 配列の組換え機構 [6]

この RecBCD 経路の組換え機構は、一度にゲノム約10KBの領域を組換えることが可能である。大腸菌 χ 配列はゲノム内で高頻度で出現し、組換え可能な範囲を10KBとすると(ゲノムの何処に傷を受けようとも、修復の際に)十分カバーできるだけの χ 配列の数が存在し、且つある程度分散している。

「高頻度で出現し、ゲノム全体に分散している」という傾向は、組換え機能を促進する χ 配列が実験的に証明されている大腸菌以外のバクテリア3種においても、共通の傾向であるといえる。

2 目的

o157 株のコンプライートゲノムが解読され、ゲノムデータが公開されたことにより 2 種の大腸菌において α 配列の比較が可能となった [7][8]。同じ大腸菌という種であることから、 α 配列は共に 5'GCTGGTGG 3'であり、これらがゲノムのどの位置に存在するかを比較を行なう。

α 配列がゲノム全体で、どの辺りに分布しているのか、また K12 株の大腸菌と比較するために以下のような解析を行った。

- ◆ α 配列の位置をプロットし、 α 配列の方向を解析する。
- ◆ 2 種大腸菌におけるの α 配列の対応を調べる。
- ◆ ゲノムにおける「 α 配列の割合」を探る。
- ◆ 2 種の大腸菌に共通な領域と、「K-island」「O-island」における α 配列の出現頻度を求める。
(2 種の大腸菌の共通領域と各大腸菌に特徴的な配列を持つ領域とに分け、各領域に α 配列がどれくらいの数で存在しているのかを求める)

3 手法

2 種の大腸菌の α 配列を対応させるための手法として、2 種の大腸菌でどの遺伝子上に存在しているかにより、GENBANK ファイルにある遺伝子名を手がかりに対応表を作成した。よって、遺伝子領域以外の α 配列に関しては図 3・4 のグラフでは表示されていない。

さらに遺伝子領域内での α 配列の方向性が保存されているかを調べるためにそれぞれの大腸菌で α 配列がどちらの方向を向いているかという情報を加えて対応する α 配列のグラフを作成した。

生物種として、この 2 種の大腸菌は非常に似通っており、さらにこの 2 種の大腸菌において (1)「共通な起源を持つ領域」(2)「K12 株に特異的な領域」(3)「o157 株に特異的な領域」と分け、それぞれにおける α 配列の数や、 α 配列が保存されているか否かに関する解析を行なった。

この領域の分け方は、前述した「遺伝子名によって対応させる」手法を用いた。

なお、O157 の論文では各大腸菌に特徴的な配列を持つ領域を表す言葉として「K-island」「O-island」という言葉を用いているが、このキーワードを使ってデータベースから切り出した配列を利用したわけではなく、今回の解析では独自に遺伝子名を対応させて共通領域を求めた。

これら単語は O157 のコンプライートゲノムファイルに「O-island」という記述があったものの、O157 ゲノムの解析が行われる前にデータベース化された K12 では「K-island」という言葉はなく、キーワードにより領域を指定することは不可能であった。故に、O157 ゲノムの論文にある「K-island」「O-island」という言葉と、今回の解析で用いたこれらの用語の定義が必ずしも一致していない。

4 結果

4.1 x 配列の「位置」を二種の大腸菌で比較する

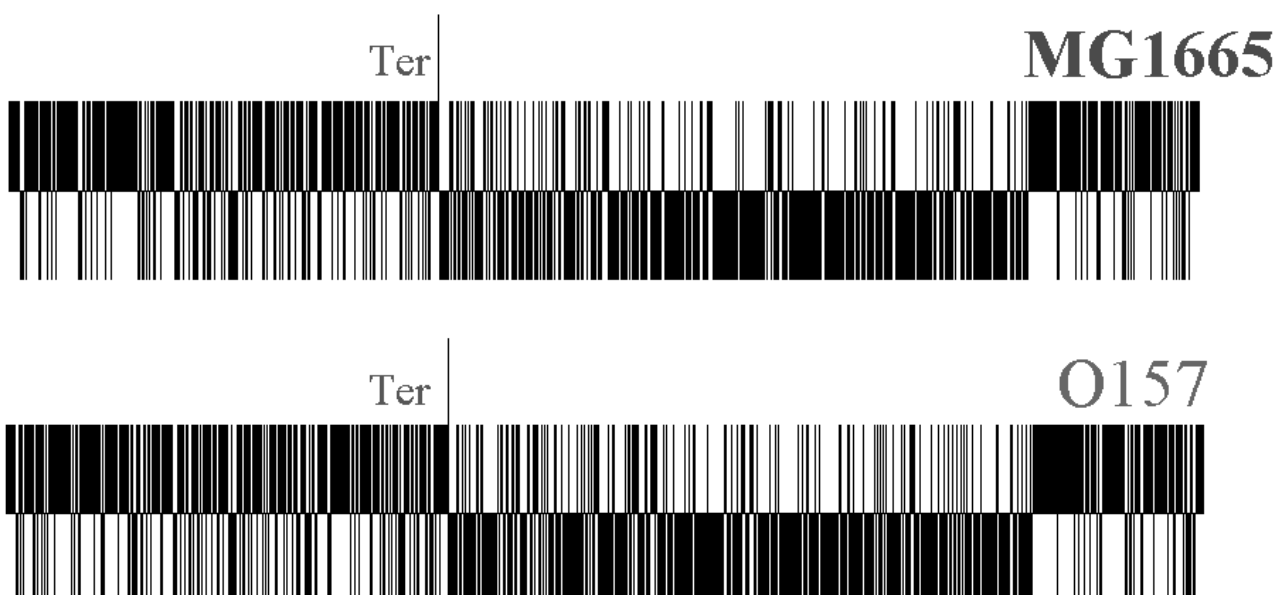


図2: 2種類の大腸菌における x 配列の存在位置

コンプライートゲノム配列を読み込み、direct strand にあるもの、complement strand にあるものを分けてグラフにプロットした。

図2を見る限りでは、大腸菌2種では x 配列の位置が非常に似通っている。大腸菌2種でゲノムそのものでも大幅に変化している領域が少ないことから、多くの x 配列が保存されていることが示唆される。

また図3・4より、K12株・O157株それぞれのゲノム上で x 配列が少ない領域には適当な間隔で（進化の過程において） x 配列が形成されてきたように思われる。線で結んで対応されていない横軸に示した点が、各大腸菌に特有の x 配列を指している。

さらに x 配列が2種の大腸菌における保存割合を調べるため、 x 配列が乗っている遺伝子名でそれぞれの大腸菌の x 配列を対応させ、グラフを作成した。図3・4から Terminus 付近でゲノムの逆位 (inversion) が起っているため [8] x 配列も逆位に伴って同様に移動している事がわかる。この際に x 配列もゲノムの逆位と同様、 x も「逆位」を起している。さらに注目すべき点はこの逆位が起った領域が Terminus 付近であるため、たとえ逆位が起っても遺伝子や x 配列の方向は GCskew に反さず、また複製方向と x 配列の向きも（逆位が起っても）相反することはない。故に x 配列は一定の方向性を保ったまま、綺麗に逆位を起していることがわかる。

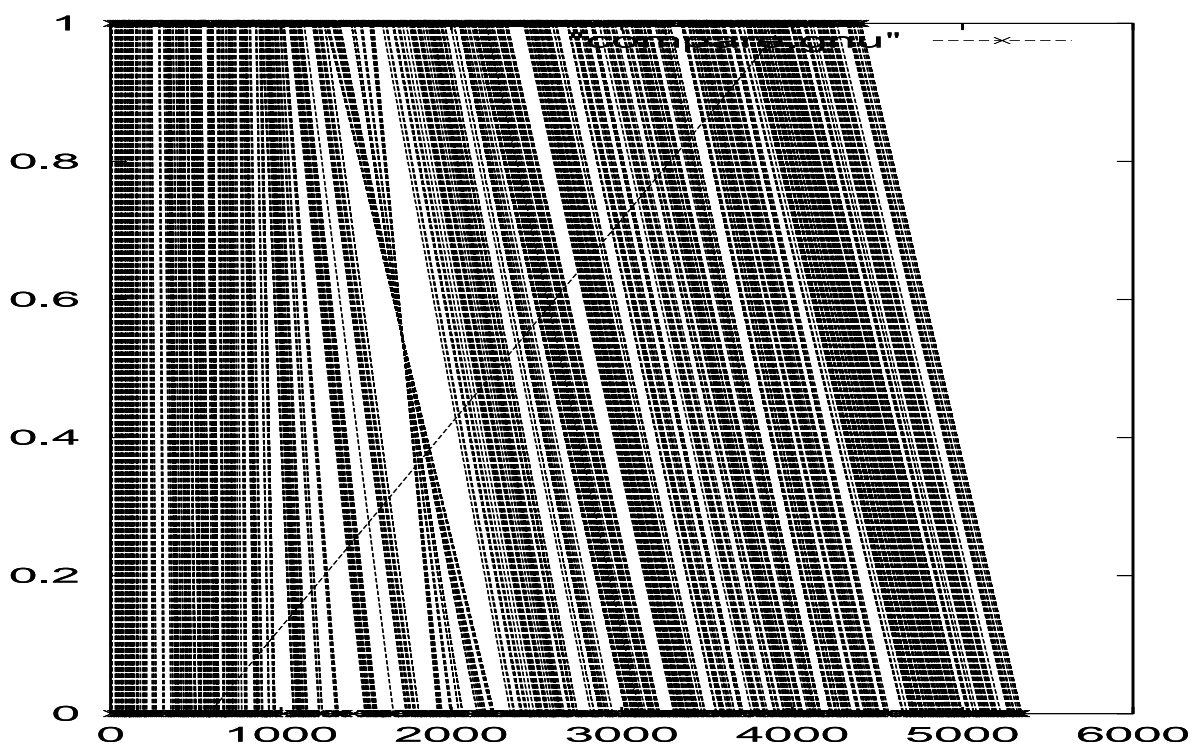


図3: λ 配列の位置を2種で比較

図4では Terminus 付近を拡大した。(横軸は base pair ではなく、遺伝子の並び順を示している)

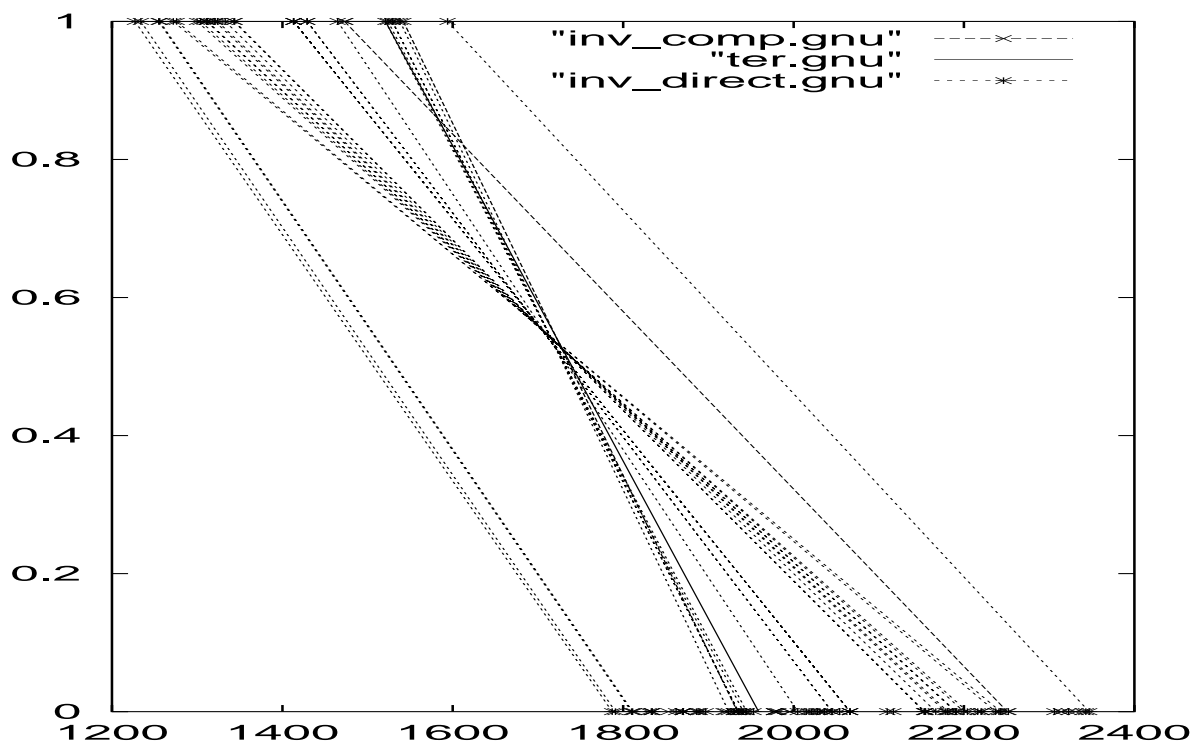
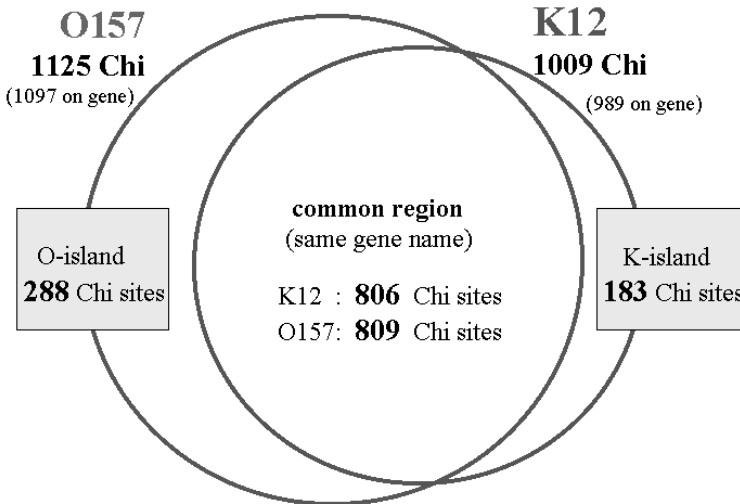
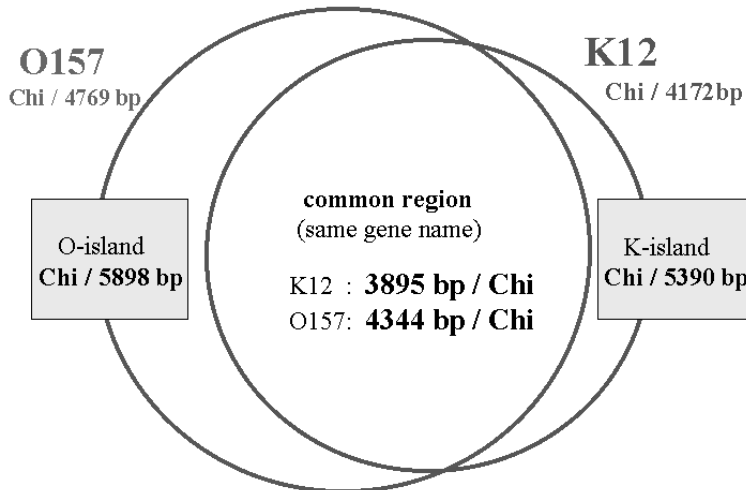


図4: Terminus 付近の λ 配列は 遺伝子と同様に inversion を起している

4.2 ゲノムにおける「 χ 配列の割合」を探る図5: χ 配列の数に関する比較データ

大腸菌2種において、共通領域（同じ遺伝子名で検索したもの）では χ 配列はほとんど保存されていることがわかる。この2種の大腸菌は、450万年前に分化したと言われている [8]。そこに水平伝播などで外来のゲノムが挿入されたなどで現在のような配列になったと考えられるため、この「O-island」や「K-island」（O株やK株にそれぞれ特異的な領域）での χ 配列の出現する割合を求めた。

図6: 新しく出来たゲノムの領域には χ 配列は少ない

上記の図より、2種の大腸菌に特異的な領域では χ 配列の出現頻度が比較的低いことがわかる。外来の遺伝子が入り込んだ際に χ 配列は自己のゲノムと他種のゲノムを区別するための配列であると言われている [?]。RecBCDがエキソヌクレアーゼの役割を果たし、外来DNAを切断する。一方、自己のゲノムでは χ 配列がゲノム上に散在しているため、エキソヌクレアーゼで切断されたとしてもすぐにDNAの切断は停止する。今回の解析における、ゲノムの長さに対する χ 配列の割合は、こういった理由のため両種に特異的な配列では χ 配列の出現頻度が低いのであろう。

5 考察

5.1 ゲノムの保存と χ 配列の方向性

K12株とO157株の比較により、複製終結点(Terminus)付近で逆位領域が発見された[8]。この「逆位」があるということは、O157のコンプリートゲノムの論文に記述があったため、逆位に伴う χ 配列の変化を探った。GCskewや χ 配列の方向性を考える際に、逆位という現象は遺伝子の向きは逆になるので、複製方向やGCskewに反することになる。

しかしO157株の場合、複製終結点周辺領域での逆位であるからこそ、GCskewにも遺伝子の方向にも反することはなかった。この領域が複製終結点を中心に逆位していることから、この位置(もしくは複製開始点)で逆位が起った場合のみGCskewに全く影響しないのである。

今回の解析結果からこのゲノムの逆位に伴って遺伝子も向きを変え、その遺伝子上に乗っていた χ 配列も方向を変えていることが明らかになった。つまり遺伝子の方向に対する χ 配列の読み枠には全く変化はなく、綺麗に逆位を起している。

こういったことからこの領域で逆位が起った理由は次の二つと考えられる。

- ・複製終結点周辺領域が(複製機構の関係で)逆位しやすいのか?
- ・複製終結点周辺領域での逆位がゲノムの痕跡として残りやすいのか?

もしこの逆位が、複製終結点以外の部分で生じていたのならばGCskewの規制やもしくは複製方向と組換えの相関の生物学的理由などにより、 χ 配列の向きが戻るなどといった何らかの圧力が配列に働いた可能性も考えられる。

ゲノム全体にかかる圧力として、GCskewの規制の度合いが種によって異なるのではないだろうか。故に、例えばGCskewがはっきり見られる大腸菌や枯草菌などにおいては、ゲノムの逆位が起ってもGCskewに従うような何らかの圧力がかかるのではないだろうか。今回のO157の場合では、GCskewがはっきり見られる大腸菌であるが、複製終結点で生じた逆位であるためGCskewに影響を与えなかったことから χ 配列もそのまま保存されたのではないかと考える。

5.2 χ 配列が増える進化的圧力が存在したか

ゲノム内に χ 配列が高頻度で出現する理由としては、進化の過程において「 χ 配列を増やす圧力」がゲノムに働いてきたとも一般的に言われている[15]。

今回の解析では、2種の大腸菌において特異的なゲノム領域では χ の存在割合が低いことから、その領域が外来DNA由来のものである可能性を示唆した。

しかしながら、仮にその領域が外来DNA由来であったものだとした場合、450万年前に分化したといわれるこの2種の大腸菌ゲノム内で前述した「 χ 配列を増やす進化的圧力」が働いているとすれば、ゲノムのどこに傷を受けても修復可能であるという面からその種にとっては有利なので χ 配列はゲノムのどこを見ても既に一定の出現頻度があっても良いのではないだろうか。ただしこの場合、以下の二つの可能性を除く必要はある。(1)「 χ 配列を増やす進化的圧力」は非常に弱いもので、450万年以上かけても不十分である。(2)これらの領域は近年挿入されて、まだ十分に χ 配列は増えきっていない。

各大腸菌の特異的な配列内でも、ある程度の x 配列は見つかっている。これらの x 配列は「進化の過程で（圧力により）増えた」結果として良いのか。この2種でゲノムが非常に保存されていたことから、ゲノムを改変する圧力として、K-island, O-island に x 配列が増えたというよりはむしろ、外部由来のDNAにも元々L-V という潜在的に x 配列になる可能性の高い並びが存在していたためではないだろうか。ただ2種の大腸菌の共通領域よりもL-Vの使用頻度が少なかったため（外来DNA由来であるため？）K-island, O-island などの特異的な領域には x の出現頻度が他の領域と比べて低くなったのではないかと考える。

外部由来のDNAが増えることによりゲノム全体のアミノ酸使用においてL, Vの割合が減ってきたのかもしれない。実際（di-codonでは見ていないが）コドンの使用頻度を調べると次のようになっている。

K12	Leu: 5.26 %	Val: 2.64 %
	↓	↓
0157	Leu: 5.10 %	Val: 2.61 %

言い換えると大腸菌ゲノムのコアなゲノム中での「L, Vが多用されている」という特徴であったものが、外来ゲノムのためにこの特徴が緩和されてしまったのではないかと示唆される。

6 まとめと展望

- 2種の大腸菌において x 配列の分布は全体を通じて非常に似通っている。
- ゲノムが逆位した領域では、 x 配列も逆位した状態で綺麗に保存されている。
- 共通なゲノムの領域では、 x 配列の保存されている割合が高い。
- 各大腸菌に固有の領域では（他の領域と比較した場合） x 配列の出現頻度が低い。

6.1 共通領域と特異領域の di-codon 使用を探る。

「進化の過程で x 配列が増える圧力」が生じたかどうか、もしくは多く存在するアミノ酸を x として利用したのかを調べるために共通領域での di-codon の出現頻度と特異領域での di-codon の出現頻度（この場合は ctg-gtg）を比較することによって、潜在的に x 配列になりうる配列の数を比較する。

6.2 外来 DNA 領域を探す手法を確立。

特異領域が外来DNAによるものであり、そのために x 配列の出現頻度が低いとするならば、 x 配列が実験的に調べられている種において、一定の WINDOW 幅を決めてその中での x の出現頻度を、ゲノム全体での x の出現頻度と比較を行なう。全体の平均値よりも低い領域は外来DNA由来のものである可能性もあるため、この領域に関して Blastなどで起源を探る。この解析を水平伝播が高頻度で起る枯草菌などで行なおうと考えている。

謝辞

富田勝教授、中山洋一専任講師、森浩禎教授、金井昭夫助教授、斎藤輪太郎氏には様々な点で指導をしていただきました。また四津谷健志氏、荒川和晴氏には技術的なサポートをしていただきました。この場をお借りしてお礼申し上げます。

関連図書

- [1] Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D., Rehrauer, W. M.; **Biochemistry of homologous recombination in Escherichia coli.**, (1994) Microbiol. 58, 401-465.
- [2] Smith, G. R., Amundsen, S. K., Dabert, P. and Taylor A. F.; **The initiation and control of homologous recombination in Escherichia coli.**, (1995) Philos Trans R Soc Lond B Biol Sci., 347, 13-20.
- [3] Dixon, D. A. and Kowalczykowski, S. C.; **Homologous Pairing In Vivo Stimulated by the Recombination Hotspot, Chi.**, (1991) Cell, 66, 361-371.
- [4] Eggleston, A. K., West, S. C.; **Recombination initiation: easy as A, B, C, D... chi?** , (1997) Curr. Biol. 7, R745-749.
- [5] Churchill, J. and Kowalczykowski, S., C.; **Identification of the RecA Protein-loading Domain of RecBCD Enzyme**, (2000) J. Mol. Biol. 297, 537-542.
- [6] Roman, L. J., Eggleston, A. K. and Kowalczykowski, S. C.; **Processivity of the DNA helicase activity of Escherichia coli RecBCD enzyme.**, (1992) Journal of Biological Chemistry 267: 4207-4214.
- [7] Blattner, F. R., Plankett III, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y.; **The Complete Genome Sequence of Escherichia coli K-12**, (1997) Science 277, 1453-1462.
- [8] Perna N. T. , Plunkett, G. 3rd, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Lim, A., Dimalanta, E. T., Potamouisis, K. D., Apodaca, J., Anantharaman, T. S., Lin, J., Yen, G., Schwartz, D. C., Welch, R. A., Blattner, F. R.; **Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.** , (2001) Nature 409 529-33.
- [9] Chedin, F., Noirot, P., BiauDET, V., Ehrich, D. S., **A five-nucleotide sequence protects DNA from exonucleolytic degradation by AddAB, the RecBC analogue of Bacillus subtilis.**, (1998) Mol. Microbiol., 29, 1369-1377.
- [10] Tillier, E. R., Collins, R. A.; **The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes.**, (2000) J. Mol. Evol. 50, 249-257.
- [11] Tracy, R. B., Chedin, F. and Kowalczykowski, S. C.; **The Recombination Hot Spot Chi Is Embedded within Islands of Preferred DNA Pairing Sequences in the E.coli Genome**, (1997) Cell 90, 205-206.

- [12] Bell, S.J., Chow, Y. C., Ho, J. Y. K. and Forsdyke, D. R.; **Correlation of Chi orientation with transcription indicates a fundamental relationship between recombination and transcription.**, (1998) *Gene* 216, 285-292.
- [13] BiauDET, V., El Karoui, M., Gruss, A.; **Codon usage can explain GT-rich islands surrounding Chi sites on the Escherichia coli genome.**, (1998) *Mol. Microbiol.* 29, 666-669.
- [14] Colbert, T., Taylor, A. F., Smith, G. R.; **Genomics, Chi sites and codons: 'islands of preferred DNA pairing' are oceans of ORFs.**, (1998) *Trends Genet.* 14, 485-488.
- [15] El Karoui, M., BiauDET, V., Schbath, S., Gruss, A.; **Characteristics of Chi distribution on different bacterial genomes.**, (1999) *Res. Microbiol.* 150, 579-587.
- [16] Colbert, T., Taylor, A. F., Smith, G. R.; **Genomics, Chi sites and codons: 'islands of preferred DNA pairing' are oceans of ORFs.**, (1998) *Trends Genet.* 14, 485-488.